THE 6TH INTERNATIONAL CONFERENCE ON INTEGRATED MODELING AND ANALYSIS IN APPLIED CONTROL AND AUTOMATION

SEPTEMBER 19-21 2012

VIENNA, AUSTRIA

IMAACA 2012

EDITED BY

AGOSTINO BRUZZONE
GENEVIÈVE DAUPHIN-TANGUY
SERGIO JUNCO
YURI MERKURYEV

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2012

ISBN 978-88-97999-04-1 (Paperback)
ISBN 978-88-97999-12-6 (PDF)

FOR PERMISSION TO PUBLISH A COMPLETE PAPER WRITE TO: DIME UNIVERSITY OF GENOA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY. ADDITIONAL COPIES OF THE PROCEEDINGS OF THE *IMAACA* ARE AVAILABLE FROM DIME UNIVERSITY OF GENOA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY.

ISBN 978-88-97999-04-1 (Paperback)
ISBN 978-88-97999-12-6 (PDF)

THE 6TH INTERNATIONAL CONFERENCE ON INTEGRATED MODELING AND ANALYSIS IN APPLIED CONTROL AND AUTOMATION

September 19-21 2012, Vienna, Austria

Organized by

DIME – University of Genoa

Liophant Simulation

Simulation Team

IMCS – International Mediterranean & Latin American Council of Simulation

DIMEG, University of Calabria

MSC-LES, Modeling & Simulation Center, Laboratory of Enterprise Solutions

Modeling and Simulation Center of Excellence (MSCOE)

Latvian Simulation Center – Riga Technical University

Logisim

LSIS – Laboratoire des Sciences de l'Information et des Systemes

MIMOS – Movimento Italiano Modellazione e Simulazione

MITIM Perugia Center – University of Perugia

Brasilian Simulation Center, LAMCE-COPPE-UFRJ

MITIM - McLeod Institute of Technology and Interoperable Modeling and Simulation – Genoa Center

M&SNet - McLeod Modeling and Simulation Network

Latvian Simulation Society

Ecole Superieure d'Ingenierie en Sciences Appliquees

Facultad de Ciencias Exactas. Ingegneria y Agrimensura

University of La Laguna

CIFASIS: CONICET-UNR-UPCAM

INSTICC - Institute for Systems and Technologies of Information, Control and Communication

National Russian Simulation Society

CEA - IFAC

Technically Co-Sponsored

IEEE – Central and South Italy Section Chapter

I3M 2012 Industrial Sponsors

CAL-TEK SRL

LIOTECH LTD

MAST SRL

I3M 2012 Media Partners

Inderscience Publishers – International Journal of Simulation and Process Modeling

Inderscience Publishers – International Journal of Critical Infrastructures

IGI Global – International Journal of Privacy and Health Information Management

Halldale Media Group: Military Simulation and Training Magazine

Halldale Media Group: The Journal for Healthcare Education, Simulation and Training

Euromerci

EDITORS

AGOSTINO BRUZZONE
MITIM-DIME, UNIVERSITY OF GENOA, ITALY
agostino@itim.unige.it

GENEVIÈVE DAUPHIN-TANGUY
ECOLE CENTRALE DE LILLE, FRANCE
genevieve.dauphin-tanguy@ec-lille.fr

SERGIO JUNCO
UNIVERSIDAD NACIONAL DE ROSARIO, ARGENTINA
sjunco@fceia.unr.edu.ar

YURI MERKURYEV
RIGA TECHNICAL UNIVERSITY, LATVIA
merkur@itl.rtu.lv

INTERNATIONAL MULTIDISCIPLINARY MODELING & SIMULATION MULTICONFERENCE, I3M 2012

GENERAL CO-CHAIRS

AGOSTINO BRUZZONE, *MITIM DIME, UNIVERSITY OF GENOA, ITALY*
YURI MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*

PROGRAM CHAIR

FRANCESCO LONGO, *MSC-LES, MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA, ITALY*

THE 6TH INTERNATIONAL CONFERENCE ON INTEGRATED MODELING AND ANALYSIS IN APPLIED CONTROL AND AUTOMATION, IMAACA 2012

CONFERENCE CHAIRS

SERGIO JUNCO, *UNIVERSIDAD NACIONAL DE ROSARIO, ARGENTINA*
GENEVIÈVE DAUPHIN-TANGUY, *ECOLE CENTRALE DE LILLE, FRANCE*

IMAACA 2012 INTERNATIONAL PROGRAM COMMITTEE

Jorge Baliño, *Univ. of São Paulo, Brazil*
Wolfgang Borutzky, *BRS-Univ. Applied Sciences, Sankt Augustin, Germany.*
Belkacem Ould Bouamama, *USTL,France*
Jean-Yves Dieulot, *Polytech'Lille, France*
Alejandro Donaire, *The University of Newcastle, Australia*
Tulga Ersal, *University of Michigan, USA*
Gianluca Gatti, *University of Calabria, Italy*
Julien Gomand, *ENSAM, Aix-en-Provence, France*
Adrian Ilinca, *Université du Québec à Rimouski, Canada*
Fabrizio Leonardi, *FEI, São Bernardo do Campo, Brazil*
Francesco Longo, *University of Calabria, Italy*
Loucas S. Louca, *University of Cyprus, School of Engineering, Cyprus*
Jacinto Marchetti, *INTEC, Santa Fe, Argentina*
Pieter Mosterman, *The Mathworks, USA*
Norberto Nigro, *CIMEC, Santa Fe, Argentina*
Rachid Outbib, *LSIS, UPCAM, France*
P. M. Pathak, *IIT, Roorkee, India*
Ricardo Pérez, *PUC, Chile*
Xavier Roboam, *INPT-LAPLACE, France*
Dominique Sauter, *CRAN, UHP Nancy, France*
Christophe Sueur, *École Centrale Lille, France*
Kazuhiro Tanaka, *Kyushu Institute of Technology, Fukuoka, Japan*
Anand Vaz, *NITJ, Jalandhar, India*
Daniel Viassolo, *Halliburton, Houston, TX, USA*
Costas Tzafestas, *NTUA, Greece*

TRACKS AND WORKSHOP CHAIRS

RENEWABLE ENERGY SYSTEMS. MODELING, SIMULATION AND CONTROL
CHAIRS: RACHID OUTBIB, LSIS, UNIVERSITE DE PROVENCE, MARSEILLE, FRANCE; ADRIAN ILINCA, LREE, UNIVERSITE DU QUEBEC A RIMOUSKI, CANADA

MODELLING AND CONTROL OF MECHATRONIC SYSTEMS
CHAIRS: JULIEN GOMAND, LSIS, ARTS ET METIERS PARISTECH, AIX-EN-PROVENCE, FRANCE; JEAN-YVES DIEULOT, LAGIS, POLYTECH-LILLE, VILLENEUVE D'ASCQ, FRANCE

SUPERVISION SYSTEM DESIGN IN INDUSTRIAL PROCESSES
CHAIRS: BELKACEM OULD BOUAMAMA, ECOLE POLYTECHNIQUE DE LILLE, FRANCE; DOMINIQUE SAUTER, CRAN, UNIVERSITE HENRI-POINCARE, NANCY, FRANCE

ALGEBRAIC CONTROL : THEORY AND APPLICATIONS
CHAIR: CHRISTOPHE SUEUR, LAGIS, ECOLE CENTRALE DE LILLE, FRANCE

MODELING, SIMULATION AND CONTROL OF NONLINEAR OSCILLATING SYSTEMS
CHAIR: GIANLUCA GATTI, DEPT. MECHANICAL ENG., UNIVERSITY OF CALABRIA, ARCAVACATA DI RENDE, ITALY

CONFERENCE CHAIRS' MESSAGE

WELCOME TO IMAACA 2012!

The IMAACA 2012 Venue is one of the most beautiful European cities, Vienna. As a city rich in history and traditions, Vienna is the perfect location for a conference like IMAACA, that year by year is consolidating its reputation as one of the important conferences dealing with Modeling and Simulation as applied to the fields of Automation and Control. As part of I3M 2012, the International Multidisciplinary Modeling and Simulation Multiconference -this year co-locating 6 thematic international conferences on Modeling & Simulation-, IMAACA provides researchers, scientists, simulationists and practitioners with an ideal framework for sharing knowledge about innovative concepts, establishing new collaborations and research projects, and identifying new relevant research areas, also as combination of multiple disciplines that share the common Simulation denominator. To this end IMAACA 2012 provides the I3M delegates with the possibility to attend different thematic tracks on the most interesting areas in automation and control including renewable energy systems, modeling and control of mechatronic systems, supervision system design in industrial processes, theory and applications of algebraic control, modeling, simulation and control of nonlinear systems.

As already happened in the past IMAACA experiences, the key of the IMAACA success is the invaluable contribution of its authors as well as the continuous work of the international program committee that together with the help of reviewers has guaranteed the very high scientific quality of the selected papers. Therefore, we would like to thank each attendee and each member of the IPC, your experience is contributing to shape the future of Modeling and Simulation in Automation and Control.

Finally, before we close we would like to wish to all the attendees a fruitful and enjoyable IMAACA 2012 in Vienna.

Sergio Junco,
Universidad Nacional
de Rosario, Argentina

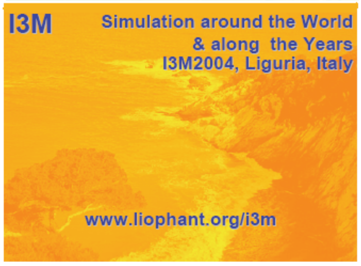
Geneviève Dauphin-Tanguy,
Ecole Centrale de Lille, France

ACKNOWLEDGEMENTS

The IMAACA 2012 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The IMAACA 2012 IPC would like to thank all the authors as well as the reviewers for their invaluable work.
A special thank goes to all the organizations, institutions and societies that have supported and technically sponsored the event.

LOCAL ORGANIZATION COMMITTEE

This International Workshop is part of the I3M Multiconference: the Congress leading **Simulation around the World and Along the Years**



Index

FINITE HORIZON H_∞ CHEAP CONTROL PROBLEM FOR A CLASS OF LINEAR SYSTEMS WITH STATE DELAYS

Valery Y. Glizer

Department of Mathematics, Ort Braude College
P.O. Box 78, Karmiel 21982, Israel

Email valery48@braude.ac.il

ABSTRACT

A finite horizon H-infinity cheap control problem with a given performance level for a linear system with state delays is considered. By a proper transformation of the control variable, this problem is converted to an H-infinity control problem for a singularly perturbed system. For this new problem, parameter-free sufficient conditions of the existence of solution and the solution (controller) itself are obtained. These results are applied to solution of a nonstandard H-infinity control problem.

Keywords: H-infinity cheap control, time-delay system, singular perturbation, nonstandard H-infinity control problem

1. INTRODUCTION

H_∞ control problems are studied extensively in the literature for systems without and with delays in the state variables (Basar and Bernard 1991, Bensoussan, Da Prato, Delfour and Mitter 1992, Doyle, Glover, Khargonekar and Francis 1989, Fridman and Shaked 1998, Fridman and Shaked 2000, van Keulen 1993). The solution of an H_∞ control problem for a linear system can be reduced to a solution of a game-theoretic Riccati equation. In the case of an undelayed system, the Riccati equation is finite dimensional (matrix one), while in the case of a delayed system, it is infinite dimensional (operator one). The operator Riccati equation can be reduced to a hybrid system of three matrix equations of Riccati type. Analysis and solution of this system are very complicated. Therefore, it is extremely important a study of classes of H_∞ control problems with delays, for which the investigation of the operator Riccati equation can be simplified. One of such classes is the class of H_∞ cheap control problems.

The H_∞ cheap control problem is an H_∞ problem with a small control cost (with respect to state and disturbance costs) in the cost functional. A cost functional with a small control cost arises in many topics of control theory. For instance, it arises in the regularization method of a singular optimal control (Bell and Jacobson 1975), in studying the limitations of optimal regulators and filters (Braslavsky, Seron, Maine and Kokotovic 1999, Kwakernaak and Sivan 1972, Seron, Braslavsky, Kokotovic and Mayne 1999), in analysis of control problems with a high control gain (Kokotovic, Khalil and O'Reilly 1986), in the investigation of inverse control problems (Moylan and. Anderson 1973), in the design of a robust control for systems with disturbances (Turetsky and Glizer 2004, Turetsky and Glizer 2011), and some others.

Cheap control problems for systems without disturbances (uncertainties) were widely investigated in the literature. The case of systems with undelayed dynamics was treated more extensively (Bikdash, Nayfeh and Cliff 1993, Jameson and O'Malley 1974/75, Kokotovic, Khalil and O'Reilly 1986, O'Malley and Jameson 1977, Sabery and Sannuti 1987, Smetannikova and Sobolev 2005). The case of systems with delayed dynamics was studied less extensively (Glizer 1999, Glizer 2005, Glizer 2006, Glizer, Fridman and Turetsky 2007, Glizer 2009a). In both cases, an optimal control problem was analyzed. H_∞ cheap control problems have been studied in the literature much less (Glizer 2009b, Toussaint and Basar 2001). It should be noted that two-player zero-sum differential games with a cheap control cost of one of the players in the performance index were analyzed in (Glizer 2000, Petersen 1986, Starr and Ho 1969, Turetsky and Glizer 2004, Turetsky and Glizer 2011). In these works, the case of an undelayed game dynamics and a cheap control cost for the player, minimizing the performance index, was analyzed, which makes the problems, considered in these works, to be close to the H_∞ cheap control problem for a system without delays.

In the present paper, a system with point-wise and distributed state delays and with a square-integrable disturbance is considered. For this system, a finite horizon H_∞ cheap control problem is formulated. A method of asymptotic analysis and solution of the considered H_∞ cheap control problem is proposed. This method is based on: (i) an equivalent transformation of the H_∞ cheap control problem to a new H_∞ problem for a singularly perturbed controlled system; (ii) an asymptotic decomposition of the resulting problem into two much simpler parameter-free subproblems, the slow and fast ones. Using controllers, solving the slow and fast subproblems, a composite controller, solving the transformed problem, is designed. The latter yields a controller, solving the original H_∞ cheap control problem. Note that the algorithm of the analysis of the finite horizon H_∞ cheap control problem, considered in this paper, is similar to that applied in (Glizer 2009b) for analysis of the infinite horizon H_∞ cheap control problem. Along with this, there are essential differences in studying these problems both in main assumptions and techniques.

The results, obtained for the H_∞ cheap control problem, are applied to the solution of a nonstandard H_∞ control problem, i.e. the problem in which the functional does not contain a quadratic control cost.

The following main notations are applied in the paper: (1) E^n is the n-dimensional real Euclidean space; (2) $\|\cdot\|$ denotes the Euclidean norm either of a vector or of a matrix; (3) the prime denotes the transposition of a matrix $A, (A')$ or of a vector $x, (x')$; (4) $L^2[b,c;E^n]$ is the Hilbert space of n-dimensional vector-valued functions square-integrable on the interval $[b,c]$, the norm in this space is denoted as $\|\cdot\|_{L^2[b,c]}$; (5) $C[b,c;E^n]$ is the Banach space of n-dimensional vector-valued functions continuous on the interval $[b,c]$, the norm in this space is denoted as $\|\cdot\|_{C[b,c]}$; (6) I_n is the n-dimensional identity matrix; (7) $\mathrm{col}(x,y)$, where $x \in E^n, y \in E^m$, denotes the column block-vector of the dimension $n+m$ with the upper block x and the lower block y, i.e., $\mathrm{col}(x,y) = (x', y')'$.

2. PROBLEM FORMULATION

2.1. H_∞ Cheap Control Problem

Consider the controlled system

$$\frac{dx(t)}{dt} = A_{11}x(t) + A_{12}y(t) + H_{11}x(t-h) + \int_{-h}^{0} G_{11}(\tau)x(t+\tau)d\tau + F_1 w(t), \tag{1}$$

$$\frac{dy(t)}{dt} = A_{21}x(t) + A_{22}y(t) + H_{21}x(t-h) + \int_{-h}^{0} G_{21}(\tau)x(t+\tau)d\tau + Bu(t) + F_2 w(t), \tag{2}$$

where $t \in [0,T]$; $x(t) \in E^n$, $y(t) \in E^m$, $u(t) \in E^m$, (u is a control), $w(t) \in E^q$, (w is a disturbance); $h > 0$ is a given constant time delay; A_{ij}, $(i,j = 1,2)$, H_{i1}, $G_{i1}(\tau)$, F_i, $(i = 1,2)$ and B are given time-invariant matrices of corresponding dimensions; B has the full rank; the matrix-functions $G_{i1}(\tau)$, $(i = 1,2)$ are piece-wise continuous for $\tau \in [-h,0]$.

Assuming that $w(t) \in L^2[0,T;E^q]$, consider the functional

$$J_\varepsilon(u,w) = \int_0^T [x'(t)D_1 x(t) + y'(t)D_2 y(t) + \varepsilon^2\|u(t)\|^2 - \gamma^2\|w(t)\|^2]dt, \tag{3}$$

where D_1 is symmetric positive-semi-definite, D_2 is symmetric positive-definite matrices; $\gamma > 0$ is a given constant; $\varepsilon > 0$ is a small parameter.

The H_∞ control problem with a performance level γ for the system (1)-(2) is to find a controller $u^*[x(\cdot), y(\cdot)](t)$ that ensures the inequality $J_\varepsilon(u^*, w) \le 0$ along trajectories of (1)-(2) for all $w(t) \in L^2[0,T;E^q]$ and for $x(t) = 0, y(t) = 0, t \le 0.$

The presence of a small multiplier ε^2 in the control cost of the functional (3) means that this problem is the H_∞ cheap control problem.

2.2. Transformation of the Cheap Control Problem

By the control transformation $u(t) = (1/\varepsilon)v(t)$, where v is a new control, the H_∞ problem (1)-(3) becomes

$$\frac{dx(t)}{dt} = A_{11}x(t) + A_{12}y(t) + H_{11}x(t-h) + \int_{-h}^{0} G_{11}(\tau)x(t+\tau)d\tau + F_1 w(t), \tag{4}$$

$$\varepsilon\frac{dy(t)}{dt} = \varepsilon\{A_{21}x(t) + A_{22}y(t) + H_{21}x(t-h) + \int_{-h}^{0} G_{21}(\tau)x(t+\tau)d\tau\} + Bv(t) + \varepsilon F_2 w(t), \tag{5}$$

$$x(t) = 0, \quad y(t) = 0, \quad t \le 0, \tag{6}$$

$$J(v,w) = \int_0^T [x'(t)D_1 x(t) + y'(t)D_2 y(t) + \|v(t)\|^2 - \gamma^2\|w(t)\|^2]dt. \tag{7}$$

Note that the system (4)-(5) is singularly perturbed (Kokotovic, Khalil and O'Reilly 1986). The state variables $x(\cdot)$ and $y(\cdot)$ are the slow and fast ones, respectively. In this system, the slow state variable is with a delay, while the fast state variable is delay free.

In the sequel, we deal with the H_∞ problem consisting of the system (4)-(5), the initial conditions (6) and the cost functional (7), which is called the original H_∞ control problem (OHICP). Once a controller of the OHICP is obtained, the corresponding controller of the H_∞ problem (1)-(3) is obtained by using the equation $u(t) = (1/\varepsilon)v(t)$.

2.3. Solvability Conditions

Consider the following $(n+m)\times(n+m)$-block-matrices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad H = \begin{pmatrix} H_{11} & 0 \\ H_{21} & 0 \end{pmatrix}, \tag{8}$$

$$G(\tau) = \begin{pmatrix} G_{11}(\tau) & 0 \\ G_{21}(\tau) & 0 \end{pmatrix}, \quad D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \tag{9}$$

and the matrix $S(\varepsilon)$

$$\begin{gathered} S(\varepsilon) = \gamma^{-2}\widetilde{F}\widetilde{F}' - \varepsilon^{-2}\widetilde{B}\widetilde{B}', \\ \widetilde{F} = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}, \ \widetilde{B} = \begin{pmatrix} 0 \\ B \end{pmatrix}. \end{gathered} \tag{10}$$

The matrix $S(\varepsilon)$ can be represented in the block form

$$\begin{gathered} S(\varepsilon) = \begin{pmatrix} \gamma^{-2}F_1F_1' & \gamma^{-2}F_1F_2' \\ \gamma^{-2}F_2F_1' & \gamma^{-2}F_2F_2' - \varepsilon^{-2}BB' \end{pmatrix} \\ \overset{\Delta}{=} \begin{pmatrix} S_1 & S_2 \\ S_2' & S_3(\varepsilon) \end{pmatrix}. \end{gathered} \tag{11}$$

Consider the following set of Riccati-type ordinary differential and two partial first-order differential equations for the matrices $P(t)$, $Q(t,\tau)$ and $R(t,\tau,\rho)$ in the domain $\Omega =$ $\{(t,\tau,\rho) : 0 \le t \le T, -h \le \tau \le 0, -h \le \rho \le 0\}$

$$\begin{gathered} \frac{dP(t)}{dt} = -P(t)A - A'P(t) \\ -P(t)S(\varepsilon)P(t) - Q(t,0) - Q'(t,0) - D, \end{gathered} \tag{12}$$

$$\begin{gathered} \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau}\right)Q(t,\tau) = -(A + S(\varepsilon)P(t))'Q(t,\tau) \\ -P(t)G(\tau) - R(t,0,\tau) \end{gathered} \tag{13}$$

$$\begin{gathered} \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau} - \frac{\partial}{\partial \rho}\right)R(t,\tau,\rho) = -G'(\tau)Q(t,\rho) \\ -Q'(t,\tau)G(\rho) - Q'(t,\tau)S(\varepsilon)Q(t,\rho). \end{gathered} \tag{14}$$

The matrices $P(t)$, $Q(t,\tau)$ and $R(t,\tau,\rho)$ satisfy the boundary conditions

$$P(T) = 0, \ Q(T,\tau) = 0, \ R(T,\tau,\rho) = 0, \tag{15}$$

$$Q(t,-h) = P(t)H, \tag{16}$$

$$\begin{gathered} R(t,-h,\tau) = H'Q(t,\tau), \\ R(t,\tau,-h) = Q'(t,\tau)H. \end{gathered} \tag{17}$$

It is seen that the matrix-functions $Q(t,\tau)$ and $R(t,\tau,\rho)$ are present in the set (12)-(14) with deviating arguments. The problem (12)-(17) is, in general, of a high dimension. Moreover, due to the expression for $S(\varepsilon)$, this problem is ill-posed for $\varepsilon \to +0$.

Lemma 2.1. *Let, for a given* $\varepsilon > 0$*, there exist a solution* $\{P(t,\varepsilon), Q(t,\tau,\varepsilon), R(t,\tau,\rho,\varepsilon)\}$ *of (12)-(17) such that*

$$\begin{gathered} P'(t,\varepsilon) = P(t,\varepsilon), \\ R'(t,\tau,\rho,\varepsilon) = R(t,\rho,\tau,\varepsilon). \end{gathered} \tag{18}$$

Then, for this ε*, the controller*

$$\begin{gathered} v^*[x(\cdot),y(\cdot)](t) = -\varepsilon^{-1}B'[P(t,\varepsilon)z(t) \\ + \int_{-h}^{0} Q(t,\tau,\varepsilon)z(t+\tau)d\tau], \ z = col(x,y) \end{gathered} \tag{19}$$

solves the OHICP.

Proof. The lemma is a direct technical extension of the result of (Glizer 2003) (Theorem 2.1 and its proof) where the case of only a point-wise state delay in the controlled system has been considered.

Remark 2.1. Due to the above mentioned features of the problem (12)-(17), verifying the existence of its solution and searching this solution are very complicated tasks.

2.4. Objectives of the Paper

The objectives of this paper are the following:

(i) to derive ε-free reduced conditions, guaranteeing the existence of a controller solving the OHICP for all sufficiently small $\varepsilon > 0$;

(ii) to design a controller much simpler than (19), which being ε-free solves the OHICP for all sufficiently small $\varepsilon > 0$;

(iii) to apply the above mentioned results to the solution of a nonstandard H_∞ control problem.

3. FORMAL CONSTRUCTING A SIMPLIFIED CONTROLLER FOR THE OHICP

In this section, we propose a method of constructing a simplified controller for the OHICP. This method is based on an asymptotic decomposition of the OHICP into two much simpler ε-free subproblems, the slow and fast ones.

3.1. Slow Subproblem

The slow subproblem is obtained from the OHICP by setting there formally $\varepsilon = 0$ and redenoting x, y, v and J by x_s, y_s, v_s and J_s, respectively. Thus, we have

$$\frac{dx_s(t)}{dt} = A_{11}x_s(t) + A_{12}y_s(t) + H_{11}x_s(t-h) + \int_{-h}^{0} G_{11}(\tau)x_s(t+\tau)d\tau + F_1 w(t), \quad t > 0, \tag{20}$$

$$Bv_s(t) = 0, \quad t \in [0,+\infty), \tag{21}$$

$$x_s(t) = 0, \quad t \le 0, \tag{22}$$

$$J_s = \int_0^T [x_s'(t)D_1x_s(t) + y_s'(t)D_2y_s(t) + \|v_s(t)\|^2 - \gamma^2\|w(t)\|^2]dt. \tag{23}$$

Since the matrix B is invertible, the equation (21) yields

$$v_s(t) = 0, \quad t \in [0,+\infty). \tag{24}$$

By substituting (24) into (23), we obtain

$$J_s = \int_0^T [x_s'(t)D_1x_s(t) + y_s'(t)D_2y_s(t) - \gamma^2\|w(t)\|^2]dt. \tag{25}$$

Since the variable $y_s(t)$ does not satisfy any equation for $t \in [0,+\infty)$, one can choose it to satisfy a desirable property of the system (20). This means that the variable $y_s(t)$ can be considered as a control variable in the system (20). Thus, the functional (25), calculated along trajectories of this system, depends on the control variable $y_s(t)$ and the disturbance $w(t) \in L^2[0,T;E^q]$, i.e., $J_s = J_s(y_s, w)$. For the system (20), we can formulate the following H_∞ control problem with a performance level γ: to find a controller $y_s[x_s(\cdot)](t)$ that ensures the inequality $J_s(y_s, w) \le 0$ along trajectories of (20),(22) for all $w(t) \in L^2[0,T;E^q]$. This H_∞ control problem is called the slow H_∞ control subproblem (SHICP) associated with the OHICP.

Consider the following set of Riccati-type matrix ordinary differential and two first-order partial differential equations with deviating arguments:

$$\frac{dP_s(t)}{dt} = -P_s(t)A_{11} - A_{11}'P_s(t) - P_s(t)S_sP_s(t) - Q_s(t,0) - Q_s'(t,0) - D_1, \tag{26}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau}\right)Q_s(t,\tau) = -(A_{11}' + P_s(t)S_s)Q_s(t,\tau) - P_s(t)G_{11}(\tau) - R_s(t,0,\tau), \tag{27}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau} - \frac{\partial}{\partial \rho}\right)R_s(t,\tau,\rho) = -G_1'(\tau)Q_s(t,\rho) - Q_s'(t,\tau)G_1(\rho) - Q_s'(t,\tau)S_sQ_s(t,\rho), \tag{28}$$

where $S_s = \gamma^{-2}F_1F_1' - A_{12}D_2^{-1}A_{12}'$.

The set of equations (26)-(28) is subject to the boundary conditions

$$P_s(T) = 0, \quad Q_s(T,\tau) = 0, \quad R_s(T,\tau,\rho) = 0, \tag{29}$$

$$Q_s(t,-h) = P_s(t)H_{11}, \tag{30}$$

$$R_s(t,-h,\tau) = H_{11}'Q_s(t,\tau), \quad R_s(t,\tau,-h) = Q_s'(t,\tau)H_{11}. \tag{31}$$

Similarly to Lemma 2.1, one has the following proposition.

Proposition 3.1. *Let there exist a solution* $\{P_s(t), Q_s(t,\tau), R_s(t,\tau,\rho)\}$ *of (26)-(31) in the domain* $\boldsymbol{\Omega}$*, such that*

$$P_s'(t) = P_s(t), \quad R_s'(t,\tau,\rho) = R_s(t,\rho,\tau). \tag{32}$$

Then, the controller

$$y_s^*[x_s(\cdot)](t) = -D_2^{-1}A_{12}' \times \left[P_s x_s(t) + \int_{-h}^{0} Q_s(\tau)x_s(t+\tau)d\tau\right] \quad (33)$$

solves the SHICP.

3.2. Fast Subproblem

The fast subproblem is obtained in the following three stages. First, the slow variable $x(\cdot)$ is removed from the equation (5) and the cost functional (7) of the OHICP. Second, the following transformation of variables is made in the resulting problem:

$$\begin{aligned} &t = \varepsilon\xi, \quad y(\varepsilon\xi) = y_f(\xi), \quad v(\varepsilon\xi) = v_f(\xi), \\ &w(\varepsilon\xi) = w_f(\xi), \\ &J(v(\varepsilon\xi), w(\varepsilon\xi)) = \varepsilon J_{f,1}(v_f(\xi), w_f(\xi)), \end{aligned} \quad (34)$$

where ξ, y_f, v_f, w_f and J_f are new independent variable, state, control, disturbance and cost functional, respectively. Thus, we obtain the system and the cost functional

$$\begin{aligned} &dy_f(\xi)/d\xi = \varepsilon A_{22}y_f(\xi) + \\ &Bv_f(\xi) + \varepsilon F_2 w_f(\xi), \quad \xi > 0, \end{aligned} \quad (35)$$

$$\begin{aligned} &J_f(v_f, w_f) = \int_0^{T/\varepsilon} [y_f'(\xi)D_2 y_f(\xi) + \\ &v_f'(\xi)v_f(\xi) - \gamma^2 \|w_f(\xi)\|^2]d\xi. \end{aligned} \quad (36)$$

Finally, neglecting formally the terms with the multiplier ε in (35) and replacing formally T/ε by $+\infty$ in (36) yields the system

$$dy_f(\xi)/d\xi = Bv_f(\xi), \quad \xi > 0. \quad (37)$$

and the cost functional

$$\begin{aligned} &J_f(v_f, w_f) = \int_0^{+\infty} [y_f'(\xi)D_2 y_f(\xi) + \\ &v_f'(\xi)v_f(\xi) - \gamma^2 \|w_f(\xi)\|^2]d\xi. \end{aligned} \quad (38)$$

For the system (37) and the cost functional (38), the H_∞ control problem with a performance level γ can be formulated as follows. To find a controller $v_f^*[y_f(\xi)]$ that stabilizes (37) and ensures the inequality $J_f(v_f^*, w_f) \le 0$ along its trajectories for all $w_f(\xi) \in L_2[0,+\infty; E^q]$ and for $y_f(0) = 0$. This H_∞ control problem is called the fast H_∞ control subproblem (FHICP) associated with the OHICP.

Let K be any $m \times m$ -matrix such that BK is a Hurwitz matrix. Then, the controller

$$v_f^*[y_f(\xi)] = Ky_f(\xi) \quad (39)$$

solves the FHICP.

Note, that the FHICP is a particular case of the infinite horizon H_∞ control problem, considered in (Basar and Bernard 1991). Due to results of this book, if there exists a solution P_f of the algebraic matrix Riccati-type equation

$$-P_f BB' P_f + D_2 = 0, \quad (40)$$

such that $-BB'P_f$ is a Hurwitz matrix, then the matrix gain K in (39) can be chosen as

$$K = -B'P_f. \quad (41)$$

Let us show that the above mentioned solution of (40) exists. Indeed, since the matrix D_2 is positive definite, then there exist the unique positive definite solution of (40)

$$\begin{aligned} &P_f = (BB')^{-1/2} \times \\ &((BB')^{1/2} D_2 (BB')^{1/2})^{1/2} (BB')^{-1/2}, \end{aligned} \quad (42)$$

where the superscript "1/2" denotes the unique symmetric positive definite square root of respective symmetric positive definite matrix, the one "-1/2" denotes the square root of respective inverse matrix.
Now, using (41) and (42), we have

$$\begin{aligned} &BK = -(BB')^{1/2} \times \\ &((BB')^{1/2} D_2 (BB')^{1/2})^{1/2} (BB')^{-1/2}. \end{aligned} \quad (43)$$

Since D_2 is positive definite and B is not singular, then the equation (43) means that the matrix BK is Hurwitz. Hence, the controller

$$v_f^*[y_f(\xi)] = -B'P_f y_f(\xi) \quad (44)$$

solves the FHICP.

3.3. Composite Controller for the OHICP

In this subsection, based on the control $v_s(t)$, given by (24), the controller $y_s^*[x_s(\cdot)](t)$, solving the SHICP, and the controller $v_f^*[y_f(\xi)]$, solving the FHICP, we construct a composite controller for the OHICP. Then, we show that this controller solves the OHICP for all sufficiently small $\varepsilon > 0$.

The composite controller is constructed in the form

$$v_c[x(\cdot), y(\cdot)](t) = v_s(t) + v_f^*[\tilde{y}(t/\varepsilon)], \tag{45}$$

where $\tilde{y}(t/\varepsilon)$ is defined as follows

$$\tilde{y}(t/\varepsilon) \overset{\Delta}{=} y(t) - y_s^*[x(\cdot)](t). \tag{46}$$

Substituting (24) and (44) into (45), and using (33), (42) and (46) yield after some rearrangement

$$\begin{aligned} v_c[x(\cdot),y(\cdot)](t) &= -B'P_f[D_2^{-1}A_{12}'P_s(t)x(t) \\ &+ y(t) + \int_{-h}^{0} D_2^{-1}A_{12}'Q_s(t,\tau)x(t+\tau)d\tau] \end{aligned} \tag{47}$$

4. ε-FREE SOLVABILITY CONDITIONS FOR THE OHICP

Theorem 4.1. *Let there exist a solution* $\{P_s(t), Q_s(t,\tau), R_s(t,\tau,\rho)\}$ *of (26)-(31) in the domain* $\boldsymbol{\Omega}$*, satisfying the conditions (32). Then, there exists a positive number* ε^**, such that the controller (47) solves the OHICP for all* $\varepsilon \in (0, \varepsilon^*]$.

Proof. The proof of the theorem consists of four parts. For the sake of saving the space of the paper, we present here a sketch of the proof.

Part I. By substituting the controller (47) into the system (4)-(5) and the cost functional (7), we obtain

$$\begin{aligned} \frac{dx(t)}{dt} &= A_{11}x(t) + A_{12}y(t) + H_{11}x(t-h) \\ &+ \int_{-h}^{0} G_{11}(\tau)x(t+\tau)d\tau + F_1 w(t), \end{aligned} \tag{48}$$

$$\begin{aligned} \varepsilon\frac{dy(t)}{dt} &= [\varepsilon A_{21} - BB'P_f D_2^{-1}A_{12}'P_s(t)]x(t) \\ &+ [\varepsilon A_{22} - BB'P_f]y(t) + \varepsilon H_{21}x(t-h) \\ &+ \int_{-h}^{0}\left[\varepsilon G_{21}(\tau) - BB'P_f D_2^{-1}A_{12}'Q_s(t,\tau)\right]\times \\ & x(t+\tau)d\tau + \varepsilon F_2 w(t), \end{aligned} \tag{49}$$

$$\begin{aligned} J(\bar{v}_c, w) &\overset{\Delta}{=} J_c(w) = \int_0^T \Big[x'(t)D_{P1}(t)x(t) \\ &+ 2x'(t)D_{P2}(t)y(t) + y'(t)D_{P3}y(t) \\ &+ 2x'(t)\int_{-h}^{0} D_{Q1}(t,\tau)x(t+\tau)d\tau \\ &+ 2y'(t)\int_{-h}^{0} D_{Q2}(t,\tau)x(t+\tau)d\tau \\ &+ \int_{-h}^{0}\int_{-h}^{0} x'(t+\tau)D_{R1}(t,\tau,\rho)x(t+\rho)d\tau d\rho \\ &- \gamma^2\|w(t)\|^2\Big]dt, \end{aligned} \tag{50}$$

where

$$D_{P1}(t) = D_1 + P_s(t)A_{12}D_2^{-1}A_{12}'P_s(t), \tag{51}$$

$$D_{P2}(t) = P_s(t)A_{12}, \; D_{P3} = 2D_2, \tag{52}$$

$$\begin{aligned} D_{Q1}(t,\tau) &= P_s(t)A_{12}D_2^{-1}A_{12}'Q_s(t,\tau), \\ D_{Q2}(t,\tau) &= A_{12}'Q_s(t,\tau), \end{aligned} \tag{53}$$

$$D_{R1}(t,\tau,\rho) = Q_s'(t,\tau)A_{12}D_2^{-1}A_{12}'Q_s(t,\rho). \tag{54}$$

Thus, the proof of the theorem is reduced to a proof of fulfilment of the following inequality for all sufficiently small $\varepsilon > 0$:

$$J_c(w) \le 0 \quad \forall w(\cdot) \in L^2[0,T;E^q], \tag{55}$$

along trajectories of the system (48)-(49) subject to the initial conditions (6).

Part II. Consider the following $(n+m)\times(n+m)$ block matrices

$$\begin{aligned} &\hat{A}(t,\varepsilon) = \\ &\begin{pmatrix} A_{11} & A_{12} \\ A_{21} - \varepsilon^{-1}P_f^{-1}A_{12}'P_s(t) & A_{22} - \varepsilon^{-1}BB'P_f \end{pmatrix}, \end{aligned} \tag{56}$$

$$\begin{aligned} &\hat{G}(t,\tau,\varepsilon) = \\ &\begin{pmatrix} G_{11}(\tau) & 0 \\ G_{21}(\tau) - \varepsilon^{-1}P_f^{-1}A_{12}'Q_s(t,\tau) & 0 \end{pmatrix}, \end{aligned} \tag{57}$$

$$\begin{aligned} D_P(t) &= \begin{pmatrix} D_{P1}(t) & D_{P2}(t) \\ D_{P2}'(t) & D_{P3}(t) \end{pmatrix}, \\ D_Q(t,\tau) &= \begin{pmatrix} D_{Q1}(t,\tau) & 0 \\ D_{Q2}(t,\tau) & 0 \end{pmatrix}, \end{aligned} \tag{58}$$

$$\begin{aligned} D_R(t,\tau,\rho) &= \begin{pmatrix} D_{R1}(t,\tau,\rho) & 0 \\ 0 & 0 \end{pmatrix}, \\ S_F &= \gamma^{-2}FF'. \end{aligned} \tag{59}$$

Consider the following system of ordinary and partial matrix differential equations of Riccati type with respect to $(n+m)\times(n+m)$-matrices $\hat{P}(t)$, $\hat{Q}(t,\tau)$ and $\hat{R}(t,\tau,\rho)$ in the domain Ω:

$$\frac{d\hat P(t)}{dt} = -\hat P(t)\hat A(t,\varepsilon) - \hat A'(t,\varepsilon)\hat P(t) - \hat P(t)S_F\hat P(t) - \hat Q(t,0) - \hat Q'(t,0) - D_P(t), \tag{60}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau}\right)\hat Q(t,\tau) = -(\hat A(t,\varepsilon) + S_F\hat P(t))'\hat Q(t,\tau) - \hat P(t)\hat G(t,\tau,\varepsilon) - \hat R(t,0,\tau) - D_Q(t,\tau), \tag{61}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial \tau} - \frac{\partial}{\partial \rho}\right)\hat R(t,\tau,\rho) = -\hat G'(t,\tau,\varepsilon)\hat Q(t,\rho) - \hat Q'(t,\tau)\hat G(t,\rho,\varepsilon) - \hat Q'(t,\tau)S_F\hat Q(t,\rho) - D_R(t,\tau,\rho). \tag{62}$$

The system (60)-(62) is considered subject to the boundary conditions

$$\hat P(T) = 0, \quad \hat Q(T,\tau) = 0, \quad \hat R(T,\tau,\rho) = 0, \tag{63}$$

$$\hat Q(t,-h) = \hat P(t)H, \quad \hat R(t,-h,\tau) = H'\hat Q(t,\tau), \quad \hat R(t,\tau,-h) = \hat Q'(t,\tau)H, \tag{64}$$

where the matrix H is given in (8).

Let us show the following. If for some $\varepsilon > 0$, the problem (60)-(64) has a solution

$\{\hat P(t,\varepsilon), \hat Q(t,\tau,\varepsilon), \hat R(t,\tau,\rho,\varepsilon)\}$ such that

$$\hat P'(t,\varepsilon) = \hat P(t,\varepsilon), \quad \hat R'(t,\tau,\rho,\varepsilon) = \hat R(t,\rho,\tau,\varepsilon), \quad (t,\tau,\rho) \in \Omega, \tag{65}$$

then, for this ε, the inequality (55) is satisfied along trajectories of the system (48)-(49) subject to the initial conditions (6).

Part III. Consider the following functional, depending on the parameter $t \in [0,T]$, on a vector $\varphi_0 \in E^{n+m}$ and on a function $\varphi_z(\theta) \in L^2[t-h,t;E^{n+m}]$:

$$V[t,\varphi_0,\varphi_z(\theta)] \overset{\Delta}{=} \varphi_0'\hat P(t,\varepsilon)\varphi_0 + 2\varphi_0'\int_{t-h}^{t}\hat Q(t,\theta-t,\varepsilon)\varphi_z(\theta)d\theta + \int_{t-h}^{t}\int_{t-h}^{t}\varphi_z'(\theta)\hat R(t,\theta-t,\sigma-t,\varepsilon)\times \varphi_z(\sigma)d\theta d\sigma. \tag{66}$$

Let, for a given $w(\cdot) \in L^2[0,T;E^q]$, the vector function $\hat z[t,w(\cdot)] = col(\hat x[t,w(t)], \hat y[t,w(\cdot)])$, $t \in [0,T]$, be the solution of the system (48)-(49) subject to the initial conditions (6). Such a solution exists and is unique. Let, for a given $t \in [0,T]$, $\hat V(t) \overset{\Delta}{=} V\big[t,\hat z[t,w(\cdot)],\hat z[\theta,w(\cdot)]\big]$. Calculating the derivative of $\hat V(t)$, we obtain after some rearrangement

$$\frac{d\hat V(t)}{dt} = -\hat z'[t,w(\cdot)]D_P(t)\hat z[t,w(\cdot)] - 2\hat z'[t,w(\cdot)]\int_{t-h}^{t}D_Q(t,\theta-t)\hat z[\theta,w(\cdot)]d\theta - \int_{t-h}^{t}\int_{t-h}^{t}\hat z'[\theta,w(\cdot)]D_R(t,\theta-t,\sigma-t)\times \hat z[\sigma,w(\cdot)]d\theta d\sigma + \gamma^2\|w(t)\|^2 - \gamma^2\|w(t) - \hat w[t,w(\cdot)]\|^2, \tag{67}$$

where

$$\hat w[t,w(\cdot)] = \gamma^{-2}F'[\hat P(t,\varepsilon)\bar z[t,w(\cdot)] + \int_{t-h}^{t}\hat Q(t,\theta-t,\varepsilon)\bar z[\theta,w(\cdot)]d\theta]. \tag{68}$$

Equation (67) directly yields the inequality $\forall t \in [0,T]$

$$\frac{d\hat V(t)}{dt} + \hat z'[t,w(\cdot)]D_P(t)\hat z[t,w(\cdot)] + 2\hat z'[t,w(\cdot)]\int_{t-h}^{t}D_Q(t,\theta-t)\hat z[\theta,w(\cdot)]d\theta + \int_{t-h}^{t}\int_{t-h}^{t}\hat z'[\theta,w(\cdot)]D_R(t,\theta-t,\sigma-t)\times \hat z[\sigma,w(\cdot)]d\theta d\sigma - \gamma^2\|w(t)\|^2 \le 0. \tag{69}$$

Integrating the inequality (69) from $t = 0$ to $t = T$ and using the conditions (6) and (63) immediately yield the fulfilment of the inequality (55) along trajectories of the system (48)-(49) subject to the initial conditions (6).

Part IV. The existence of solution to the problem (60)-(64) for all sufficiently small $\varepsilon > 0$, which satisfies the conditions (65), is shown by formal constructing and justifying the zero-order asymptotic solution to the problem (60)-(64). This asymptotic solution can be obtained in the way similar (but not identical) to (Glizer 1999). This completes the proof of the theorem.

Corollary 4.1. *Under the conditions of Theorem 4.1, the controller*

$$u_c[x(\cdot),y(\cdot)](t)=(1/\varepsilon)v_c[x(\cdot),y(\cdot)](t)$$

solves the H_∞ *control problem (1)-(3) for all* $\varepsilon\in(0,\varepsilon^*]$.

5. NONSTANDARD H_∞ CONTROL PROBLEM

In this section, we consider the following functional for the system (1)-(2):

$$J(u,w)=\int_0^T[x'(t)D_1x(t)+y'(t)D_2y(t)-\gamma^2\|w(t)\|^2]dt. \qquad (70)$$

It is seen that the functional (70) does not contain a quadratic control cost.

The nonstandard H_∞ control problem with a performance level γ for the system (1)-(2) (NHICP) is to find a controller $u^*[x(\cdot),y(\cdot)](t)$ that ensures the inequality $J(u^*,w)\le 0$ along trajectories of (1)-(2) for all $w(t)\in L^2[0,T;E^q]$ and for $x(t)=0,\ y(t)=0,\ t\le 0.$

Since the functional (70) does not contain a quadratic control cost, the approach, proposed in Lemma 2.1 for the solution of the H_∞ control problem (4)-(7), is not applicable for the solution of the NHICP. In order to solve the NHICP, we replace the functional (70) with the cheap control functional (3). Such a replacing leads to the H_∞ control problem (4)-(7), for which Corollary 4.1 gives ε-free reduced-order solvability conditions, as well as the controller solving this problem. Due to this corollary, the employing the controller $u(t)=u_c[(x(\cdot),y(\cdot)](t)$ in the system (1)-(2) subject to the initial conditions $x(t)=0,y(t)=0,t\le 0$ yields the following inequality for all $w(t)\in L^2[0,T;E^q]$ and all $\varepsilon\in(0,\varepsilon^*]$:

$$J(u_c,w)+\int_0^T\|v_c[x(\cdot),y(\cdot)](t)\|^2dt\le 0, \qquad (71)$$

where the arguments $x(\cdot)$ and $y(\cdot)$ for v_c constitute the solution $col(x(t,\varepsilon),y(t,\varepsilon))$ of the system (48)-(49) subject to the initial conditions (6).

From the inequality (71), one directly has $J(u_c,w)\le 0$, which implies that the controller $u(t)=u_c[(x(\cdot),y(\cdot)](t)$ solves the NHICP for all $\varepsilon\in(0,\varepsilon^*]$ if there exist a solution $\{P_s(t),Q_s(t,\tau),R_s(t,\tau,\rho)\}$ of (26)-(31) in the domain Ω, satisfying the conditions (32).

Remark 5.1. The inequality (71) yields a stronger inequality than $J(u_c,w)\le 0$. Namely,

$$J(u_c,w)\le-\int_0^T\|v_c[x(\cdot,\varepsilon),y(\cdot,\varepsilon)](t)\|^2dt. \qquad (72)$$

The integral in the right-hand side of this inequality depends on $\varepsilon\in(0,\varepsilon^*]$. The following theorem gives an estimate of this integral for small enough $\varepsilon>0$.

Theorem 5.1. *Let there exist a solution* $\{P_s(t),Q_s(t,\tau),R_s(t,\tau,\rho)\}$ *of (26)-(31) in the domain* Ω*, satisfying the conditions (32). Then there exists a positive number* ε_1^**,* $(\varepsilon_1^*\le\varepsilon^*)$*, such that, for any given* $w(t)\in L^2[0,T;E^q]$ *and all* $\varepsilon\in(0,\varepsilon_1^*]$*, the following inequality is satisfied:*

$$0\le\int_0^T\|v_c[x(\cdot,\varepsilon),y(\cdot,\varepsilon)](t)\|^2dt\le a\varepsilon\left(\|w(t)\|_{L^2[0,T]}\right)^2, \qquad (73)$$

where $a>0$ *is some constant independent of* ε.

Proof. In order to save the space, we present here a sketch of the proof.

The left-hand inequality in (73) is obvious. Proceed to the proof of the right-hand one.

Asymptotic analysis of the problem (48)-(49), (6) leads to the existence of a constant $0<\varepsilon_1^*\le\varepsilon^*$ such that, for any given $w(t)\in L^2[0,T;E^q]$ and all $\varepsilon\in(0,\varepsilon_1^*]$, the following inequalities are valid:

$$\|x(t,\varepsilon)-\bar{x}(t)\|_{C[0,T]}\le a_1\varepsilon\|w(t)\|_{L^2[0,T]}, \qquad (74)$$

$$\|y(t,\varepsilon)-\bar{y}(t)\|_{C[0,T]}\le a_1\varepsilon^{1/2}\|w(t)\|_{L^2[0,T]}, \qquad (75)$$

where $a_1>0$ is some constant independent of ε,

$$\bar{x}(t)=\int_0^t\overline{\Phi}_x(t,s)F_1w(s)ds, \qquad (76)$$

$$\bar{y}(t)=\int_0^t\overline{\Phi}_y(t,s)F_1w(s)ds, \qquad (77)$$

the $n\times n$-matrix-valued function $\overline{\Phi}_x(t,s)$ is the solution of the following problem for $0\le s\le t\le T$:

$$\frac{d\overline{\Phi}_x(t,s)}{dt}=\left(A_{11}-A_{12}D_2^{-1}A_{12}'P_s(t)\right)\overline{\Phi}_x(t,s)$$
$$+H_{11}\overline{\Phi}(t-h,s)+\int_{-h}^0\left(G_{11}(\tau)\right.$$

$$-A_{12}D_2^{-1}A'_{12}Q_s(t,\tau)\Big)\overline{\Phi}_x(t+\tau,s)ds, \tag{78}$$

$$\overline{\Phi}_x(t,s)=0,\quad t<s;\quad \overline{\Phi}_x(s,s)=I_n, \tag{79}$$

and the $m\times n$-matrix-valued function $\overline{\Phi}_y(t,s)$ has the form

$$\overline{\Phi}_y(t,s)=-D_2^{-1}A'_{12}P_s(t)\overline{\Phi}_x(t,s)$$

$$-\int_{-h}^{0}D_2^{-1}A'_{12}Q_s(t,\tau)\overline{\Phi}_x(t+\tau,s)d\tau. \tag{80}$$

Now, by using the inequalities (74)-(75) and the equations (76)-(80), one obtains after some rearrangement the inequality

$$\begin{aligned}&\left\|v_c[x(\cdot,\varepsilon),y(\cdot,\varepsilon)](t)\right\|_{C[0,T]}\le\\ &a_2\varepsilon^{1/2}\left\|w(t)\right\|_{L^2[0,T]},\end{aligned} \tag{81}$$

where $a_2>0$ is some constant independent of ε.

The inequality (81) directly yields the right-hand inequality in (73).

6. CONCLUSIONS

In this paper, a linear controlled system with point-wise and distributed state delays and a square-integrable disturbance is considered. For the sake of simplicity, it is assumed that this system consists of two modes. One of them is controlled directly, while the other is controlled through the first one. Moreover, it is considered the case where the state variable of the directly controlled mode has no delays. For this system, the finite horizon H_∞ control problem with a given performance level is studied. The control cost in the cost functional of this problem is assumed to be small with respect to the state and disturbance costs, i.e., the considered problem is the H_∞ cheap control problem. By using a simple control transformation, this problem is converted to the H_∞ control problem for a system with a small multiplier $\varepsilon>0$ for a part of the derivatives, i.e., for a singularly perturbed system. In this singularly perturbed system, the slow state variable has delays, while the fast state variable has not. This new H_∞ control problem, considered as an original one, is analyzed in the sequel of the paper. For this problem, reduced-order solvability conditions, valid for any positive small enough ε, are derived. The ε-free controller, solving this problem for all sufficiently small values $\varepsilon>0$, also is designed. This controller, being multiplied by $1/\varepsilon$, yields the controller, solving the H_∞ cheap control problem. These results are applied to the solution of the nonstandard H_∞ control problem. It is shown that the controller, solving the H_∞ cheap control problem, also solves the nonstandard H_∞ control problem. Moreover, it is shown that this controller ensures the cost functional of the nonstandard H_∞ control problem to be smaller than the negative function of ε, the absolute value of which is of order of ε for all sufficiently small values $\varepsilon>0$.

ACKNOWLEDGMENTS

This research was partially supported by KAMEA fund (Ministry of Immigrant Absorption, State of Israel).

REFERENCES

Basar, T., Bernard, P., 1991. *H^∞ - Optimal control and related minimax design problems: a dynamic games approach*. Boston: Birkhauser.

Bell, D.J., Jacobson, D.H., 1975. *Singular optimal control problems*. New York: Academic Press.

Bensoussan, A., Da Prato, G., Delfour, M.C., Mitter, S.K., 1992. *Representation and control of infinite dimensional systems*. Boston: Birkhauser.

Bikdash, M.U., Nayfeh, A.H., Cliff, E.M., 1993. Singular perturbation of the time-optimal soft-constrained cheap-control problem. *IEEE Transactions on Automatic Control,* 38 (3), 466−469.

Braslavsky, J.H., Seron, M.M., Maine, D.Q., Kokotovic, P.V., 1999. Limiting performance of optimal linear filters. *Automatica*, 35 (2), 189−199.

Doyle, J.C. , Glover, K., Khargonekar, P.P., . Francis, B., 1989. State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control,* 34 (8), 831−847.

Fridman, E., Shaked, U., 1998. H_∞ state-feedback control of linear systems with small state-delay. *Systemsand Control Lett*ers, 33 (3), 141−150.

Fridman, E., Shaked, U., 2000. Finite horizon H_∞ state-feedback control of continuous-time systems with state delays. *IEEE Transactions on Automatic Control,* 45 (12), 2406−2412.

Glizer, V.Y., 1999. Asymptotic solution of a cheap control problem with state delay. *Dynamics and. Control*, 9 (4), 339−357.

Glizer, V.Y., 2000. Asymptotic solution of zero-sum linear-quadratic differential game with cheap control for minimizer. *Nonlinear Differential Equations and Applications,* 7 (2), 231−258.

Glizer, V.Y., 2003. Asymptotic analysis and solution of a finite-horizon H_∞ control problem for singularly-perturbed linear systems with small state delay. *Journal of Optimization Theory and Applications*, 117 (2), 295−325.

Glizer, V.Y., 2005. Suboptimal solution of a cheap control problem for linear systems with multiple state delays. *Journal of Dynamical and Control Systems,* 11 (4), 527−574.

Glizer, V.Y., 2006.. Cheap control problem of linear systems with delays: a singular perturbation approach. In: F. Ceragioli, A. Dontchev, H. Furuta, K. Marti and L. Pandolfi, eds. *Systems, control, modeling and optimization*. IFIP Series 202. New York: Springer, 183−193.

Glizer, V.Y., Fridman, L.M., Turetsky, V., 2007. Cheap suboptimal control of an integral sliding mode for uncertain systems with state delays. *IEEE Transactions on Automatic Control,* 52 (10), 1892−1898.

Glizer, V.Y., 2009a. Infinite horizon cheap control problem for a class of systems with state delays. *International Journal of Nonlinear and Convex Analysis*, 10 (2), 199−233.

Glizer, V.Y., 2009b. H_∞ cheap control for a class of linear systems with state delays. *International Journal of Nonlinear and Convex Analysis*, 10 (2), 235−259.

Jameson, A., O'Malley, R.E., 1974/75. Cheap control of the time-invariant regulator. *Applied Mathematics and Optimization,* 1 (4), 337−354.

Keulen, B. van, 1993. H_∞− *Control for distributed parameter systems: a state-space approach.* Boston: Birkhauser.

Kokotovic, P.V., Khalil, H.K., O'Reilly, J., 1986. *Singular perturbation methods in control: analysis and design*. London: Academic Press.

Kwakernaak, H., Sivan, R., 1972. The maximally achievable accuracy of linear optimal regulators and linear optimal filters. *IEEE Transactions on Automatic Control,* 17 (1), 79−86.

Moylan, P.J., Anderson, B.D.O., 1973. Nonlinear regulator theory on an inverse optimal control problem. *IEEE Transactions on Automatic Control,* 18 (5), 460−465.

O'Malley, R.E., Jameson, A., 1977. Singular perturbations and singular arcs, II. *IEEE Transactions on Automatic Control,* 22 (3), 328−337.

Petersen, I.R., 1986. Linear quadratic differential games with cheap control. *Systems and Control Letters ,* 8 (2), 181−188.

Sabery, A., Sannuti, P., 1987. Cheap and singular controls for linear quadratic regulators. *IEEE Transactions on Automatic Control,* 32 (3), 208−219

Seron, M.M., Braslavsky, J.H., Kokotovic, P.V., Mayne, D.Q., 1999. Feedback limitations in nonlinear systems: from Bode integrals to cheap control. *IEEE Transactions on Automatic Control,* 44 (4), 829−833.

Smetannikova, E.N., Sobolev, V.A., 2005. Regularization of cheap periodic control problems. *Automation and Remote Contro,* 66 (6), 903−916.

Starr, A.W., Ho, Y.C., 1969. Nonzero-sum differential games. *Journal of Optimization Theory and Applications,* 3 (3), 184−206.

Toussaint, G., Basar, T., 2001. Achieving nonvanishing stability regions with high-gain cheap control using H_∞ techniques: the second order case. *Systems and Control Letters*, 44 (2), 79−90.

Turetsky, V., Glizer, V.Y., 2004. Robust state-feedback controllability of linear systems to a hyperplane in a class of bounded controls. *Journal of Optimization Theory and Applications,* 123 (3), 639−667.

Turetsky, V., Glizer, V.Y., 2011. Robust trajectory tracking problem: cheap control approach. *Proceedings of the 5th International Conference on Integrated Modeling and Analysis in Applied Control and Automation*, pp. 26-34. September 12-14, Rome (Italy).

Valery Y. Glizer received the M.Sc. degree in Mathematics from the Dniepropetrovsk State University (Dniepropetrovsk, Ukraine) in 1972. Dr. Valery Glizer received his Ph.D. degree in Mathematics and Physics from the Moscow State University (Moscow, Russia) in 1980. From 1990 to 2007, Dr. Glizer served as a Senior Research Fellow at the Technion – Israel Institute of Technology (Haifa, Israel). In 2007, Dr. Glizer joined the Ort Braude College (Karmiel, Israel). Currently, he is an Associate Professor at the Department of Mathematics in this college. His main research interests are: dynamic games, optimal control, robust control, singular perturbations, and time delay systems. Prof. Glizer is an author/coauthor of more than 160 scientific publications, including the book "Robust Controllability of Linear Systems", Nova Science Publishers, New York, NY, 2012. He is a Member of the Editorial Board for the journals "The Open Automation and Control Systems Journal", "The Open Operational Research Journal", "ISRN Mathematical Analysis Journal" and "Conference Papers in Mathematics". He also serves as a referee for about 30 international scientific journals. Prof. Glizer was awarded for the Excellence in Research at the Ort Braude College in 2009 - 2012. He received The Best Paper Award of The 2011 International Conference of Applied and Engineering Mathematics of The World Congress of Engineering (WCE 2011), London, U.K. Prof. Glizer is a member of the International Society of Dynamic Games.

ROBUST CONTROLLABILITY SETS OF LINEAR AND SATURATED LINEAR STRATEGIES WITH DISCONTINUOUS GAINS

Valery Y. Glizer $^{(a)}$, **Vladimir Turetsky** $^{(b)}$, **Josef Shinar** $^{(c)}$

$^{(a)}$) Department of Mathematics, Ort Braude College, P.O.B. 78, Karmiel 21982, Israel
$^{(b)}$ Department of Mathematics, Ort Braude College, P.O.B. 78, Karmiel 21982, Israel
$^{(c)}$ Faculty of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel

$^{(a)}$) turetsky1@braude.ac.il, $^{(b)}$ valery48@braude.ac.il, $^{(c)}$ aer4301@aerodyne.technion.ac.il

ABSTRACT

A linear system with scalar control and disturbance is considered. The robust controllability of this system to a given target set is studied in the classes of linear and saturated linear strategies. In previous works, the analysis of this problem was limited to continuous sign-constant gains. Such an approach has been inspired by those practical problems, where the control coefficient function in the scalar system, corresponding to the original one, is sign-constant. For systems of sign-varying control coefficient functions using a sign-varying discontinuous gain is proposed. It is shown that using such a gain considerably increases the robust controllability set of the corresponding strategy.

Keywords: linear controlled system, disturbance, linear feedback strategy, saturated linear feedback strategy, robust controllability set

1. INTRODUCTION

Controllability is one of the basic system properties. This property has been well studied for systems without uncertainties by using an open-loop control (Kalman, 1960; Kwakernaak and Sivan, 1972; Bryson and Ho, 1975; Gabasov and Kirillova, 1976). However, this elegant theory is not applicable to real-life systems, affected by unmeasurable input parameters (uncertainties). For such systems, controllability should be augmented by the robustness property with respect to any admissible uncertainty realization. As a rule, the robust controllability can be realized by a feedback control. In the framework of differential games of kind, this property (called sometimes playability) was studied extensively (Isaacs, 1965; Blaquiere et al., 1969; Krasovskii and Subbotin, 1988; Lewin, 1994). There, the input uncertainty (disturbance) is considered as the control of an opponent. Various types of robust feedback controllability of systems with uncertainties were investigated in (Petersen et al., 1992; Savkin, 1997; Savkin and Petersen, 1999; Turetsky and Glizer, 2004; Ganebny et al., 2006).

The general robust controllability concept does not take into account possible control constraints, although such constraints are indispensable part of most practical control problems. For analysis of the robust controllability problem with control constraints, in (Glizer and Turetsky, 2012), it was developed a concept of a *robust controllability set*. According to this concept, the robust controllability set is constructed for a so-called robust transferring feedback strategy, which steers the closed-loop system from the maximal possible set of initial positions to a prescribed target set against any admissible disturbance. The time realizations of such a strategy may violate the prescribed hard control constraints for some initial positions and for some admissible disturbances. By taking into account the hard constraints, the maximal possible set of initial positions is reduced to the robust controllability set. If the system trajectory emanates from any point of this set, the corresponding time realization of the robust transferring strategy satisfes the control constraint, robustly with respect to all admissible disturbances.

In (Glizer and Turetsky, 2012), two classes of robust transferring strategies – linear and saturated linear – have been studied. The gains of these strategies were assumed to be non-zero and smooth, which means that they are sign-constant. Such gains are effective in the case where the control coefficient function in the scalar system, corresponding to the original one, is sign-constant. However, a non-minimum phase controller, which can be found in some applications, leads to a sign- varying control coefficient function. In this case, the robust controllability set of a strategy with a sign-constant gain becomes small or even empty. In this paper, we propose to use non-zero sign-varying gains with the sign opposite to the sign of the control coefficient function. Note that such gains are necessarily discontinuous. It is shown that such extension of the class of admissible gains enlarges considerably the robust controllability sets of linear strategies and of saturated linear strategies.

2. PROBLEM STATEMENT

2.1. Original Controlled System

Consider a controlled dynamic system

$$\dot{x} = A(t)x + b(t)u + c(t)v + f(t), \quad 0 \le t \le t_f, \quad (1)$$

where $x \in R^n$ is a state vector; $u \in R^1$ and $v \in R^1$ are the control and the disturbance, respectively; t_f is a prescribed final instant of time; the matrix-valued function $A(t)$ and the vector-valued functions $b(t)$, $c(t)$, $f(t)$ are continuous on $[0, t_f]$.

The control and the disturbance satisfy the constraints

$$|u| \le \rho_u, \quad (2)$$

$$|v| \le \rho_v, \quad (3)$$

where ρ_u, ρ_v are given positive constants.

Definition 1 *A function* $u = u(t,x)$, $(t,x) \in \mathcal{S}_x \triangleq \{(t,x) : t \in [0,t_f), x \in R^n\}$, *is called an admissible feedback strategy for the system (1), if the corresponding closed-loop system has a unique absolutely continuous solution* $x(t)$, $t \in [0,t_f)$, *for any admissible disturbance* $v(t)$ *and for any initial condition* $x(t_0) = x_0$, $(t_0,x_0) \in \mathcal{S}_x$. *It is also assumed that there exists*

$$x(t_f) \triangleq \lim_{t \to t_f - 0} x(t). \quad (4)$$

The target set is the linear manifold in (t,x)-space

$$\mathcal{T}_x = \{(t,x) : t = t_f, d^T x + d_0 = 0\}, \quad (5)$$

where $d = (d_1, d_2, \ldots, d_n)^T \in R^n$ is a prescribed non-zero vector, d_0 is a prescribed scalar.

The control objective is to bring the system (1) from a given initial position $x(t_0) = x_0$, $(t_0,x_0) \in \mathcal{S}_x$, to the target set (5), respecting the control constraint (2), by means of an admissible feedback strategy $u(t,x)$, for all admissible disturbances $v(t)$.

Definition 2 *An admissible strategy is called robust transferring from a given initial position* $(t_0,x_0) \in \mathcal{S}_x$ *to* $\mathcal{T}_x$, *if for any admissible* $v(t)$: $x(t_f) \in \mathcal{T}_x$. *It is called robust transferring from* $\mathcal{M}_x \subseteq \mathcal{S}_x$ *to* $\mathcal{T}_x$, *if it is robust transferring from any point* $(t_0,x_0) \in \mathcal{M}_x$ *to* $\mathcal{T}_x$.

Let, for a given admissible strategy, the set $\mathcal{M}_x^{\max} \subseteq \mathcal{S}_x$ be the maximal set, from which it is robust transferring to $\mathcal{T}_x$. The set $\mathcal{M}_x^{\max} = \mathcal{M}_x^{\max}(u(\cdot))$ is called the *robust transferrable set* of the strategy $u(\cdot)$.

Definition 3 *The subset* $\mathcal{C}_x = \mathcal{C}_x(u(\cdot))$ *of the robust transferrable set* $\mathcal{M}_x^{\max}$ *is called the robust controllability set of* $u(\cdot)$, *if:*

(i) *for any initial point* $(t_0,x_0) \in \mathcal{C}_x$ *and any admissible disturbance, the time realization of* $u(t,x)$ *along the trajectory* $x = x(t)$ *satisfies the control constraint (2):*

$$|u(t,x(t))| \le \rho_u, \quad t \in [t_0,t_f). \quad (6)$$

(ii) *for any initial point* $(t_0,x_0) \in \mathcal{M}_x^{\max} \setminus \mathcal{C}_x$ *there exist an admissible disturbance and a time moment* $t_1 \in [t_0,t_f)$, *such that*

$$|u(t_1,x(t_1))| > \rho_u. \quad (7)$$

2.2. Scalarization

Let $\Phi(t,\tau)$, $0 \le \tau \le t \le t_f$, be the fundamental matrix of the homogeneous system corresponding to (1). By the non-homogenous transformation of the state variable in (1),

$$z = z(t,x) = d^T\left(\Phi(t_f,t)x + \int_t^{t_f} \Phi(t_f,\tau) f(\tau) d\tau\right) + d_0, \quad (8)$$

this system is reduced (Glizer and Turetsky, 2012; Gutman, 2006) to the scalar equation

$$\dot{z} = h_1(t)u + h_2(t)v, \quad (9)$$

where

$$h_1(t) = d^T \Phi(t_f,t) b(t), \qquad h_2(t) = d^T \Phi(t_f,t) c(t). \quad (10)$$

Note that due to the continuity of $A(t)$, $b(t)$ and $c(t)$, the functions $h_1(t)$ and $h_2(t)$ are continuous on $[0,t_f]$. For the scalar system (9), the target set (5) becomes

$$\mathcal{T}_z = \{t_f, 0\}. \tag{11}$$

For such scalar systems, the definitions 1 – 3 are reformulated.

Definition 4 *A function* $u = u(t, z)$, $(t, z) \in \mathcal{S}_z \triangleq \{(t, z) : t \in [0, t_f), z \in R^z\}$, *is called an admissible feedback strategy for the system (9), if the corresponding closed-loop system has the unique absolutely continuous solution* $z(t)$, $t \in [0, t_f)$, *for any admissible disturbance* $v(t)$ *and for any initial condition* $z(t_0) = z_0$, $(t_0, z_0) \in \mathcal{S}_z$. *It is also assumed that there exists*

$$z(t_f) \triangleq \lim_{t \to t_f - 0} z(t). \tag{12}$$

Definition 5 *An admissible strategy* $u(t, z)$ *is called robust transferring from a given initial position* $(t_0, z_0) \in \mathcal{S}_z$ *to* $\mathcal{T}_z$, *if for any admissible* $v(t)$: $z(t_f) \in \mathcal{T}_z$. *It is called robust transferring from* $\mathcal{M}_z \subseteq \mathcal{S}_z$ *to* $\mathcal{T}_z$, *if it is robust transferring from any point* $(t_0, z_0) \in \mathcal{M}_z$ *to* $\mathcal{T}_z$,

For a given admissible strategy, let the set $\mathcal{M}_z^{\max} \subseteq \mathcal{S}_z$ be the maximal set, from which it is robust transferring to $\mathcal{T}_z$.

Definition 6 *The set* $\mathcal{C}_z = \mathcal{C}_z(u(\cdot)) \subseteq \mathcal{M}_z^{\max}$ *is called the robust controllability set of* $u(\cdot)$, *if for any initial point* $(t_0, z_0) \in \mathcal{C}_z$ *and any admissible disturbance, the time realization of* $u(t, z)$ *along the trajectory* $z = z(t)$ *satisfies the control constraint (2):*

$$| u(t, z(t)) | \le \rho_u, \quad t \in [t_0, t_f). \tag{13}$$

Remark 1 Let the strategy $u(t, z)$ be robust transferring for the system (9) from $\mathcal{M}_z \subseteq \mathcal{S}_z$ to $\mathcal{T}_z$. In (Glizer and Turetsky, 2012), it is proved that if the strategy

$$\tilde{u}(t, x) = u(t, z(t, x)), \tag{14}$$

where $z(t, x)$ is given by (8), is admissible for the system (1), then it is robust transferring for this system from

$$\mathcal{M}_x = \{(t_0, x_0) : (t_0, z(t_0, x_0)) \in \mathcal{M}_z\} \tag{15}$$

to the target set (5). This yields that if $\mathcal{C}_z(u(\cdot))$ is the robust controllability set of $u(t, z)$, then the set

$$\mathcal{C}_x = \{(t_0, x_0) : (t_0, z(t_0, x_0)) \in \mathcal{C}_z(u(\cdot))\}, \tag{16}$$

is the robust controllability set of $\tilde{u}(t, x)$, given by (14).

Remark 1 allows to confine the following analysis only to the scalar case.

2.3. Previous Results

In this section, the main results of the book (Glizer and Turetsky, 2012, Chapters 2 – 3) on the construction of the robust controllability sets for linear and saturated linear robust transferring strategies are briefly summarized.

2.3.1. Linear Strategy

Let us introduce the characteristic numbers $N_1 \ge 0$, C_1, $N_2 \ge 0$, C_2 of the coefficient functions $h_1(t)$ and $h_2(t)$ of (9): assuming that that the limit exists,

$$\lim_{t \to t_f - 0} \frac{h_i(t)}{(t_f - t)^{N_i}} \triangleq C_i \ne 0, i = 1, 2. \tag{17}$$

It is assumed that

$$N_2 \ge N_1. \tag{18}$$

Consider the linear strategy

$$u(t, z) = K(t) z, \tag{19}$$

where the gain function $K(t)$ satisfies the following conditions:

(I) $K(t) \ne 0$ for $t \in [0, t_f)$.

(II) $K(t)$ is continuously differentiable for $t \in [0, t_f)$.

(III) one of two following limit conditions is satisfied:

$$\lim_{t \to t_f - 0} K(t) = +\infty, \tag{20}$$

or

$$\lim_{t \to t_f - 0} K(t) = -\infty, \tag{21}$$

(IV) there exists $N_K > 1$ such that

$$\lim_{t \to t_f - 0} \dot{K}(t)(t_f - t)^{N_K} = C \ne 0, \tag{22}$$

(V) either

$$N_K > N_1 + 2, \text{ and } CC_1 < 0, \tag{23}$$

or

$$N_K = N_1 + 2, \text{ and } CC_1 < -(N_K - 1)^2, \tag{24}$$

The set of all gains $K(t)$, satisfying the conditions (I) – (V), is denoted as $\mathcal{K}$. If the condition (18) is satisfied, then the linear strategy (19) with the gain $K(t) \in \mathcal{K}$ is robust transferring from $\mathcal{S}_z$ to $(t_f, 0)$.

Let us introduce the function

$$\begin{aligned}&\mathcal{P}(t, K(t)) = \\ &\dot{Z}^* - \rho_u(\mathrm{sign}K(t))h_1(t) - \rho_v\,|\,h_2(t)\,|,\end{aligned} \tag{25}$$

where

$$Z^*(t) \triangleq \frac{\rho_u}{|\,K(t)\,|}. \tag{26}$$

Assume that the set of zeros of the function $\mathcal{P}(t, K(t))$ on $(0, t_f)$ is finite (including empty). Due to this assumption, there exists $0 < \delta \le t_f$ such that two cases can be distinguished:

Case 1:

$$\mathcal{P}(t, K(t)) > 0, \;\; t \in (t_f - \delta, t_f). \tag{27}$$

Case 2:

$$\mathcal{P}(t, K(t)) < 0, \;\; t \in (t_f - \delta, t_f). \tag{28}$$

Let for a given $K(t)$, the set $\mathcal{T}$ consist of all distinct zeros of $\mathcal{P}(t, K(t))$ with positive slope. If $\mathcal{T} \ne \varnothing$, it can be written as 2, where $t_1 < t_2 < \ldots < t_p$. Let in this case $Z_i(t_0)$, $i = 1, \ldots, p$, be the solution of the terminal value problem

$$dZ/dt_0 = K(t_0)h_1(t_0)Z + \rho_v\,|\,h_2(t_0)\,|, \tag{29}$$

$$Z(t_i) = Z^*(t_i), \tag{30}$$

on the interval $[0, t_i]$. In Case 2, an additional function $Z_{p+1}(t_0)$ is defined as the solution of the equation (29) with the initial condition

$$Z_{p+1}(0) = \lim_{t \to t_f - 0} F(t), \tag{31}$$

where

$$F(t) \triangleq \frac{\rho_u - \rho_v\,|\,K(t)\,|\displaystyle\int_{t_0}^{t} G(t,\xi)\,|\,h_2(\xi)\,|\,d\xi}{|\,K(t)\,|\,G(t,0)}, \tag{32}$$

$$G(t,\xi) = \exp\!\Big(\int_{\xi}^{t} K(\eta)h_1(\eta)d\eta\Big). \tag{33}$$

Let r be the maximal index ($r \in \overline{1,p}$ in Case 1 or $r \in \overline{1,p+1}$ in Case 2), such that the trajectory $z_0 = Z_r(t_0)$ intersects the t_0-axis, and $t_c = t_c(K(\cdot)) \in (0, t_r)$ be the last time moment such that $Z_r(t_c + 0) > 0$, $Z_r(t_c - 0) < 0$. If no trajectory intersects the t_0-axis, then $t_c = 0$. If $t_c = t_f$, then the robust controllability set $\mathcal{C}_z(K(\cdot))$ of the strategy (19) is empty.

Theorem 1 *Let for a given gain* $K(t) \in \mathcal{K}$, *Case 1 be valid. Then* $\mathcal{C}_z(K(\cdot)) \ne \varnothing$.

If $\mathcal{T} \ne \varnothing$, *then*

$$\begin{aligned}&\mathcal{C}_z(K(\cdot)) = \{(t_0, z_0):\; t_c \le t_0 < t_f, \\ &|\,z_0\,| \le \min\{Z^*(t_0), Z_r(t_0), \ldots, Z_p(t_0)\}\}.\end{aligned} \tag{34a}$$

If $\mathcal{T} = \varnothing$, *then*

$$\mathcal{C}_z(K(\cdot)) = \{(t_0, z_0):\; t_c \le t_0 < t_f,\; |\,z_0\,| \le Z^*(t_0)\}. \tag{34b}$$

Theorem 2 *Let for a given gain* $K(t) \in \mathcal{K}$, *Case 2 be valid. If* $\mathcal{C}_z(K(\cdot)) \ne \varnothing$, *it is given as follows.*

If $\mathcal{T} \ne \varnothing$, *then*

$$\begin{aligned}&\mathcal{C}_z(K(\cdot)) = \{(t_0, z_0):\; t_c \le t_0 < t_f, \\ &|\,z_0\,| \le \min\{Z^*(t_0), Z_r(t_0), \ldots, Z_p(t_0), Z_{p+1}(t_0)\}\}.\end{aligned} \tag{35a}$$

If $\mathcal{T} = \varnothing$, *then*

$$\begin{aligned}&\mathcal{C}_z(K(\cdot)) = \{(t_0, z_0):\; t_c \le t_0 < t_f, \\ &|\,z_0\,| \le \min\{Z^*(t_0), Z_{p+1}(t_0)\}\}.\end{aligned} \tag{35b}$$

Remark 2 It follows from Theorems 1 and 2 that the robust controllability set $\mathcal{C}_z(K(\cdot))$ is symmetric with respect to the t_0-axis, and is described as

$$\begin{aligned}&\mathcal{C}_z(K(\cdot)) = \\ &\{(t_0, z_0):\; t_c \le t_0 < t_f,\; |\,z_0\,| \le Z_b(t_0)\}.\end{aligned} \tag{36}$$

The function $Z_b(t_0)$ denotes the respective boundary functions in (34) – (35).

2.3.2. Saturated Linear Strategy

Let us consider the saturation of (19) for $(t, z) \in \mathcal{S}_z$:

$$u_{sat}(t,z) = \mathrm{Sat}(K(t)z) = \begin{cases} \rho_u, & K(t)z > \rho_u, \\ K(t)z, & |K(t)z| \le \rho_u, \\ -\rho_u, & K(t)z < -\rho_u. \end{cases} \tag{37}$$

Remark 3 This saturated linear strategy respects by definition the control constraint along any trajectory of (9). Therefore, its robust controllability set $\mathcal{C}_z^{sat} = \mathcal{C}_z^{sat}(K(\cdot))$ consists of all the points $(t_0, z_0) \in \mathcal{S}_z$, from which this strategy is robust transferring. In general, the strategy (37) is not robust transferring from the entire set $\mathcal{S}_z$ to the target point $(t_f, 0)$. However, it is robust transferring at least from the robust controllability set $\mathcal{C}_z$ of the strategy (19). In this subset, $K(t)z$ is robust transferring and $|K(t)z(t)| \le \rho_u$. Thus, for any gain $K(t) \in \mathcal{K}$,

$$\mathcal{C}_z^{sat}(K(\cdot)) \supseteq \mathcal{C}_z(K(\cdot)). \tag{38}$$

The inclusion (38) is illustrated by Fig. 1, where $t_f = 4$. The strip $\mathcal{S}_z = \{(t_0, z_0) : t_0 \in [0,4],\ z_0 \in (-\infty, +\infty)\}$ is the robust transferrable set of a linear robust transferring strategy $u(t,z) = K(t)z$. The set, denoted as *I*, consists of all initial positions, from which the time realizations of the strategy $u = K(t)z(t)$ satisfy the constraint (2). This also implies that the saturated linear stratefgy $u = \mathrm{Sat}\left(K(t)z(t)\right)$ is robust transferring from this set. The set *II* is the set of all initial positions, for which $|K(t)z(t)| > \rho_u$, while the saturated linear strategy still remains robust transferring. The set *III* contains all the points of $\mathcal{S}_z$, for which $|K(t)z(t)| > \rho_u$ and the saturated linear strategy is not robust transferring. In other words, $I = \mathcal{C}_z(K(\cdot))$, $II = \mathcal{C}_z^{sat}(K(\cdot)) \setminus \mathcal{C}_z(K(\cdot))$ and $III = \mathcal{S}_z \setminus \mathcal{C}_z^{sat}(K(\cdot))$. From the practical viewpoint, the inclusion (38) means that implementing the saturated linear strategy is preferable than the corresponding linear strategy.

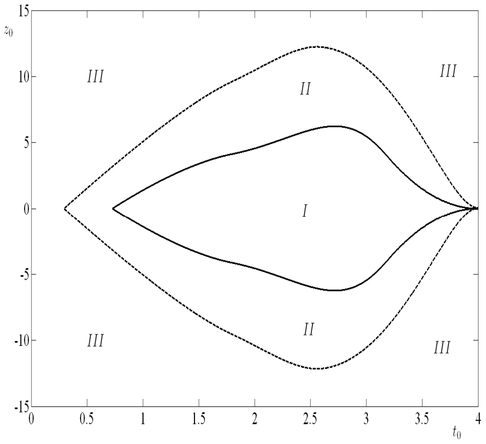


Figure 1: Illustration of the Inclusion (38)

Let $z(t; \bar{t}, \bar{z})$, $t \in [\bar{t}, t_f)$, denote the solution of the initial value problem

$$\dot{z} = h_1(t) u_{sat}(t,z) + h_2(t) v,\ \ z(\bar{t}) = \bar{z}. \tag{39}$$

Theorem 3 *If* $\mathcal{C}_z^{sat}(K(\cdot)) \ne \varnothing$, *it is given by*

$$\mathcal{C}_z^{sat}(K(\cdot)) = \{(t_0, z_0):\ t_0 \in [t_c^s, t_f), |z_0| \le Z_0^s(t_0)\}, \tag{40}$$

where

$$t_c^s = \min\{t_0 \in [0, t_f) : \exists\, z_0 \ge 0 : z(t_f; t_0, z_0) = 0\}, \tag{41}$$

$$Z_0^s(t_0) = z(t_0; t_c^s, z_0^s), \tag{42}$$

$$z_0^s = \max\{z_0 \ge 0 : z(t_f; t_c^s, z_0) = 0\}. \tag{43}$$

Introduce the function

$$Z_m(t) \triangleq \int_t^{t_f} (\rho_u |h_1(\xi)| - \rho_v |h_2(\xi)|) d\xi. \tag{44}$$

Theorem 4 *Let*

$$h_1(t) \ne 0,\ \ t \in [0, t_f). \tag{45}$$

and

$$Z_m(t) > 0,\ \ t \in [0, t_f). \tag{46}$$

Then

$$\mathcal{C}_z^{sat}(K(\cdot)) \subseteq \mathcal{C}_z^m \triangleq \{(t_0, z_0) : t_0 \in [0, t_f), |z_0| \le Z_m(t_0)\}. \tag{47}$$

Moreover,

$$\mathcal{C}_z^{sat}(K(\cdot)) = \mathcal{C}_z^{m}, \qquad (48)$$

if and only if

$$\rho_u / |K(t)| \le Z_m(t),\ t \in [0, t_f). \qquad (49)$$

Remark 4 Due to (Glizer and Turetsky, 2008), for any admissible robust transferring strategy $u(t,z)$:

$$\mathcal{C}_z(u(\cdot)) \subseteq \mathcal{C}_z^{m}. \qquad (50)$$

This means that, subject to the conditions (45), (46) and (49), the robust controllability set $\mathcal{C}_z^{sat}(K(\cdot))$ is maximal for the system (9).

3. NEW RESULTS

For any $K(t) \in \mathcal{K}$, let define the sign-varying gain

$$\tilde{K}(t) = \begin{cases} -\mathrm{sign} h_1(t)\,|K(t)|, & h_1(t) \ne 0, \\ -\mathrm{sign} h_1(t+0)\,|K(t+0)|, & h_1(t) = 0, \end{cases}$$
$$t \in [0, t_f) \qquad (51)$$

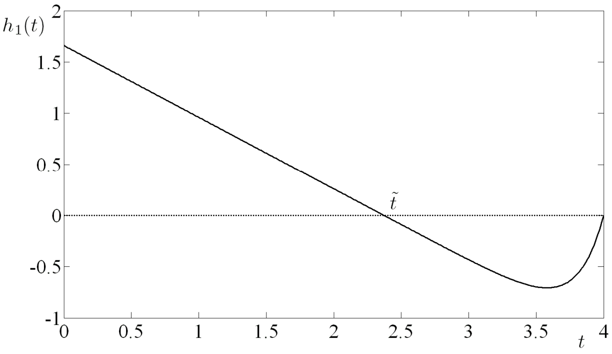


Figure 2a: Illustration of the Equation (51)

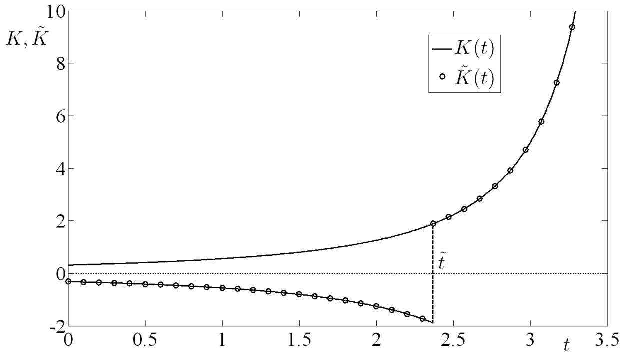


Figure 2b: Illustration of the Equation (51)

The construction of the gain $\tilde{K}(t)$ is illustrated by Fig. 2. The set of all such functions is denoted $\tilde{\mathcal{K}}$.

Remark 5 Due to (51), for all $\tilde{K}(t) \in \tilde{\mathcal{K}}$: $|\tilde{K}(t)| \in \mathcal{K}$.

Let extend the set of admissible gains to the set

$$\mathcal{K}_1 = \mathcal{K} \cup \tilde{\mathcal{K}}. \qquad (52)$$

For $K(t) \in \mathcal{K}$, the robust controllability set of a linear and a saturated linear strategies is constructed by Theorems 1 – 2 and Theorem 3, respectively. In this section, the sets $\mathcal{C}_z(K(\cdot))$ and $\mathcal{C}_z^{sat}(K(\cdot))$ are constructed for $\tilde{K}(t) \in \tilde{\mathcal{K}}$.

3.1. Linear Strategy

Lemma 1 *Let* $K(t) \in \mathcal{K}$ *and* $\tilde{K}(t) \in \tilde{\mathcal{K}}$ *correspond to* $K(t)$ *by (51). Then for any initial position* $(t_0, z_0) \in \mathcal{S}_z$ *and for any admissible disturbance* $v(t)$*, the initial value problem*

$$\dot{z} = h_1(t)\tilde{K}(t)z + h_2(t)v,\ \ z(t_0) = z_0, \qquad (53)$$

is equivalent to the initial value problem

$$\dot{z} = -|h_1(t)||K(t)|z + h_2(t)v,\ \ z(t_0) = z_0. \qquad (54)$$

Proof. This lemma directly follows from the definition (51) of the gain $\tilde{K}(t)$.

□

Theorem 5 *Let* $K(t) \in \mathcal{K}$ *and* $\tilde{K}(t) \in \tilde{\mathcal{K}}$ *correspond to* $K(t)$ *by (51). Let* $h_1(t)$ has only a finite number of distinct zeros on the interval $[0, t_f]$. *Then the robust controllability set* $\mathcal{C}_z(\tilde{K}(\cdot))$ *is constructed by applying Theorems 1 –2 to the system*

$$\dot{z} = -|h_1(t)|u + h_2(t)v, \qquad (55)$$

with the strategy

$$u(t,z) = |K(t)|z. \qquad (56)$$

Moreover,

$$\mathcal{C}_z(\tilde{K}(\cdot)) \supseteq \mathcal{C}_z(K(\cdot)). \qquad (57)$$

Proof. The first statement of the theorem is a direct consequence of Remark 5 and Lemma 1. In order to prove the inclusion (57), it is sufficient to show that if $(t_0, z_0) \in \mathcal{C}_z(K(\cdot))$, then $(t_0, z_0) \in \mathcal{C}_z(\tilde{K}(\cdot))$. For this, let us show that any trajectory of the system (53), starting from $(t_0, z_0) \in \mathcal{C}_z(K(\cdot))$, does not leave $\mathcal{C}_z(K(\cdot))$. Assume the opposite, i.e. that for some initial point $(t_0, z_0) \in \mathcal{C}_z(K(\cdot))$ and for some admissible disturbance $v(t)$, the trajectory $z_2(t; t_0, z_0, v(\cdot))$ of the system (53) leaves $\mathcal{C}_z(K(\cdot))$ through its upper boundary. This means that there exist the time moment $t_1 \in [t_0, t_f)$ and a number $\delta > 0$ such that $z_2(t_1; t_0, z_0, v(\cdot)) = Z_b(t_1)$ and for $t \in (t_1, t_1 + \delta)$:

$$z_2(t; t_0, z_0, v(\cdot)) > Z_b(t). \tag{58}$$

Note that

$$-|h_1(t)K(t)|z \leq h_1(t)K(t)z, \quad z \geq 0. \tag{59}$$

Let $z_1(t; t_1, Z_b(t_1), v(\cdot))$ denote the trajectory of the system (9) for $u = K(t)z$ and the same disturbance $v(t)$ as in (53), starting from the boundary point $(t_1, Z_b(t_1))$. Then, by applying Differential Inequality Theorem (Hartman, 1964) to the systems (53) and (9) with $u = K(t)z$, and by using the inequality (59),

$$\begin{aligned} &z_2(t; t_0, z_0, v(\cdot)) \leq z_1(t; t_1, Z_b(t_1), v(\cdot)), \\ &t \in (t_1, t_1 + \delta_1), \end{aligned} \tag{60}$$

where $0 < \delta_1 \leq \delta$. By definition of the robust controllability set $\mathcal{C}_z(K(\cdot))$,

$$\begin{aligned} &z_1(t; t_1, Z_b(t_1), v(\cdot)) < Z_b(t), \\ &t \in (t_1, t_1 + \delta_1). \end{aligned} \tag{61}$$

The inequalities (60) – (61) contradict the inequality (58), meaning that the trajectory $z_2(t; t_0, z_0, v(\cdot))$ cannot leave $\mathcal{C}_z(K(\cdot))$ through its upper boundary.

The fact that it also cannot leave $\mathcal{C}_z(K(\cdot))$ through its lower boundary, is proved similarly by using the inequality

$$-|h_1(t)K(t)|z \geq h_1(t)K(t)z, \quad z < 0. \tag{62}$$

□

3.2. Saturated Linear Strategy

Similarly to Lemma 1 and Theorem 5 in the case of a linear strategy, the following lemma and theorem hold.

Lemma 2 *Let* $K(t) \in \mathcal{K}$ *and* $\tilde{K}(t) \in \tilde{\mathcal{K}}$ *correspond to* $K(t)$ *by (51). Then for any initial position* $(t_0, z_0) \in \mathcal{S}_z$ *and for any admissible disturbance* $v(t)$, *the initial value problem*

$$\dot{z} = h_1(t)\mathrm{Sat}(\tilde{K}(t))z + h_2(t)v, \; z(t_0) = z_0, \tag{63}$$

is equivalent to the initial value problem

$$\begin{aligned} &\dot{z} = -|h_1(t)|Sat(|K(t)|z) + h_2(t)v, \\ &z(t_0) = z_0. \end{aligned} \tag{64}$$

Theorem 6 *Let* $K(t) \in \mathcal{K}$ *and* $\tilde{K}(t) \in \tilde{\mathcal{K}}$ *correspond to* $K(t)$ *by (51). Let* $h_1(t)$ has only a finite number of distinct zeros on the interval $[0, t_f]$. *Then the robust controllability set* $\mathcal{C}_z^{sat}(\tilde{K}(\cdot))$ *is constructed by applying Theorem 3 to the system*

$$\dot{z} = -|h_1(t)|u + h_2(t)v, \tag{65}$$

with the strategy

$$u(t, z) = Sat(|K(t)|z). \tag{66}$$

Moreover,

$$\mathcal{C}_z^{sat}(\tilde{K}(\cdot)) \supseteq \mathcal{C}_z^{sat}(K(\cdot)). \tag{67}$$

The following theorem is a direct consequence of Theorem 4 and Lemma 2.

Theorem 7 *Subject to the condition (46), the robust controllability set* $\mathcal{C}_z^{sat}(\tilde{K}(\cdot))$ *is maximal for the system (9) if and only if the inequality (49) is satisfied.*

4. INTERCEPTION PROBLEM

In this section, the results of Section 3 are applied to an interception problem with non-minimum phase controllers. Consider a planar engagement between two point-mass objects (pursuer and evader). The velocities V_p and V_e and the bounds of the lateral acceleration commands $a_p^{\max}$ and $a_e^{\max}$ of the objects are constant. The geometry of such planar engagement is presented in Fig. 3.

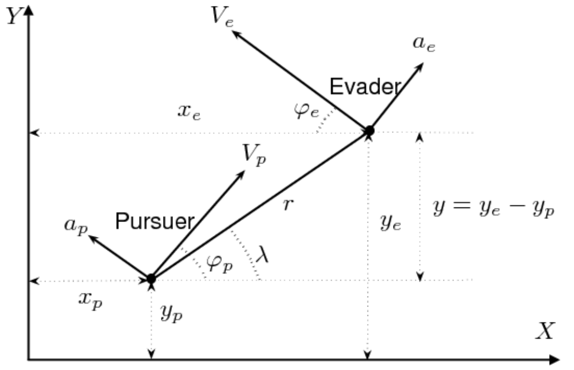


Figure 3: Interception Geometry

In this paper, in contrast with (Glizer and Turetsky, 2012; Shinar, 1981), it is assumed that the controllers of the pursuer and the evader are described by non-minimum phase transfer functions

$$H_i(s) = \frac{s - b_i}{1 + \tau_i s}, i = p, e, \tag{68}$$

where b_p, τ_p, b_e and τ_e are positive constants.

Assuming that the aspect angles φ_p and φ_e are small, the engagement can be modeled by the system (1). In this system, x_1 is the relative separation between the objects, normal to the initial line-of-sight; x_2 is the relative normal velocity. Due to the non-minimum phase form of the transfer functions (68), the variables x_3 and x_4 are connected to the lateral accelerations of the evader and the pursuer by

$$a_e = x_3 + \frac{1}{\tau_e} v, \tag{69}$$

$$a_p = x_4 + \frac{1}{\tau_p} u, \tag{70}$$

The controls of the pursuer u and the evader v are the lateral acceleration commands, satisfying the constraints (2) – (3) with $\rho_u = a_p^{\max}$ and $\rho_v = a_e^{\max}$, respectively. The final time is $t_f = r_0 / (V_p + V_e)$, where r_0 is the initial range between the objects. In this example, $V_p = 700$ m/s, $V_e = 800$ m/s, $r_0 = 6$ km, and, consequently, $t_f = 4$ s. The matrix A is

$$A(t) \equiv \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1/\tau_e & 0 \\ 0 & 0 & 0 & -1/\tau_p \end{bmatrix}, \tag{71}$$

the vectors b and c are

$$\begin{aligned} b(t) &\equiv (0, -1/\tau_p, 0, -(1 + b_p/\tau_p))^T, \\ c(t) &\equiv (0, 1/\tau_e, -(1 + b_e/\tau_e), 0)^T, \end{aligned} \tag{72}$$

$$\begin{aligned} f(t) &\equiv 0, \; x_0 = (0, x_{20}, 0, 0)^T, \\ x_{20} &= V_e \varphi_e(0) - V_p \varphi_p(0). \end{aligned} \tag{73}$$

The objective of the pursuer is to nullify the miss distance $|x_1(t_f)|$, i.e. in the target hyperplane, $d = (1,0,0,0)^T$, $d_0 = 0$.

In the scalar system (9):

$$h_1(t) = (1 + \tau_p b_p)\psi((t_f - t)/\tau_p) - \frac{t_f - t}{\tau_p}, \tag{74}$$

$$h_2(t) = (1 + \tau_e b_e)\psi((t_f - t)/\tau_e) - \frac{t_f - t}{\tau_e}, \tag{75}$$

where $\psi(\xi) \triangleq \exp(-\xi) + \xi - 1$.

For these coefficient functions, $N_1 = N_2 = 1$, $C_1 = -1/\tau_p$, $C_2 = -1/\tau_e$, i.e. the condition (18) holds.

Proposition 1 *If*

$$b_p > \frac{1 - \exp(-t_f/\tau_p)}{\tau_p(\exp(-t_f/\tau_p) + t_f/\tau_p - 1)}, \tag{76}$$

then the function $h_1(t)$, given by (74), changes its sign once in the interval $(0, t_f)$. Moreover, if the inequality(76) is not satisfied, then $h_1(t)$ is sign-constant for $t \in (0, t_f)$.

Proof. Let start with the first statement of the theorem. The inequality (76) is equivalent to the inequality

$$h_1(0) > 0. \tag{77}$$

Note that

$$h_1(t_f) = 0. \tag{78}$$

The derivative of the function $h_1(t)$ is

$$\dot{h}_1(t) = \left(\frac{1}{\tau_p} + b_p\right)\exp(-(t_f - t)/\tau_p) - b_p, \tag{79}$$

yielding

$$\dot{h}_1(t_f) = \frac{1}{\tau_p} > 0. \tag{80}$$

The relations (77) – (78) and (80) guarantee that the function $h_1(t)$ changes its sign at least once in the interval $(0, t_f)$.

Let prove that $h_1(t)$ changes its sign exactly once in the interval $(0, t_f)$. Taking into account the equation (85), it is sufficient to show that the derivative $\dot{h}_1(t)$ has no more than one zero in the interval $(0, t_f)$. Indeed, due to (79), the equation $\dot{h}_1(t) = 0$ can be rewritten as

$$\exp(-(t_f - t)/\tau_p) = \tau_p b_p / (1 + \tau_p b_p), \tag{81}$$

which has no more than one zero in the interval $(0, t_f)$. This completes the proof of the first statement of the proposition. The second statement is proved by similar arguments. □

Consider the linear feedback strategy (19) with the gain

$$K(t) = A / (t_f - t)^2, \tag{82}$$

where $A > 0$. Note that the gain (82) satisfies the conditions (I) – (III). The characteristic numbers of this gain, defined by (22), are $N_K = 3$, $C = 2A$. Note that $N_K = N_1 + 2$ and the condition $CC_1 = -2A/\tau_p < -(N_K - 1)^2 = -4$ is satisfied for

$$A > 2\tau_p. \tag{83}$$

Thus, for such a gain the conditions (IV) – (V) are also satisfied and $K(t) \in \mathcal{K}$. Therefore, the strategy

$$u(t, z) = Az / (t_f - t)^2, \tag{84}$$

where A satisfies (83), is robust transferring from $\mathcal{S}_z = \{(t, z) : t \in [0, t_f),\ z \in R^1\}$ to $(t_f, 0)$.

Example. Let $t_f = 4$ s, $\tau_p = 0.2$ s, $a_p^{\max} = 30$ m/s^2, $b_e = 0.5$ s^{-1}, $\tau_e = 0.2$ s, $a_e^{\max} = 10$ m/s^2, $A = 5$ s. For these parameters, the inequality (76) becomes $b_p > 0.263$. In this example, b_p is chosen as $b_p = 0.7$ s^{-1}. The closed-loop system, corresponding to the gain (82), is

$$\dot{z} = \frac{5h_1(t)}{(4-t)^2} z + h_2(t)v. \tag{85}$$

The graph of the function $h_1(t)$ is depicted in Fig. 2. It is seen that this function changes the sign from positive to negative at $t = \tilde{t} = 2.37$. Thus, due to (51),

$$\tilde{K}(t) = \begin{cases} -5/(4-t)^2, & 0 \le t < \tilde{t}, \\ 5/(4-t)^2, & \tilde{t} \le t < 4. \end{cases} \tag{86}$$

The equivalent closed-loop system (54), corresponding to the gain (86), is

$$\dot{z} = -5\,|h_1(t)|\,z/(4-t)^2 + h_2(t)v. \tag{87}$$

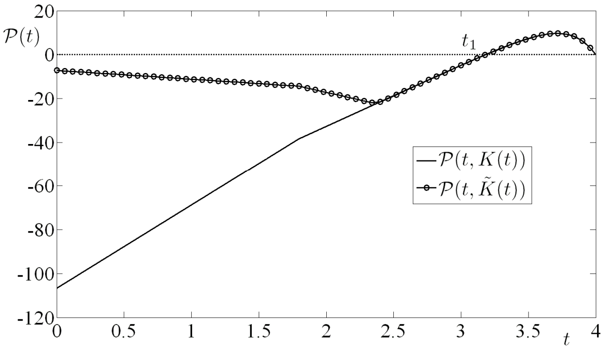


Figure 5: Interception Problem: the Functions $\mathcal{P}(t, K(t))$ and $\mathcal{P}(t, \tilde{K}(t))$

In Fig. 5, the functions $\mathcal{P}(t, K(t))$ and $\mathcal{P}(t, \tilde{K}(t))$ are depicted. It is seen that Case 1 (see (27)) is valid for both gains and $\mathcal{T} = \{t_1\} = \{3.19\}$. In Fig. 6, the robust controllability sets of the linear control strategies $\mathcal{C}_z(K(\cdot))$ and $\mathcal{C}_z(\tilde{K}(\cdot))$ are depicted, demonstrating the advantage of the discontinous-sign gain $\mathcal{C}_z(\tilde{K}(\cdot)) \supset \mathcal{C}_z(K(\cdot))$.

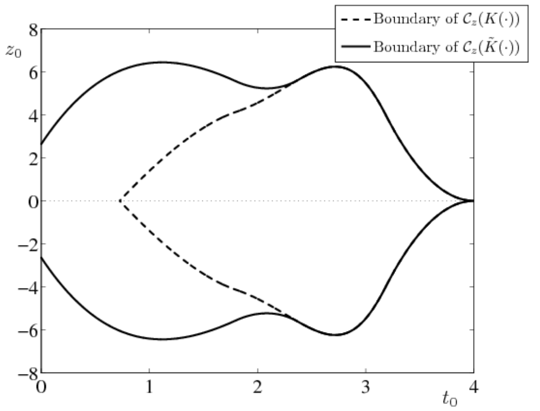


Figure 6: Robust Controllability Sets of the Linear Control Strategies $\mathcal{C}_z(K(\cdot))$ and $\mathcal{C}_z(\tilde{K}(\cdot))$

In Fig. 7, the robust controllability sets of the saturated linear control strategies $\mathcal{C}_z^{sat}(K(\cdot))$ and $\mathcal{C}_z^{sat}(\tilde{K}(\cdot))$ are depicted, showing that, similarly to the case of linear strategies, $\mathcal{C}_z^{sat}(\tilde{K}(\cdot)) \supset \mathcal{C}_z^{sat}(K(\cdot))$.

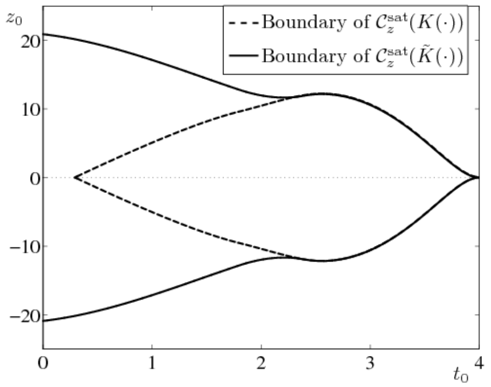


Figure 7: Robust Controllability Sets of the Saturated Linear Control Strategies $\mathcal{C}_z^{sat}(K(\cdot))$ and $\mathcal{C}_z^{sat}(\tilde{K}(\cdot))$

5 CONCLUSIONS

In the paper, a linear controlled system having a sign-varying control coefficient with bounded disturbance is considered. Using for such systems a linear, or a saturated linear control strategy with continous-sign gain leads to a small (sometimes even empty) robust controllability set. Earlier results for constructing the robust controllability sets of linear and saturated linear transferring strategies are extended to the case of a sign-varying discontinuous gain. Such an extension is based on the reducing the sytem with discontinuous gain control to the equivalent system with a corresponding continuous gain control. It is shown that by replacing the continous gain with a properly chosen discontinous gain the robust controllability set is substantially enlarged. It is also shown that the robust controllability set of a saturated linear control strategy is larget than the robust controllability set of the corresponding linear control strategy. These results are illustrated by the example of an interception problem with non-minimum phase controllers of the pursuer and the evader.

REFERENCES

Blaquiere, A., Gerard, F., and Leitmann, G.,1969. *Quantitative and Qualitative Games*. New York: Academic Press.

Bryson, A.E., and Ho, Y.C., 1975. *Applied Optimal Control*. New York, NY: Hemisphere.

Gabasov, R., and Kirillova, F.,1976. *The Qualitative Theory of Optimal Processes. Control and Systems Theory,* vol. 3. New York-Basel: Marcel Dekker, Inc.

Ganebny, S. A., Kumkov, S. S., and Patsko, V. S.,2006. Construction of a Control in Problems with an Unknown Level of Dynamic Noise. *Journal of Applied Mathematics and Mechanics*, 70, 680 – 695.

Glizer, V. Y., and Turetsky, V. 2008. Complete Solution of a Differential Game with Linear Dynamics and Bounded Controls. *Applied Mathematics Research eXpress,* 2008. article ID: abm012, 49 pages.

Glizer, V. Y., and Turetsky, V., 2012. *Robust Controllability of Linear Systems*. New York, NY: Nova Science Publishers.

Gutman, S., 2006. *Applied Min-Max Approach to Missile Guidance and Control.* Progress in Astronautics and Aeronautics, vol. 209. Reston, VI: AIAA, Inc.

Hartman, P., 1964. *Ordinary Differential Equations*. New York - London - Sidney: John Wiley & Sons.

Isaacs, R., 1965. *Differential Games*. New York, NY: John Wiley.

Kalman, R. E., 1960. Contributions to the Theory of Optimal Control. *Boletín de la Sociedád Matematica Mexicana*, 5, 102 – 119.

Krasovskii, N.N., and Subbotin, A.I., 1988. *Game-Theoretical Control Problems.* New York, NY: Springer Verlag.

Kwakernaak, H., and Sivan, R., 1972. *Linear Optimal Control Systems*. New York: Wiley-Interscience.Lewin, J.,1994. *Differential Games*. London: Springer Verlag.

Petersen, I.R., Corless, M., and Ryan, E.P., 1992. A Necessary and Suffcient Condition for Quadratic Finite Time Feedback Controllability. In: Balemi, S., Truol, W., and Mansour, M. (ed), *Robustness of Dynamic Systems with Parameter Uncertainties: Proceedings of the International Workshop on Robust Control, Ascona, Switzerland*. Basel: Birkhauser, 165 – 173.

Savkin, A.V., 1997. Robust Output Feedback Constrained Controllability of Uncertain Linear Time-Varying Systems. *Journal of Mathematical Analysis and Applications*, 215, 376 – 387.

Savkin, A.V., and Petersen, I.R., 1999. Weak Robust Controllability and Observability of Uncertain Linear Systems. *IEEE Transactions on Automatic Control,* 44, 1037 – 1041.

Shinar, J., 1981. Solution techniques for realistic pursuit-evasion games, in: C. Leondes (Ed.), Advances in Control and Dynamic Systems, 17, Academic Press, New York, NY, 1981, pp. 63 – 124.

Turetsky, V., and Glizer, V. Y., 2004. Robust State-Feedback Controllability of Linear Systems to a Hyperplane in a Class of Bounded Controls. *Journal of Optimization Theory and Applications*, 123, 639 – 667.

MODEL PREDICTIVE CONTROL FOR FORMATION KEEPING IN AN ORBIT

Adel Abdulrahman (a), **Mohamad Bagash** (b), **Ossama Abdelkhalik** (c)

(a) Mechanical Engineering Department, Sana'a University, Yemen
(b) Industrial Engineering Department, Taiz University, Yemen
(c) Engineering Mechanics-Mechanical Engineering Department, MTU, USA

(a) galil12@yahoo.com,

ABSTRACT

The MPC algorithm concept is widely used in the process industry, but its application in the formation flaying control is rare. This paper presents a MPC algorithm for the formation in an orbit based on the leader-following approach, the linear model implemented in the MPC algorithm is based on the kepler's nonlinear dynamic equation for the relative position. In the suggested control algorithm, a control is to be applied as long as the formation is moving in a prescribed target interval. As the formation leaves that interval, the formation can be left to move naturally after imposing the proper initial states to cause the formation to return back to that interval with approximately the required configuration.

Keywords: Formation keeping, model predictive control (MPC), Lyapunov function

1. INTRODUCTION

Formation flying has been identified as an enabling technology for many of NASA's twenty-first-century space and earth science missions. These missions will help to revolutionize our understanding of the origin, environment, and the evolution of planetary systems (Mesbahi and Hadaegh 2001). The Air Force has Also identified formation flying as a key technology for the 21st century.

According to (Lawton 2000), three principal approaches have been developed to coordinate spacecraft in formation. These are leader-following, behavior-based, and virtual structure. In the leader-following (LF) approach, one vehicle is chosen to be the leader while the remaining vehicles are designated as followers. The leader is responsible for achieving the position and attitude goals of the formation mission while the followers are responsible for achieving the formation keeping objectives. In other words, the leader tracks a prescribed trajectory while the followers track the leader position and attitude with a prescribed offset.

(Kapila, Sparks, Buffington, and Yan 1999), developed a control for low-earth orbit formation flying in a circular orbit. The Clohessy-Wiltshire (C-W) linear dynamic equations are used as a model for the relative position. These equations were originally developed in the context of the spacecraft rendezvous problem. A pulse-based, discrete time feedback control strategy is developed based on full state feedback control, and a linear quadratic regulator (LQR) approach is used to calculate the gains.

(Queiroz, Kapila, and Yan 2000), proposed an adaptive nonlinear control for the problem of formation keeping and its stability was proved using Lyapunov approach. The full nonlinear position equations were used for the descriptions of the position of the leader and flower spacecrafts.

(McInnes, 1995), used simple analytic commands to bring a loose ring of satellites into a perfect ring formation with uniform intersatellite spacing in a circular orbit. For each spacecraft, the Keplarian equations of motion are used. A potential function is constructed to maintain the relative orientation of spacecraft. A control law is selected such that this potential function is negative definite.

(Abdelkhalik and Alberts 2004), developed a controller for the formation in an elliptic orbit based on the leader following approach. The model of the formation flying used for the controller is the Keplarian's nonlinear dynamic equations for the relative position, the inverse dynamic techniques was applied for developing the control low for the formation flaying problem.

(Manikonda, Arambel, Gopinathan, Mehra, and Hadaegh 1999), combined the feedback linearization and model predictive control (MPC) to design a controller for space formation keeping and attitude control, the model used for the purpose of designing the MPC controller is based on the assumption of no coupling between each space craft. Moreover, (Breger, How and Richards 2005), used Hill's equations of relative motion in circular orbit that governs the spacecraft to remain inside a specified error box for a formation flying control, the model with an assumed noise were implemented in the MPC algorithm for a formation flying control.

Formation members will, in general, naturally drift away from each other when moving in separate orbits. If they were given proper initial relative velocities that are corresponding to their initial relative positions then they will return to their initial configuration after an orbital period. If formation is required to maintain station keeping over a certain target area then the formation can be controlled during this period only and then the formation will be driven to the appropriate initial states for the free flying period.

The MPC algorithm concept is widely used in the process industry (Henson 1998, AbdulRahman, Mokbel and Soufian 2002, Rodrigues and Odloak 2000), but its application in the formation flaying control is rare. This paper presents a MPC algorithm for the formation in an orbit based on the leader-following approach. In the suggested control algorithm, a control is to be applied as long as the formation is moving in a prescribed target interval. As the formation leaves that interval, the formation can be left to move naturally after imposing the proper initial states to cause the formation to return back to that interval with approximately the required configuration. The linear model implemented in the MPC algorithm is the same one that is used by (Abdelkhalik and Alberts 2004). The performance of the MPC algorithm is compared with the performance of the nonlinear control technique based on the inverse dynamic to Keplarian's nonlinear dynamics relative motion.

2. RELATIVE ERROR DYNAMIC MODEL EQUATION

As the leader satellite moves in orbit (figure 1), a certain desired location for the follower satellite also moves with some offset from the leader position. Let the position of the leader satellite be $\vec{r}_l$, the desired position of the follower be $\vec{r}_{des}$, and the follower satellite position be $\vec{r}_f$. The position of the desired position relative to the follower position is:

$$\vec{r}_{df} = \vec{r}_{des} - \vec{r}_f \tag{1}$$

This may be called the error in follower relative position. The desired follower position relative to the leader position is:

$$\vec{r}_{dl} = \vec{r}_{des} - \vec{r}_l \tag{2}$$

The acceleration of the error in follower relative position can then be written as:

$$\ddot{\vec{r}}_{df} = \ddot{\vec{r}}_{des} - \ddot{\vec{r}}_f = \ddot{\vec{r}}_l + \ddot{\vec{r}}_{dl} - \ddot{\vec{r}}_f \tag{3}$$

Recall from Kepler dynamics for two body motion:

$$\ddot{\vec{r}}_l = -\mu \frac{\vec{r}_l}{r_l^3} \tag{4}$$

$$\ddot{\vec{r}}_f = -\mu \frac{\vec{r}_f}{r_f^3} + \vec{u} \tag{5}$$

where $\vec{u}$ is the control thrust vector. By assuming that $\ddot{\vec{r}}_{dl} = 0$, this is forced by control objective.

$$\therefore \ddot{\vec{r}}_{df} = \mu \frac{\vec{r}_f}{r_f^3} - \vec{u} - \mu \frac{\vec{r}_l}{r_l^3} \tag{6}$$

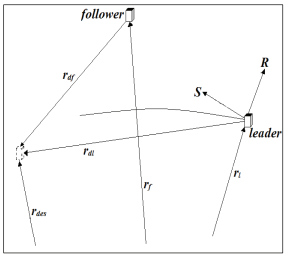


Figure 1 Relative positions of satellite in an orbit

Let x, y, and z be the components of the vector $\ddot{\vec{r}}_{df}$ expressed in the RSW coordinate frame as shown in figure 1. The center of the RSW frame is located at the leader satellite center, where R is a unit vector pointing in direction from Earth center to satellite center, S is a unit vector in the velocity direction normal to R, and W completes the orthonormal set. Then the above dynamic model can be linearized in a similar way to that mentioned in (Inalhan, Tillerson and How 2002) to yield:

$$\ddot{x} = -f_x + 2\omega\dot{y} + \dot{\omega}y + \left(\omega^2 - \frac{2\mu}{r_{tgt}^3}\right)x$$

$$\ddot{y} = -f_y - 2\omega\dot{x} - \dot{\omega}x + \left(\omega^2 + \frac{\mu}{r_{tgt}^3}\right)y \tag{7}$$

$$\ddot{z} = -f_z + \left(\frac{\mu}{r_{tgt}^3}\right)z$$

According to the model given above in equation 7, the z dynamics are decoupled from the orbital plane dynamics and so can be controlled separately. In this deployment only the orbital plane dynamics controlled.

3. MODEL PREDICTIVE CONTROL ALGORITHM

The core of the MPC algorithm is the model of the plant, which can be in the form of a discrete state as follows:

$$x(k+1) = f(x(k), \Delta u(k))$$
$$y(k) = h(x(k), \Delta u(k)) \tag{8}$$

With this model form, the future output response of the plant can be predicted p-step ahead into the future $\hat{y}(k+l)$, where l = 1,2,……,p. The prediction value $\hat{y}(k+l)$ depends on the past actuation and the planned m-step ahead actuation {Δu(k+j), j =1,2, ..., m-1, m < p}. The planned moves {Δu(k+j), j=1,2,....,m-1} are determined as a solution to the following optimization problem.

$$J=\gamma^{y}\sum_{i=1}^{p}(e(k+i/k))^{2}+\gamma^{u}\sum_{i=0}^{m-1}(\Delta u(k+i/k))^{2} \tag{9}$$

Where, it can be noticed that the cost function index J incorporates the errors e(k+i/k) which is the difference between the future reference trajectory r(k+i/k) and the predicted output of the system $\hat{y}(k+i/k)$ equation 10, the change in the actuation moves Δu(k+i/k), and the weighting output γy and input γu.

$$e(k+i/k)=r(k+i/k)-\hat{y}\,(k+i/k) \qquad \textit{subject to} \tag{10}$$

$$x(k+i/k)=f(x(k+i-1/k),\Delta u(k+i-1/k)) \;\; \textit{for } i\geq 0$$

Outside the control horizon m, the actuation moves are constant and their change Δu(k+i/k) = 0. The first element of the minimizing control sequence is implemented on the actual plant. Then the whole cycle of output measurement, prediction, and input trajectory determination is repeated. This procedure is repeated one sampling interval later with a new prediction horizon, control horizon and reference trajectory defined and new output measurement. Because the prediction horizon remains of the same length as for the previous sampling interval, but slides along by one sampling interval at each step, this way of control is called receding horizon strategy; the receding horizon strategy makes a closed loop control law from the original open loop using the actual state and output measurement of the plant under control.

The optimal control sequence depends on the current measurement y(k/k), the prediction horizon p, the control horizon m, and the weights γ^y and γ^u . One of the advantages of the MPC algorithm is its applicability to handle in straightforward way multivariable interactive control problems, and to extend to constrained control problems.

4. LINEAR CONTROL BASED ON LYAPUNOV FUNCTION

For the time variant system (LTV) in equation 7, assume a Laypunov function (Abdelkhalik and Alberts 2004) of the form:

$$V=\frac{1}{2}\left(k_x x^2+k_y y^2\right)+\frac{1}{2}\left(\dot{x}^2+\dot{y}^2\right)$$

$$\therefore \frac{dV}{dt}-\left(k_x x\dot{x}+k_y y\dot{y}\right) \tag{11}$$

Substituting for acceleration $\ddot{x},\ddot{y}$ from equation 7 yields,

$$\frac{dV}{dt}=\left(k_x x\dot{x}+k_y y\dot{y}\right)-\dot{x}f_x-\dot{y}f_y+\dot{\omega}(\dot{x}y-x\dot{y})$$
$$+\omega^2(x\dot{x}+y\dot{y})-\frac{\mu}{r_{tgt}^3}(2x\dot{x}-y\dot{y}) \tag{12}$$

Let the control be as follow:

$$f_x=\left(k_x+\omega^2-\frac{2\mu}{r_{tgt}^3}\right)x+k_{dx}\dot{x}+\dot{\omega}y+2\omega\dot{y}$$

$$f_y=\left(k_y+\omega^2+\frac{\mu}{r_{tgt}^3}\right)y+k_{dy}\dot{y}-\dot{\omega}x-2\omega\dot{x} \tag{13}$$

$$\therefore \frac{dV}{dt}=-k_{dx}\dot{x}^2-k_{dy}\dot{y}^2$$

Which is negative semi-definite, the equilibrium state can be easily checked by setting: $\frac{dV}{dx}=\frac{dV}{d\dot{x}}=\frac{dV}{dy}=\frac{dV}{d\dot{y}}=0.$ This makes $x=0,y=0,\dot{x}=0,\dot{y}=0$.

By applying the controls to the equation of motion 7, the closed loop system is:

$$\ddot{x}=-k_{dx}\dot{x}-k_x x \tag{14}$$

$$\ddot{y}=\quad k_{dy}\dot{y}\quad k_y y$$

5. SIMULATION RESULTS

A simulation tool was developed based on the MPC Toolbox in the MatLab/Simulink environment. First, an open response of the system to error initial conditions was obtained using the linear model of the system based on controller gains corresponding to $\omega_{nx}=\omega_{ny}=0.0005$ rad/second, and $\xi_x=\xi_y=0.65$. I t can be noticed from figure 2 that the time for the positions and acceleration of the system to return back to their zero initial conditions is long enough.

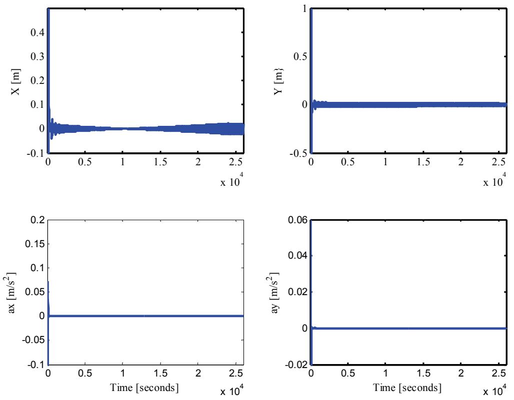


Figure 2 linear system response for errors initial conditions

The objective of this work is to test the ability and cabaplity of the linear controllers to bring the follower to be in distance of 100 m in the x-dirextion and 150 in the y-direction from the leader. An elliptical orbit for the leader was selected with the following parameters (Abdelkhalik and Alberts 2004): semi-major axis is 6.7781e+006 m, eccentricity 0.005 and inclination of 96 □, the follower positon is given as the initial conditions for the x and y positions.

The closed loop response of the system having parameters similar to those implementd in figure 2 shows unstability to bring the system to the desired values. Hence the controller gains were modified to be $\omega_{nx} = \omega_{ny} = 0.003$ rad/second, and $\xi_x = \xi_y = 0.65$. These parameters give a good and fast closed loop response of the system as shown in the following figure 3.

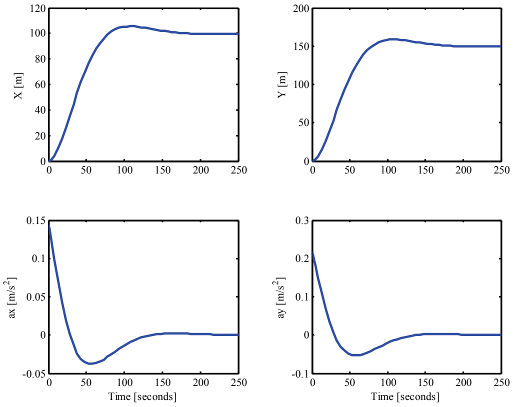


Figure 3 closed loop system response based on Lyapunov function

The model predictive control (MPC) algorithm is applied to the system in the interest to get improved trajectroy tracking. For the purpose of testing the MPC algorithm, the same formation configuration as in the previous controller case is used, which means consideing the model as time inveriant model by making $\omega_{nx} = \omega_{ny} = 0.003$, using a sampling period of 1 to make the model discrete, and do not consider controlling the system in z direction. Moreover, the output weighting matrix for the relative position in x and y direction has been chosen to be $\gamma^y = 2$ and for acceleration in both mentioned direction $\gamma^y = 1$, while the input weighting $\gamma^u = 0.9$. in addition to the previous mentioned configuration and parameters, the prediction horizon has been chosen p = 5 and the control horizon m = 2. The closed loop system response based on the MPC algorithm is shown in figure 4. It can be noticed from figure 4 that the MPC drove the system to the desired x and y positions in short time compared to the controller based on Lyapunov function and maintain zero error in the control interval.

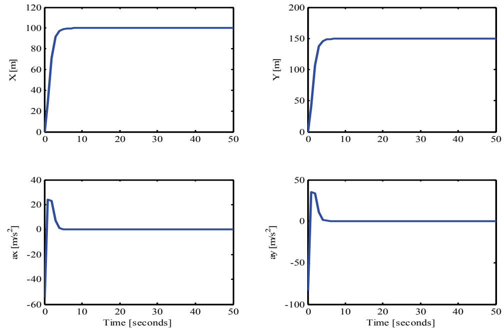


Figure 4 closed loop system response based on MPC algorithm

CONCLUSION

This work demonstrated the feasibility of maintaining the formation conditions in an eccentric orbit in a prescribed interval. Two controllers are evaluated; a linear Lyapunov function type controller and model predictive control algorithm via simulation in the MatLab/Simulink environment. Both controllers show ability to control the formation and correct the initial errors. MPC takes less time to reach the desired positions comparing to the controller based on Lyapunov function, Moreover, the control gains for the controller based on Lyapunov function needs to be tuned by simulation to meet the prescribed behavior, and some gains may lead to instability.

REFERENCES

Abdelkhalik, O. and Alberts, T., 2004. Interval control of formations in eccentric orbits, 14th AAS/AIAA Space Flight Mechanics Conference, Maui, Hawaii, February 8-12, pp. 1 - 12.

AbdulRahman, A., Mustapha Soufian, Mokbel, A. and Majeed Soufian. 2002. Neural network for model based predictive control of a polymerisation reactor, Proceedings of 4th international conference on Recent Advances in Soft Computing (RASC 2002), Nottingham, 12 - 13 December UK , pp. 13 - 19,

Breger, L., How, J. and Richards, A., 2005. Model predictive control of spacecraft formations with sensing noise, Procedings of the American Control Conference, Portland, OR, USA, June 8 - 10, pp. 2385 - 2390.

de Queiroz, M.S., Kapila, V. and Yan, Q., 2000. Adaptive nonlinear control of multiple spacecraft formation flying, Journal of Guidance, Control and Dynamics, Vol. 23,pp. 385 - 390.

Henson, M. A., 1998. Non-linear model predictive control: current status and future directions, Computers and Chemical Engineering, Vol. 23, pp. 187 - 202.

Inalhan, G., Tillerson, M. and How, J.P., 2002. Relative dynamics and control of spacecraft formations in eccentric orbits, Journal of Guidance, Control and Dynamics, Vol. 25, pp.48 - 59.

Kapila, V., Sparks, A,G. and Buffington, J.M., and Yan, Q., 1999. Spacecraft formation flying: Dynamics and Control, American Control Conference, San Diego, CA, pp. 4137-4141.

Lawton, J., 2000. A behavior-based approach to multiple spacecraft formation flying, PhD Dissertation, Faculty of Brigham Young University, Department of Electrical and Computer Engineering.

Manikonda, V., Arambel, P. O., Gopinathan, M., Mehra, R. K. and Hadaegh, F. Y., 1999. A model predictive control-based approach for spacecraft formation keeping and attitude control, Procedings of the American Control Conference, San Diego, California, June, pp. 4258 - 4262.

McInnes, C.R., 1995. Autonomous ring formation for a planar constellation of satellites, Journal of Guidance, Control and Dynamics, Vol. 18,Engineering Notes, pp. 1215 – 1217.

Mesbahi, M. and Hadaegh, F.Y., 2001. Formation flying control of multi spacecraft via graphs, matrix inequalities, and switching, Journal of Guidance, Control and Dynamics, Vol. 24, No. 2, pp. 369 – 379.

Rodrigues, M. A. and Odloak, D., 2000. Output feedback MPC with guaranteed robust stability, Journal of Process Control, Vol. 10, pp. 557 - 572.

Vallado, D., 2001. Fundamentals of astrodynamics and applications, Space Technology Library, 2001.

MODAL ANALYSIS REDUCTION OF MULTI-BODY SYSTEMS WITH GENERIC DAMPING

Loucas S. Louca

Department of Mechanical and Manufacturing Engineering, University of Cyprus

lslouca@ucy.ac.cy

ABSTRACT

Modal analysis of multi-body systems is broadly used to study the behavior and controller design of dynamic systems. In both cases, model reduction that does not degrade accuracy is necessary for the efficient use of these models. Previous work by the author addressed the reduction of modal representations by eliminating entire modes or individual modal elements (inertial, compliant, resistive). In that work, the bond graph formulation was used to model the system and the modal decomposition was limited to systems with proportional damping. The objective of the current work is to develop a new methodology such that model reduction can be implemented to modal analysis of multi-body systems with non-proportional damping that are not modeled using bond graphs. This extension also makes the methodology applicable to realistic systems where the importance of modal coupling terms is quantified and potentially eliminated. The new methodology is demonstrated through an illustrative example.

Keywords: multi-body systems, non-proportional damping, modal analysis, model reduction,

1. INTRODUCTION

Modeling and simulation have yet to achieve wide utilization as commonplace engineering tools. One reason is that current modeling and simulation techniques are inadequate. Specifically, a major disadvantage is that they require sophisticated users who are often not domain experts and thus lack the ability to effectively utilize the model and simulation tools to uncover the important design trade-offs. Another drawback is that models are often large and complicated with many parameters, making the physical interpretation of the model outputs, even by domain experts, difficult. This is particularly true when "unnecessary" features are included in the model.

A variety of algorithms have been developed and implemented to help automate the production of proper models of dynamic systems. Wilson and Stein (1995) developed MODA (Model Order Deduction Algorithm) that deduces the required system model complexity from subsystem models of variable complexity using a frequency-based metric. They also defined proper models as the models with physically meaningful states and parameters that are of necessary but sufficient complexity to meet the engineering and accuracy objectives. Additional work on deduction algorithms for generating proper models in an automated fashion has been reported by Ferris et al. (1998), Ferris and Stein (1995) and Walker et al. (1996). These algorithms have also been implemented and demonstrated in an automated modeling environment (Stein and Louca 1996).

In an attempt to overcome the limitations of the frequency-based metrics, Louca et al. (1997) introduced a new model reduction technique that also generates proper models. This approach uses an energy-based metric (element activity) that in general, can be applied to nonlinear systems (Louca et al. 2010), and considers the importance of all energetic elements (generalized inductance, capacitance and resistance). The contribution of each energy element in the model is ranked according to the activity metric under specific excitation. Elements with small contribution are eliminated in order to produce a reduced model. The activity metric was also used as a basis for even further reduction, through partitioning a model into smaller and decoupled submodels (Rideout et al. 2007).

Beyond the physical-based modeling, modal decomposition is also used to model and analyze continuous and discrete systems (Meirovitch 1967). One of the advantages of modal decomposition is the ability to straightforwardly adjust (i.e., reduce) model complexity since all modes are orthogonal to each other. The reduction of such modal decomposition models is mostly based on frequency, and the user defined frequency range of interest (FROI) determines the frequencies that are important for a specific scenario. In this case, modes with frequencies within the FROI are retained in the reduced model and modes outside this range are eliminated. As expected, mode truncation introduces error in the predictions that can be measured and adjusted based on the accuracy requirements (Li and Gunter 1981; Liu et al. 2000).

The element activity metric provides more flexibility than frequency-based metrics, which address the issue of model complexity by only the frequency content of the model. In contrast, the activity metric considers the energy flow in the system, and therefore, the importance of all energy elements in the model can be described. Previous work by the author addressed the development and reduction of modal representations using the bond graph formulation and using the activity metric (Louca 2006). This work introduced a methodology that reduces the model complexity by eliminating entire modes or partial modes through modal elements (inertial, compliant, resistive). The identification and elimination of insignificant elements was performed with the use of the activity metric and Model Order Reduction Algorithm (MORA). This approach has advantages over frequency-based reduction techniques; however, it has a significant limitation in that it can only be applied to systems with proportional damping and thus is not able to be applied to realistic systems.

The objective of the current work is to develop a methodology that overcomes the limitations of the author's previous work, such that modal analysis and model reduction can be applied to a more general class of systems with non-proportional damping. In addition, the activity metric will be formulated for systems that are modeled with second order ordinary differential equations (ODE), rather than bond graphs and first order ODEs that were used in previous work. Second order ODEs are typically derived from Lagrange's equations or Newton's Law. These two additions will make the activity metric a more appealing model reduction methodology that can be applied to realistic systems.

This paper is organized as follows: first, background about the energy-based metric is provided, along with the reduction algorithm. Next, the equation formulation and modal decomposition of multi-body systems with non-proportional damping is presented. Then, the activity analysis of all modal elements is introduced, along with the closed-form expressions of the steady state activities. An illustrative example of a linear quarter car model is also presented, in order to demonstrate the development of its modal decomposition and the evaluation of the coupling terms' importance using the activity metric. Finally, in the last section, discussion and conclusions are given.

2. BACKGROUND

The original work on the energy-based metric for model reduction is briefly described here since it is the foundation of the contributions in this paper. The main idea behind this model reduction technique is to evaluate the "element activity" of the individual energy elements of a complex system model under a stereotypic set of inputs and initial conditions. The activity of each energy element establishes an importance hierarchy for all elements. Those below a user-defined threshold of acceptable level of activity are eliminated from the model. A reduced model is then generated and a new set of governing differential equations is derived. More details, extensions, and applications of this approach are given in previous publications (Louca and Stein 2002; Louca et al. 2004; Louca and Stein 2009; Louca et al. 2010).

The activity metric has been previously formulated for systems with nonlinearities in both the element constitutive laws and kinematics. In this work, the activity metric is applied to linear systems for which analytical expressions for the activity can be derived, and therefore, avoid the use of numerical time integration that could be cumbersome. The analysis is further simplified if, in addition to the linearity assumption, the system is assumed to have a single sinusoidal excitation, and only the steady state response is examined. These assumptions are motivated from Fourier analysis where an arbitrary function can be decomposed into a series of harmonics. Using this decomposition, the activity analysis can be performed as a function of frequency in order to study the frequency dependency of the energy elements in a dynamic system.

2.1. Element Activity for Linear Systems

The starting point of a model reduction process is a system model, which typically includes complexity that has minimal contribution to the accuracy of the model's dynamic response. Thus, the goal of a reduction algorithm is to identify this "unnecessary" complexity in order to generate a reduced model that is easier to analyze yet accurate. One approach to accomplish this is the activity metric that was previously defined by the author (Louca et al. 2010). The activity metric is outlined below as it was originally developed for models of multi-energy systems that are represented by first order ODEs.

Models of dynamic systems consist of physical energy elements that can store (inertia and stiffness) or dissipate (resistance) energy. In addition, these can be considered as generalized elements in order to have the ability to model multi-energy systems, i.e., translational mechanical, rotational mechanical, electrical and hydraulic. A quantity that can be defined for all energy elements in a model and it is independent of its energy domain, is power. However, power is time dependent and thus not a suitable modeling metric.

A measure of the power response of a dynamic system, which has physical meaning and a simple definition, is used to develop the modeling metric, element activity (or simply "activity"). Element activity, A, is defined for each energy element as:

$$A = \int_0^{\tau} \left| \mathbb{P}(t) \right| \cdot dt \tag{1}$$

where $\mathbb{P}(t)$ is the element power and τ is the time over which the model has to predict the system behavior. The activity has units of energy, representing the amount of energy that flows in and out of the element over the given time τ. The energy that flows in and out of an element is a measure of how active this element is (how much energy passes through it), and consequently the quantity in Eq. (1) is termed activity.

Element power is the product of generalized effort and flow, which for the linear mechanical domain is the product of force and velocity. Given this definition the activity can then be rewritten as:

$$A = \int_0^{\tau} \left| e(t) \cdot f(t) \right| \cdot dt \tag{2}$$

For calculating the activity, the respective efforts and flows must be known throughout the window of interest $[0,\tau]$. Given the system linearity assumption, the efforts and flows can be calculated using linear system analysis theory. The state equations of a linear system, along with the output equations, can be written in matrix form as given below:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\cdot\mathbf{x}+\mathbf{b}\cdot u\\ \mathbf{y} &= \mathbf{C}\cdot\mathbf{x}+\mathbf{d}\cdot u\end{aligned} \tag{3}$$

The above system is assumed to have n states, one input, and k outputs (one for each energy element in the system).

The system has k_I inertial, k_C compliant and k_R resistive elements, thus, $k = k_I + k_C + k_R$. Element power is calculated by using either the effort or flow and the constitutive law, and therefore, only k outputs are required for calculating the activity of all elements in the system. The outputs are selected to be flow, effort, and flow for inertial, compliant, and resistive elements, respectively. The duals of these variables can also be used for calculating element power. Also, the parameters r_I, r_C, r_R are known constants representing the linear constitutive law coefficients of inductance, compliance and resistance, respectively. With the above definitions, the power of each energy element is calculated as:

$$\begin{aligned}&\text{Inertia:} && \mathbb{P}_I = e_I\cdot f_I = \left(r_I\cdot\dot{f}_I\right)\cdot f_I\\ &\text{Compliance:} && \mathbb{P}_C = e_C\cdot f_C = e_C\cdot\left(r_C\cdot\dot{e}_C\right)\\ &\text{Resistance:} && \mathbb{P}_R = e_R\cdot f_R = \left(r_R\cdot f_R\right)\cdot f_R\end{aligned} \tag{4}$$

2.2. Activity Index and MORA

The activity as defined in Eq. (1) is a measure of the absolute importance of an element as it represents the amount of energy that flows through the element over a given time period. In order to obtain a relative measure of the importance, the element activity is compared to a quantity that represents the "overall activity" of the system. This "overall activity" is defined as the sum of all the element activities of the system, is termed total activity (A^{Total}) and is given by:

$$A^{Total} = \sum_{i=1}^{k} A_i \tag{5}$$

where A_i is the activity of the i^{th} element given by Eq. (1). Thus a normalized measure of element importance, called the element activity index or just activity index, is defined as:

$$AI_i = \frac{A_i}{A^{Total}} = \frac{\displaystyle\int_0^{\tau}\left|\mathbb{P}(t)\right|\cdot dt}{\displaystyle\sum_{i=1}^{k}\left\{\int_0^{\tau}\left|\mathbb{P}_i(t)\right|\cdot dt\right\}} \tag{6}$$

The activity index, AI_i, is calculated for each element in the model and it represents the portion of the total system energy that flows through a specific element.

With the activity index defined as a relative metric for addressing element importance, the Model Order Reduction Algorithm (MORA) is constructed. The first step of MORA is to calculate the activity index for each element in the system for a given system excitation and initial conditions. Next, the activity indices are sorted to identify the elements with high activity (most important) and low activity (least important). With the activity indices sorted, the model reduction proceeds given the desired engineering specifications. These specifications are defined by the modeler who then converts them into a threshold of the total activity (e.g., 99%) that he/she wants to include in the reduced model. This threshold defines the borderline between the retained and eliminated model elements. The elimination process is shown in Figure 1 where the sorted activity indices are summed starting from the most important element until the specified threshold is reached. The element which, when included, increments the cumulative activity above the threshold, is the last element to be included in the reduced model. The elements that are above this threshold are removed from the model. The elimination of low activity elements is done by removing these elements from the model description.

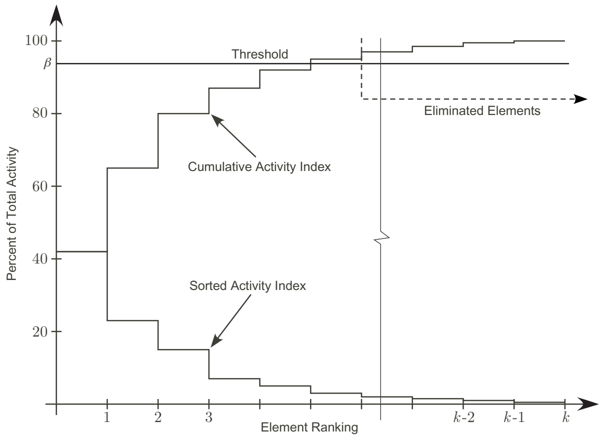


Figure 1: Activity index sorting and elimination

3. MULTI-BODY SYSTEMS

The activity metric as described in the previous section will now be introduced for multi-body systems. The definition of the activity remains the same as stated in Eq. (1), however, the modeling approaches for multi-body systems typically result in second order ODEs. In addition, the generalized effort and flow variables used for defining the activity now become force and velocity given that we have a mechanical system. Modeling of a multi-body system through either Lagranges's equations or Newton's Law generates differential equations in the independent Degrees of Freedom (DOF). For the purposes of the current work, the system is assumed to have no constraints, thus, represented by a set of explicit second order ODEs. In addition, the system is assumed to have generic damping that is not necessarily proportional. For such a system with n DOF the set of differential equations is given by:

$$\mathbf{M}\cdot\ddot{\mathbf{x}}+\mathbf{B}\cdot\dot{\mathbf{x}}+\mathbf{K}\cdot\mathbf{x}=\mathbf{f}\cdot u(t) \tag{7}$$

In the above equation the matrices $\mathbf{M},\mathbf{B},\mathbf{K}\in\Re^{n\times n}$ are constant, square, and non-diagonal and $\mathbf{x}\in\Re^{n}$ is the vector with the physical coordinates or DOF. The system has only one input, $u(t)$, which is exciting the masses through the input vector, $\mathbf{f}\in\Re^{n}$. The input is a continuous function in time. Note that this system is linear given that the matrices are constant.

The above equation can be used to obtain the time response of the system that is required for calculating the activity. However, the objective of the paper is to reduce the modal decomposition model of a multi-body system and not its physical model as given in Eq. (7). Therefore, before continuing with the activity analysis, the modal decomposition of the physical coordinate ODEs must be generated.

This can be done with the modal analysis of the undamped system, i.e., $\mathbf{B}=\mathbf{0}$. The first step is to calculate the natural frequencies and eigenvectors $(\omega_{n,i},\mathbf{v}_i)$, $i=1,\ldots,n$, of the undamped system, through the solution of the eigenvalue problem. The system eigenvectors, through a coordinate transformation, can then be used to generate a set of decoupled ODEs in the modal coordinates, $\mathbf{z}$. The physical and modal coordinates are related with the following transformation:

$$\mathbf{x}=\mathbf{V}\mathbf{z} \tag{8}$$

where $\mathbf{V}=[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n]$ is the orthogonal eigenvector matrix.

In order to generate a set of decoupled equations, the same coordinate transformation of Eq. (8) is applied to the original system with damping in Eq. (7). However, this is only possible when the system has proportional damping, i.e., $\mathbf{B}=\alpha\mathbf{M}+\delta\mathbf{K}$ where $\alpha,\delta\in\Re$. In the general case considered in this work, the equations cannot be decoupled, due to the non-proportional damping, and have the following form:

$$\hat{\mathbf{M}}\cdot\ddot{\mathbf{z}}+\hat{\mathbf{B}}\cdot\dot{\mathbf{z}}+\hat{\mathbf{K}}\cdot\mathbf{z}=\mathbf{V}^{T}\cdot\mathbf{f}\cdot u(t)=\hat{\mathbf{f}}\cdot u(t) \tag{9}$$

In the above equation the modal mass and stiffness matrices are diagonal, however, the modal damping matrix is not diagonal and is given by $\hat{\mathbf{B}}=\mathbf{V}^{T}\cdot\mathbf{B}\cdot\mathbf{V}$. The diagonal elements of the modal mass matrix are equal to one and the diagonal elements of the modal stiffness matrix are the squares of the system natural frequencies. The off-diagonal elements of the modal damping matrix create the coupling between the differential equations in Eq. (9) and their significance will now be evaluated using the activity metric.

3.1. Activity Analysis

The modal decomposition of the physical system produces a set of differential equations that are coupled only through the off-diagonal elements of the modal damping matrix. More specifically, Eq. (9) has the following form:

$$\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1\end{bmatrix}\ddot{\mathbf{z}}+\begin{bmatrix}b_{11}&b_{12}&\cdots&b_{1n}\\b_{21}&b_{22}&\cdots&b_{2n}\\\vdots&\vdots&\ddots&\vdots\\b_{n1}&b_{n2}&\cdots&b_{nn}\end{bmatrix}\dot{\mathbf{z}}$$
$$+\begin{bmatrix}\omega_{n,1}^{2}&0&\cdots&0\\0&\omega_{n,2}^{2}&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\omega_{n,n}^{2}\end{bmatrix}\mathbf{z}=\hat{\mathbf{f}}u(t) \tag{10}$$

The only way to decouple these equations is to eliminate the off-diagonal terms of the modal damping matrix. However, the significance of these terms, to the system dynamic behavior and accuracy, must be first evaluated before eliminated. The key contribution of this work is to use the activity metric to evaluate the importance of these terms to the overall system dynamics.

In the background section the activity was introduced as a modeling metric that can be calculated for physical energy elements. The model in Eq. (10) does not have any physical elements. However, it has elements that exhibit the same characteristics as in the physical domain. The model has modal mass, damping, and stiffness that can be considered as energy elements. Therefore, for each of these modal energy elements, the activity can be calculated in a similar way that is given in Eq. (4).

For the modal mass element the power, which is needed for calculating the activity, is equal to the mass velocity multiplied by the inertial force. Therefore, the power for the i^{th} modal mass is:

$$\mathbb{P}_{M,i}=F_{M,i}\cdot v_i=(\hat{m}_{ii}\cdot\ddot{z}_i)\cdot\dot{z}_i=\ddot{z}_i\cdot\dot{z}_i \tag{11}$$

Similarly, the power of the modal stiffness element is equal to the spring force multiplied by the spring velocity. Therefore, the power for i^{th} modal stiffness is:

$$\mathbb{P}_{K,i}=F_{K,i}\cdot v_i=(\hat{k}_{ii}\cdot z_i)\cdot\dot{z}_i=\omega_{n,i}^{2}\cdot z_i\cdot\dot{z}_i \tag{12}$$

Next, the power of the modal damping terms must be calculated. A possible approach for calculating the power would be to treat the system damping as a single element, which would result in a single activity to consider in the model reduction process. This approach would be easy to implement, however, it is limiting given that it can only assess if the whole damping matrix is important or not according to the activity metric. A different approach, which was previously used for calculating the activity of rigid bodies moving in space, can be considered for more flexibility. For rigid bodies in translation, the inertial forces were evaluated for each DOF and the activity was separately calculated for each DOF (Louca et al. 2004).

A similar and less restraining approach is to treat each term of the modal damping matrix separately and calculate the power and activity for each of these terms. For the i^{th} mode the total damping force acting on the modal mass is given by:

$$F_{B,i} = b_{i1} \cdot \dot{z}_1 + b_{i2} \cdot \dot{z}_2 + \ldots + b_{in} \cdot \dot{z}_n \tag{13}$$

Based on this force the total damping power can be calculated by multiplying the above force with the modal velocity as shown in the equation below:

$$\mathbb{P}_{B,i} = \left(b_{i1} \cdot \dot{z}_2 + b_{i2} \cdot \dot{z}_2 + \ldots + b_{in} \cdot \dot{z}_n\right) \cdot \dot{z}_i \tag{14}$$

This power consists of n terms and each term is considered separately in order to maximize the flexibilty with the elements that can be considered for elimination. Therefore, the j^{th} term of the power of the i^{th} damping force is:

$$\mathbb{P}_{B,ij} = b_{ij} \cdot \dot{z}_j \cdot \dot{z}_i \tag{15}$$

This approach allows the use of the activity to eliminate unimportant terms from the modal damping matrix, which can decouple the modal equations in Eq. (10). In addition, for calculating the above power no additional variables are required since the modal damping power is a function of the modal velocities, which are available after solving the differential equations. Finally, given that the modal damping matrix is symmetric, the power of symmetric terms is equal as it can be seen from Eq. (15), i.e., $\mathbb{P}_{R,ij} = \mathbb{P}_{R,ji}$. Therefore, calculating the activity of the upper off-diagonal terms is sufficient for the analysis, i.e., $i, j = 1, \ldots, n \middle| j \geq i$.

Also the activity of each energy element, the activity of the complete mode can also be calculated. The mode activity is used to access the importance of a mode, and then can be eliminated from the model if it has relatively low activity. The modal power is equal to the modal force (excitation) multiplied by the modal velocity. Thus, the power for the i^{th} mode is:

$$\mathbb{P}_i = F_i \cdot v_i = \left(\hat{f}_i \cdot u(t)\right) \cdot \dot{z}_i \tag{16}$$

3.2. Steady-State Harmonic Activity

To calculate the activity of the elements and terms described in the previous section, the linearity of the system is exploited in order to derive closed-form expressions. In addition, it is assumed that the system excitation, $u(t)$, is a single harmonic with unit amplitude:

$$u(t) = \sin \omega t \tag{17}$$

where $\omega \in \Re$ is the excitation frequency. The steady state response of the modal coordinates in Eq. (9) is calculated using linear system analysis theory. This gives the following closed-form expression for the steady state time response of the i^{th} modal coordinate:

$$z_i\left(t, \omega\right) = Z_i\left(\omega\right) \cdot \sin\left(\omega t + \phi_i\left(\omega\right)\right) \tag{18}$$

where:

$\mathbf{Z}\left(\omega\right) = \left|\mathbf{G}\left(j\omega\right)\right|$, is the steady state amplitude

$\boldsymbol{\Phi}\left(\omega\right) = \sphericalangle\mathbf{G}\left(j\omega\right)$, is the steady state phase shift

$\mathrm{G}\left(s\right) = \left(s^2\hat{\mathbf{M}} + s\hat{\mathbf{B}} + \hat{\mathbf{K}}\right)^{-1} \cdot \hat{\mathbf{f}}$

The activity is calculated by Eq. (1) and the power flow expressions in Eq. (11), (12), and (15), but first the upper bound of the integral must be specified. For this case, the steady state and periodicity of the response are exploited. A periodic function repeats itself every T seconds, and therefore, a single period of this function contains the required information about the response. For this reason, the upper bound of the integral is set to one period of the excitation, $\tau = T = 2\pi/\omega$. Thus, the steady state activity for the i^{th} modal mass is given by:

$$\begin{aligned} A_{M,i}^{ss} &= \int_0^T \left|\mathbb{P}_{M,i}(t)\right| dt = \int_0^T \left|\ddot{z}_i \dot{z}_i\right| dt \\ &= Z_i^2 \omega^3 \int_0^T \left|\sin\left(\omega t + \phi_i\right) \cos\left(\omega t + \phi_i\right)\right| dt \\ &\Rightarrow A_{M,i}^{ss} = 2 Z_i^2\left(\omega\right) \omega^2 \end{aligned} \tag{19}$$

Using the same approach as above the steady state activity of the i^{th} modal stiffness is given by:

$$\begin{aligned} A_{K,i}^{ss} &= \int_0^T \left|\mathbb{P}_{K,i}(t)\right| dt = \int_0^T \left|\omega_{n,i}^2 z_i \dot{z}_i\right| dt \\ &= Z_i^2 \omega_{n,i}^2 \int_0^T \left|\sin\left(\omega t + \phi_i\right) \cos\left(\omega t + \phi_i\right)\right| dt \\ &\Rightarrow A_{K,i}^{ss} = 2 Z_i^2\left(\omega\right) \omega_{n,i}^2 \end{aligned} \tag{20}$$

For the modal damping the activity is calculated for each term of the matrix. Therefore, for the (i, j) term of the matrix the steady state activity is given by:

$$
\begin{aligned}
A^{ss}_{B,ij} &= \int_0^T \left|\mathcal{P}_{B,ij}(t)\right| dt = \int_0^T \left|b_{ij}\dot{z}_j\dot{z}_i\right| dt \\
&= b_{ij} Z_i Z_j \omega^2 \int_0^T \left|\cos\left(\omega t + \phi_i\right)\cos\left(\omega t + \phi_j\right)\right| dt \\
\Rightarrow A^{ss}_{B,ij} &= 2 b_{ij} Z_i Z_j \omega \left(\left(\frac{\pi}{2} - \theta\right)\cos\theta + \sin\theta\right) \\
\text{where} \quad & \theta(\omega) = \left|\phi_i(\omega) - \phi_j(\omega)\right|, \quad \theta \in \left[0, \pi\right]
\end{aligned}
\tag{21}
$$

The activity of an entire mode is calculated the same way. Thus, the activity of the i^{th} mode is given by:

$$
\begin{aligned}
A^{ss}_{i} &= \int_0^T \left|\mathcal{P}_{i}(t)\right| dt = \int_0^T \left|\hat{f}_i \dot{z}_i u(t)\right| dt \\
&= \hat{f}_i Z_i \omega \int_0^T \left|\cos\left(\omega t + \phi_i\right)\sin \omega t\right| dt \\
\Rightarrow A^{ss}_{i} &= 2 \hat{f}_i Z_i \left(\phi_i \sin\phi_i + \cos\phi_i\right) \\
\text{where} \quad & \phi_i \in \left[-\pi/2, \pi/2\right]
\end{aligned}
\tag{22}
$$

The methodology presented in this section allows engineers to evaluate the importance of damping coupling terms of modal decompositions. The modal decompositions are of multi-body system models that have generic damping and are not limited to proportional damping. The calculated element activities in Eq. (19), (20), and (21) can be used in MORA to identify insignificant elements and eliminate them from the model. The possible elimination of coupling terms will generate a set of decoupled differential equations that are easier to analyze and solve. In addition, mode activity in Eq. (22) can be used to evaluate the importance of an entire mode and possibly eliminate modes with low activity. In the next section, an example is used to demonstrate this new methodology.

4. ILLUSTRATIVE EXAMPLE

A quarter car model, which is extensively used in the automotive industry, is chosen to apply the developed methodology. The model consists of the sprung mass, namely, the major mass supported by the suspension, and the unsprung mass, which includes the wheel and axle masses supported by the tire. The suspension is modeled as a spring and a damper in parallel, which are connected to the unsprung mass. The tire is also modeled as a spring and a damper in parallel, which transfer the road force to the unsprung mass (i.e., wheel hub). The tire is assumed to always be in contact with the road. The input to the system is a force, $F(t)$, applied to the sprung mass. Such force can represent dynamic engine loads that are transmitted to the vehicle body at the mounting points. The system is composed of six linear ideal energy elements described by an equal number of parameters. The ideal physical model of the system is depicted in Figure 2 and the parameters representing a medium sized passenger car are given in the Appendix. For the set of parameters in the Appendix the system does not have proportional damping.

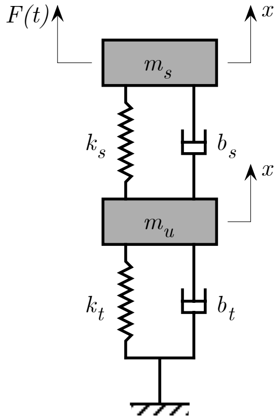


Figure 2: Quarter car model

The system has two DOF ($n = 2$) and the equations of motion are generated using Newton's Law. The DOF are the sprung mass and unsprung mass position, i.e., $\mathbf{x} = \left\{x_s, x_u\right\}^T$. The differential equations for this set of coordinates are:

$$
\begin{bmatrix} m_s & 0 \\ 0 & m_u \end{bmatrix}\ddot{\mathbf{x}} + \begin{bmatrix} b_s & -b_s \\ -b_s & b_s + b_t \end{bmatrix}\dot{\mathbf{x}} + \begin{bmatrix} k_s & -k_s \\ -k_s & k_s + k_t \end{bmatrix}\mathbf{x} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} F\left(t\right)
\tag{23}
$$

The first step of the proposed methodology is to calculate the undamped natural frequencies and eigenvectors. The parameters from the Appendix are substituted in Eq. (23) and the analysis is performed numerically using Matlab. The eigenvalue analysis gives the following two natural frequencies and eigenvectors:

$$
\begin{aligned}
\omega_{n,1} &= 7.996 \text{ rad/s}, \quad \mathbf{v}_1 = \begin{Bmatrix} \text{-61.17} \\ \text{-5.451} \end{Bmatrix} 10^{-3} \\
\omega_{n,2} &= 76.27 \text{ rad/s}, \quad \mathbf{v}_2 = \begin{Bmatrix} \text{-2.018} \\ 165.2 \end{Bmatrix} 10^{-3}
\end{aligned}
\tag{24}
$$

Then, the procedure outlined in Section 3 is used to decouple the equations of motion in Eq. (23). However, the equations cannot be completely decoupled since the system has non-proportional damping. Therefore, the equations as given by Eq. (10), after applying the coordinate transformation given by Eq. (8), become:

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\ddot{\mathbf{z}} + \begin{bmatrix} 1.06 & 2.99 \\ 2.99 & 14.97 \end{bmatrix}\dot{\mathbf{z}} + \begin{bmatrix} 63.94 & 0 \\ 0 & 5816.6 \end{bmatrix}\mathbf{z} = \begin{Bmatrix} 0.0612 \\ 0.0020 \end{Bmatrix} F\left(t\right)
\tag{25}
$$

The equations in the modal coordinates are coupled through the off-diagonal terms of the modal damping matrix. As expected, the mass and stiffness modal matrices are diagonal. The activity metric will be used next to evaluate the importance of a complete mode or individual modal elements.

The steady state activity is calculated over the range of frequencies $10^{-2}-10^{4}$ rad/s that are well below the low and well above the high natural frequencies. The steady state response of the modal coordinates is calculated using Eq. (18) for this frequency range and it is shown in Figure 3. Note that the second mode has two peaks since the modal coordinates are not decoupled. The steady state modal amplitude and phase shift will be used for calculating the activities.

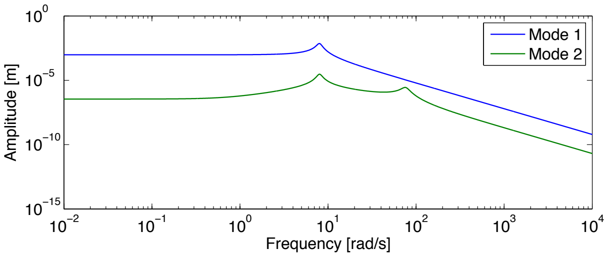


Figure 3: Modal coordinate frequency response

First, the activity of the complete modes is calculated using Eq. (22). The frequency response of the two modal activity indices is given in Figure 4. Mode 1 has the highest activity over the whole frequency with an activity index of almost one. The second mode has very low activity index ($\sim 10^{-5}$) at low frequencies, which varies until it reaches a higher value ($\sim 10^{-3}$) at high frequencies. These activity indices can be used to generate a reduced model, which in this case, could be a model that includes either mode 1, or mode 2, or mode 1 and 2.

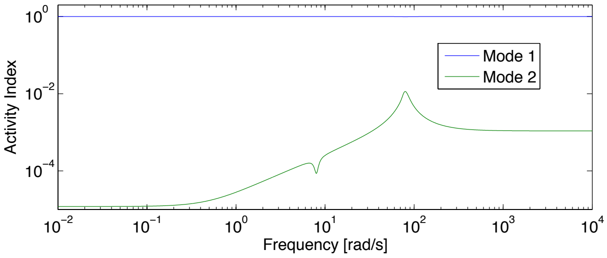


Figure 4: Modal activity index

The reduction is performed using MORA and a reduction threshold is set to 99.95%. Given this threshold, at low frequencies, a reduced model that includes only mode 1 is necessary. This reduced model is sufficient to accurately predict the response up to an excitation frequency of 17.86 rad/s. After this frequency both modes are needed in the model according to MORA. In summary, MORA generates one reduced model that includes only mode 1, which is valid for low frequencies. The second model incudes both modes (no reduction) and it is valid for higher excitation frequencies.

To validate this result, a reduced model with just mode 1 is generated and compared to the full model. The comparison is made between the steady state amplitude of the two DOF as shown in Figure 5. Note that the top plot of this figure has the responses of both the full and reduced model that overlay one another. The reduced model accurately predicts the response of both system coordinates for low input frequencies and up to the frequency calculated by MORA (17.86 rad/s). After this frequency the reduced model still accurately predicts the amplitude of the sprung mass position, x_s, however, it does a poor job in predicting the amplitude of the unsprung mass position, x_u. The reduced model predicts the amplitude of the sprung mass position with a maximum error of 0.75% over the whole range of excitation frequencies.

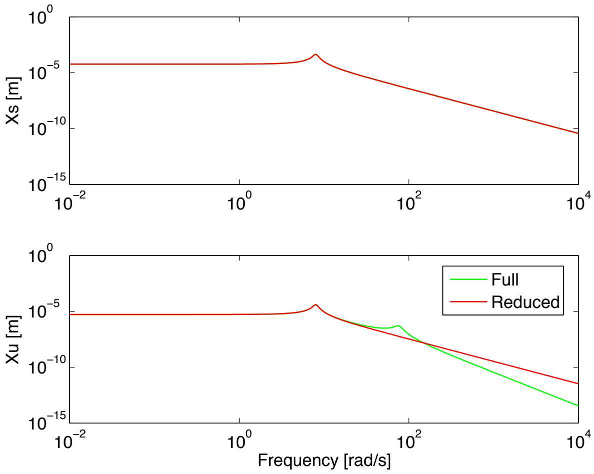


Figure 5: Reduced model comparison

Next, the activities of the individual modal elements are considered, as introduced in Section 3.1, which allow a "finer" activity analysis and model reduction. The steady state activities are calculated using Eq. (19), (20), and (21). Then the activity indices are calculated and their frequency dependancy is depicted in Figure 6. At low frequencies the most active elements are the modal stiffness and diagonal modal damping (b_{11}) of the first mode. These elements handle the bulk of the energy flow induced into the system by the input. As the frequency increases the activity of all elements increases, and around the first resonance the four modal elements of the first mode (m_1, k_1, b_{11}, b_{12}) are the most active. At this frequency, the second mode elements have lower activity since most of the energy flow is handled by the first mode. Around the second resonance the elements of the first mode are still the most active, however, the activity of the second mode elements have a peak. As the frequency continues to increase, the activity of elements drops except for the two modal masses for which the activity reaches a steady state value. These results agree with the previous activity analysis for the entire modes since at least one element from the first mode has the highest activity at all frequencies. At low frequencies the first modal stiffness is the dominant element, where at high frequencies the first modal mass has the highest activity.

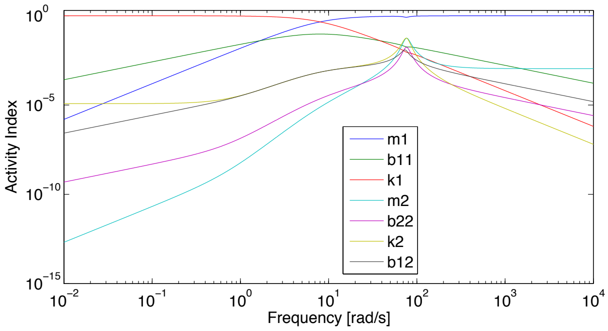


Figure 6: Element activity index

MORA is then applied to generate a series of reduced models. Again, based on a 99.95% threshold, the model complexity is determined as a function of the engine load frequency. The first plot in Figure 7 shows the range of frequencies over which each element is included in the reduced model according to MORA. The y-axis represents each element, where the thick line defines the range of frequencies, over which the element should be included, while no line implies the range of frequencies over which the element is eliminated. The second plot shows the number of elements included in the reduced model.

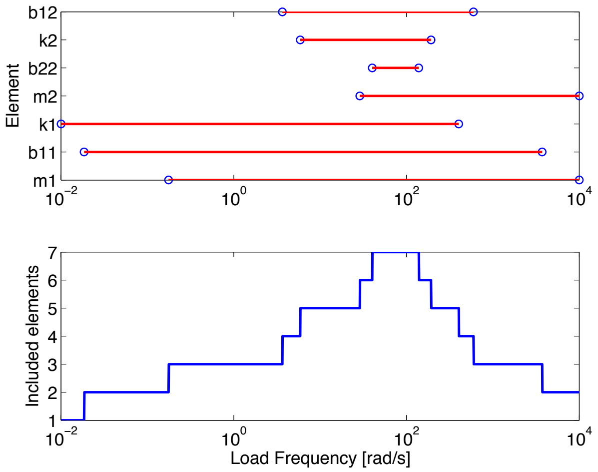


Figure 7: Included modal elements, Threshold = 99.95%

As shown in Figure 7, a low frequency excitation requires only the modal stiffness of the first mode. This is in agreement with theory given that this is a static case (low frequency), where there is no motion and stiffness characteristics dominate the response. Also, the modal stiffness of the second mode is not important since it is about two orders of magnitude stiffer than the first mode and its deflection is insignificant.

As the frequency increases, the first diagonal modal damping is included and then the first modal mass must also be added. The next element to be included, before the first natural frequency at 3.6 rad/s, is the off-diagonal element of the modal damping. Right before the first natural frequency at 5.9 rad/s, the second modal stiffness is added. The same model is valid until the second resonance where gradually all elements are included.

For higher frequencies above the second natural frequency the activity of some elements decreases and they are eliminated from the model. For high frequency input the modal masses have the highest activities and they are the only elements included in the reduced model. The individual element activity analysis agrees with the analysis performed for the entire mode in Figure 4, however, more reduction is accomplished due to the finer calculation of the activity distribution. For example, at low frequencies instead of including the entire first mode only the modal stiffness is needed. Similarly, at high frequencies only the two modal masses are included instead of the entire modes.

The above analysis generates ten reduced models that are necessary to accurately predict the system's response over the whole frequency range. Each of these models is unique and includes different modal elements over specific frequency ranges. All these models are not generated but one reduced model is developed and compared with the full model.

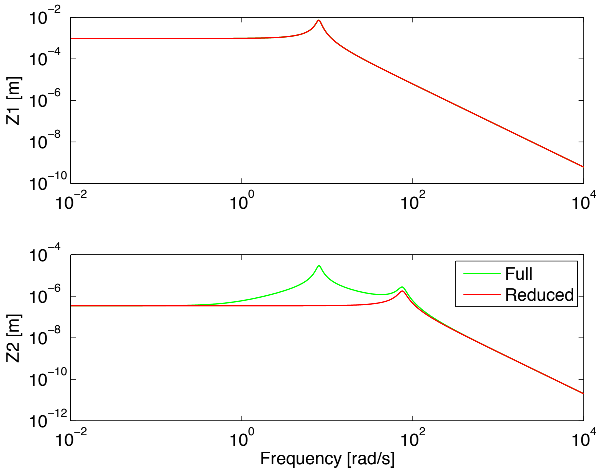


Figure 8: Mode comparison: Full vs decoupled

The key objective of this work is to evaluate the importance of the coupling (off-diagonal) terms of the modal damping matrix as it was shown in Figure 7. Therefore, a reduced model without the off-diagonal damping terms is generated, which consists of two decoupled under-damped modes. This model is valid for low frequencies up to 3.6 rad/s and from 597 rad/s up to high frequencies. The reduced model over these ranges of frequencies according to MORA has more elements removed, however, for simplicity only the off-diagonal damping terms are eliminated. The comparison of the steady state amplitude between the reduced and full model is shown in Figure 8. The reduced model is accurately predicting the steady state amplitude of both modes over the whole frequency range, except over the frequencies (3.6 - 597 rad/s) where the off-diagonal damping term has high activity and should be included in the model.

5. DISCUSSION AND CONCLUSIONS

A new methodology for decoupling modal decompositions of multi-body systems is developed in this paper. The methodology uses the activity metric to quantify the importance of coupling terms, which can then be eliminated using MORA if they fall below the user-specified threshold. The activity metric can also be used to identify unimportant elements within a mode and achieve even further reduction in the model size and complexity. The new methodology overcomes the limitation of a previously developed procedure and gives engineers the ability to reduce modal decompositions of real systems with non-proportional damping.

The importance of coupling terms is measured using the activity metric. This enables the elimination of these terms when their activity index is low compared to the other elements in the system. This is the main contribution of this work and it was shown that the new methodology generates simple, more computationally efficient, yet accurate models. In addition, the decoupled equations provide more flexibility when they are used for analysis or design purposes.

It is shown that when considering the sinusoidal steady state response, the derivation of analytical expressions for the activity as a function of the input frequency is possible. It is also shown that the activity varies with the frequency content of the excitation. This is expected as different input frequencies excite different system dynamics; nevertheless, the activity metric quantifies the contribution of each energy element to the system dynamic response.

The developed methodology is valid for a single harmonic input, however this is not a limitation since generic inputs can also be addressed. For generic inputs, the reduced model can be generated using a Fourier expansion of the input into a series of harmonics. Using this decomposition, the activity analysis is first carried out for a single harmonic excitation under steady state conditions. Then, since the system is linear, the effects of each harmonic are superposed to get the aggregate response caused by this generic input. The same approach can be used for controller design in order to use reduced models that are accurate given the controller and disturbance bandwidth.

The importance of individual modal elements, in contrast with a complete mode, can be addressed and lead to a significant reduction in model size. The benefits from this reduction are twofold. First the reduced size of the model should result in an equivalent reduction in computational cost. Second, the identification of low activity, and thus unimportant elements, reduces the development time since for these elements less accurate parameters can be used. In the example provided in this work the parameters were calculated and acquiring these parameters was inexpensive, however, in other cases modal parameters are measured experimentally, which can be an expensive process.

The results of this paper provide more insight into the nature of the reduced ordered models produced by MORA, and therefore, demonstrate that MORA is an even more useful tool than previously realized for the production of proper models of nonlinear systems. Also, a new methodology for addressing the importance of modes in modal decomposition has been developed. This methodology can handle complete modes, individual modal elements, as well as modal damping coupling terms.

APPENDIX

Quarter Car Parameters

Sprung Mass	m_s	=	267 kg
Suspension Stiffness	k_s	=	18,742 *N/m*
Suspension Damping	b_s	=	340 N·s/m
Unsprung Mass	m_u	=	36.6 kg
Tire Stiffness	k_t	=	193,915 N/m
Tire Damping	b_t	=	200 N·s/m

REFERENCES

Ferris, J.B., J.L. Stein, and M.M. Bernitsas, 1998. "Development of Proper Models of Hybrid Systems." *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Vol. 120, No. 3, pp. 328-333. Published by ASME, New York, NY.

Ferris, J.B. and J.L. Stein, 1995. "Development of Proper Models of Hybrid Systems: A Bond Graph Formulation." *Proceedings of the 1995 International Conference on Bond Graph Modeling*, pp. 43-48, January, Las Vegas, NV. Published by SCS, ISBN 1-56555-037-4, San Diego, CA.

Li, D.F. and E.J. Gunter, 1981. "Study of the Modal Truncation Error in the Component Mode Analysis of a Dual-Rotor." *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power*, Vol. 104, No. 3, p 525-532. Published by ASME, ISSN 0022-0825, New York, NY.

Liu, Dauh-Churn, Chung, Hsing-Liao and Chang, Wei-Min, 2000. "Errors Caused by Modal Truncation in Structure Dynamic Analysis." *Proceedings of the International Modal Analysis Conference - IMAC*, v 2, p 1455-1460. Published by Society for Experimental Mechanics Inc., ISSN 1046-6770, Bethel, CT.

Louca, L.S., J.L. Stein, G.M. Hulbert, and J.K. Sprague, 1997. "Proper Model Generation: An Energy-Based Methodology." *Proceedings of the 1997 International Conference on Bond Graph Modeling*, pp. 44-49, Phoenix, AZ. Published by SCS, ISBN 1-56555-103-6, San Diego, CA.

Louca, L.S. and J.L. Stein, 2002. "Ideal Physical Element Representation from Reduced Bond Graphs". *Journal of Systems and Control Engineering*, Vol. 216, No. 1, pp. 73-83. Published by the Professional Engineering Publishing, ISSN 0959-6518, Suffolk, United Kingdom.

Louca, L.S., D.G. Rideout, J.L. Stein, and G.M. Hulbert, 2004. "Generating proper dynamic models for truck mobility and handling". *International Journal of Heavy Vehicle Systems (Special Issue on Advances in Ground Vehicle Simulation),* Vol. 11, No. 3/4 pp. 209-236. Published by Inderscience Enterprises Ltd., ISSN 1744–232X, St. Helier, United Kingdom.

Louca, L.S., 2006. Bond Graph Based Modal Representations and Model Reduction of Lumped Parameter Systems. *Proceedings of the 20th European Conference on Modeling and Simulation*, pp. 196-203, Bonn, Sankt Augustin, Germany. Published by ECMS, ISBN 0-9553018-0-7.

Louca, L.S. and J.L Stein, 2009. "Energy-Based Model Reduction of Linear Systems". Proceedings of the *6th International Symposium on Mathematical Modeling*, Vienna, Austria. Published in the series ARGESIM-Reports no. 35, ISBN 978-3-901608-35-3, Vienna, Austria.

Louca, L.S., J.L. Stein, and G.M. Hulbert, 2010. "Energy-Based Model Reduction Methodology for Automated Modeling". *Journal of Dynamic Systems Measurement and Control,* Vol. 132, No. 6, 061202 (16 pages). Published by the American Society of Mechanical Engineers, ISSN Print 0022-0434, ISSN Online 1528-9028, New York, NY.

Meirovitch, L., 1967. *Analytical Methods in Vibrations*. Macmillan Publishing Inc., New York, NY

Rideout, D.G., J.L. Stein, and L.S. Louca, 2007. "Systematic Identification of Decoupling in Dynamic System Models". *Journal of Dynamic Systems Measurement and Control,* Vol. 129, No. 4, pp. 503 – 513. Published by the American Society of Mechanical Engineers, ISSN Print 0022-0434, ISSN Online 1528-9028, New York, NY.

Stein, J.L. and L.S. Louca, 1996. "A Template-Based Modeling Approach for System Design: Theory and Implementation." *TRANSACTIONS of the Society for Computer Simulation International*. Published by SCS, ISSN 0740-6797/96, San Diego, CA.

Walker, D.G., J.L. Stein, and A.G. Ulsoy, 1996. An Input-Output Criterion for Linear Model Deduction. *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Vol. 122, No. 3, pp. 507-513. Published by ASME, New York, NY.

Wilson, B.H. and J.L. Stein, 1995. "An Algorithm for Obtaining Proper Models of Distributed and Discrete Systems." *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Vol. 117, No. 4, pp. 534-540. Published by ASME, New York, NY.

AUTHOR BIOGRAPHY

Loucas S. Louca received his Diploma in Mechanical Engineering from the National Technical University of Athens, Greece, in 1992. He then moved to the University of Michigan where he received his M.S.E. in 1994 and Ph.D. in 1998, both in Mechanical Engineering. During his graduate studies at the University of Michigan he received a Fulbright scholarship (Cyprus-America Scholarship Program). He continued to work in the Mechanical Engineering department at the University of Michigan as a Research Fellow until 2000 when he joined the research faculty of the Mechanical Engineering department as an Assistant Research Scientist. He was contacting research and advising students in the area of intelligent vehicle system dynamics and control within the Automotive Research Center at the University of Michigan. He joined the faculty of the University of Cyprus in January 2004 and he is currently and Assistant Professor.

His research interests lie in the areas of system dynamics and control, bond graph theory, physical system modeling and model reduction of large scale systems, modeling of automotive systems, multi-body dynamics, computer aided modeling and simulation, and haptic interfaces and rehabilitation. He is the author of CAMBAS (Computer Aided Model Building Automation System), an automated modeling software that enables the rapid development of efficient models for linear systems and its used for the teaching of courses in modeling of dynamic systems.

He is an active member of the bond graph research community and organizes focused sessions at modeling related conferences. He is also a member of the American Society of Mechanical Engineers (ASME) and the vice-chairman of the Modeling and Identification technical panel of the Dynamic Systems and Control Division. He is also a member of the Institute of Electrical and Electronics Engineers (IEEE), Society for Modeling & Simulation International (SCS), and Society of Automotive Engineers (SAE).

REVIEW OF "FAULT DETECTION, DIAGNOSIS AND DECISION SUPPORT METHODS" IN INDUSTRY

Rambabu Kandepu[(a)], Tor Inge Waag[(b)], Karl Magnus Laundal[(c)], Anis Yazidi[(d)]

(a), (b),(c), (d) Teknova AS, Norway

(a)rka@teknova.no, (b)tiw@teknova.no, (c)kml@teknova.no (d)aya@teknova.no

ABSTRACT

This paper focuses on some specific industrial cases in the area of condition monitoring, fault detection, diagnosis and decision support system methods. The paper briefly describes the cases, describes the existing methods, if any, and lists the specific challenges.

Keywords: Condition monitoring of industrial equipment, Review of industrial methods.

1. INTRODUCTION

Fault detection and diagnosis has been an active and important field for petrochemical plants. It has not been given the same level of attention in other process industries; especially metallurgical and material processing plants. But as the competition is growing in the world market, the fault detection and diagnosis is becoming important in all types of industries to reduce unexpected situations and to avoid unplanned shutdowns. The other important parameter for the motivation in this direction is health and safety issues arising from unexpected faults.

The work presented here is a part of an initiative taken by Teknova to work together with local industries in southern Norway with a focus on methods and tools for fault detection, diagnosis and decision support systems. The first step in this initiative is to survey the existing methods and tools in practice. The survey was focused to categorize the methods in three categories: measurement methods, fault detection & diagnosis (condition monitoring) and decision support system. The survey is intended to look at the methods applied for identifying faults both in process and operating equipment. However, most of the work presented here is focused on the equipment aspects.

In order to have a fault detection & diagnosis (or condition monitoring system) it is very important to have enough measurements from a process or equipment. However it is not always possible to measure key variables/parameters to monitor a process or equipment well enough. Especially the process plants being surveyed were metallurgical plants, where the operating temperatures for some process operations exceed 2000°C. Furthermore, the corrosive environment, dangerous fluids/gases and closed systems make it even more difficult to measure the key variables. For some types of equipment, e.g. pipes handling different types of fluids, it may be cumbersome to have manual periodic monitoring equipment. There can be situations where it is better to have contact-free sensors with regard to health and safety issues. Even though there are lots of advancements in various measuring technologies, the application of these advanced technologies for metallurgical plants has to overcome a lot of challenges.

In some cases, it is almost impossible to apply any technology to measure key process variables or performance parameters of equipment. In these situations, modeling plays an important role to develop soft sensors to estimate the immeasurable variables. In traditional process industry, where application of models for control and state estimation purposes has not been a practice, it is a challenge to make the transition. Especially, in situations where mathematical models cannot be applied and knowledge based and/or data-based models are useful, it is even harder to make engineers interested in developing and implementing these kinds of models.

For some cases, there are measurements available, but it is a challenge to have dedicated human resources to use the measurements for the purpose of condition monitoring of the equipment. It appears very common that there is a lot of historical data of the process measurements which is not being used for any purpose. The field of industrial diagnosis is not being applied at the same level in different process plants. The reasons can be many; the size of the plant, the age of the plant and the experience from before, availability of human and economic resources etc. Even in the case where there is an interest for using the data to gain useful knowledge, it is a challenge to convince other engineers/operators to have a transition from more human control to less human control in the plant.

Decision support system plays an important role once a fault is detected. For the plants operated by operators, it is very important to have an optimal design of decision support systems; design of Human Machine Interface (HMI) screens, placement of screens, color design, alarm level design and display etc. For some critical equipment failure and unexpected process operating conditions, the decision support system plays a key role for the operators and engineers to make important decisions in a short time interval.

1.1. Overview of the state of the art

At this juncture, we shall give a brief overview of the most prominent state of the art techniques used in the field fault detection and diagnostics. In one hand, fault detection and diagnosis methods fall into two main families, namely, Quantitative methods

(Venkatasubramanian et al 2003a) and Qualitative methods (Venkatasubramanian et al 2003b). In the other hand, other taxonomies divide the later techniques into History based approaches and Model based approaches (Venkatasubramanian et al 2003c). The cross-combination of these aforementioned taxonomies results into four classes illustrated in Figure 1.

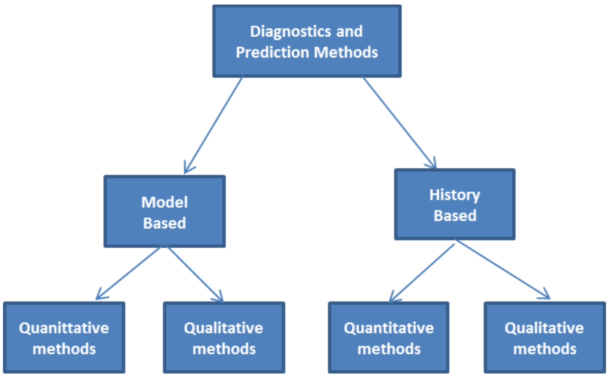


Figure 1: Classification of failure detection and prediction techniques.

Failure isolation refers to the ability of the diagnostic methodology to discriminate between the different types of failures and trace their root causes (Dash and Venkatasubramanian 2000). Fault isolation seeks to identify a cause-effect relationship that is susceptible of explaining the reason behind system deviation. In this perspective, Fault trees (Lapp and Powers 1977) have been extensively used for failure isolation. They are based on backward reasoning where the symptoms of the failure are first identified and then traced back using backward chaining to a possible root cause that can justify the process's mal-functioning.

Bayesian networks are a popular technique for fault diagnosis which bears similarities with the fault tees (Przytula and Thompson 2000). Statistical knowledge of the process and failure occurrences are summarized in a graph which edges represent cause-effect relationships. In the absence of statistical knowledge, the graph can be constructed by knowledge experts that specify the conditional probabilities of the occurrence of a system event given observations (symptoms) in a form of tables attached to each node. Bayesian networks have shown great permit in complex systems where human expert attention might easily become overwhelmed by the overabundant possible root causes of a failure. Usually, the hypothesis that dominates the other hypothesizes in terms of probability is identified as the root cause of the failure. Nevertheless, a sorted list of the eventual root causes of the failure ordered by their respective probabilities given the symptoms can assist the human expert in the presence of uncertainties. Case Based Reasoning (Guiu et al 1999) is a widely used technique for identifying faulty situations by leveraging historical knowledge of the process. In Case Based Reasoning, the status of the monitored process is compared to previously encountered failure situations from historical data accumulated over experience. A distance similarity measure is employed in order to situate the current status vis-à-vis previously encountered failures. The later distance is usually computed in form of a weighted combination of the difference between the current status of the monitored process variables and the historical process data. Nevertheless, the main shortcoming of Case based reasoning is its inability to predict and recognize novel classes of failures that are not present in the historical data. It is worth mentioning that Bayesian Networks and Case based reasoning belong to the class of qualitative based methods. Knowledge based Expert Systems (Becraft and Lee 1993) are another widely used method for failure diagnostics and prediction. Knowledge based Expert Systems are constructed via collaboration between a knowledge engineer and domain experts. The simplest form of knowledge based expert system is defined by antecedent part and a consequence part which maps these system observations to a fault mode. These rules reflect the expert knowledge experience about the system. There are three main challenges when constructing such expert systems: the issue of completeness of the set of rules to cover all states of the system, the eventual huge number of rules and the possibility of creating conflicting rules. Neural networks (Hoskins and Himmelblau 1998) are a form of quantitative methods that are based on the history of the process. Neural Networks are particularly useful in the absence of a model of a physical model of the process being monitored. Neural networks are able to extract hidden knowledge from process data by deducing a mathematical mapping between the inputs of the process and the classes of failures. When it comes to failure diagnostics and prediction, the Neural Network is fed by training data from the different failure modes as well as from the normal operations modes. The Neural network consists of different layers of neurons where the input of one layer serves as output to the next layer. Back-propagation is probably the most successful technique for updating the weights of the neurons in a Neural Network. In Back-Propagation update mode, the weights are adjusted using a gradient approach in order to mitigate the error between the computed output of the neural network and the expected output from the labeled training data. The main shortcomings of Neural Networks are twofold. First, Neural networks give good performance in already known situations while usually fail to predict novel failures for which there is no training data. In addition, it is not possible to express and extract the learned rules of the neural networks in a human readable form.

Machine Learning techniques have been also widely deployed in the area failure diagnosis and prediction. The adopted Machine Learning techniques can be broadly divided into two main classes, namely Regression techniques and Classification techniques. Regression paradigms aim to deduce mathematical expressions that fit the training data. Principal Component Analysis /Partial Least Squares (Wold et al

1984) have received a lot of attention when it comes to creating a statistical model of the failures occurrences by reducing the dimension of the space containing the variable of the monitored process and then applying multivariate regression techniques in order to relate the process input variables to the output variables. Gaussian Processes (Rasmussen and Williams 2006) have recently emerged as intriguing technique for non-linear regression that started to gain popularity. Pattern recognition classifiers have also been used in the field. These techniques include Support Vector Machines (Pöyhönen et al 2002) and Decision trees. In some cases, the different classes of faults cannot be separated in a low dimensional space by a linear classifier; therefore Kernel based techniques are used to map the low dimensional space into a high dimensional space where the faults classes are separable. Other statistical methods include Clustering and Data Mining techniques. Clustering and Data Mining techniques are two forms of quantitative based techniques that have found many applications in the field of failure diagnosis. It is worth mentioning that clustering is a form of unsupervised learning in contrast to classification techniques such as Support vector machines and decision trees that are forms of supervised learning. In Data Mining, association mining rules is a known method (Agrawal and Srikant 1994) that permits to discover frequent episodes in event sequences which are able to predict the failure in advance.

Model based techniques are based on deep knowledge of the mathematical model that governs the monitored process. Kalman filters (Frank et al 2000) are probably the most popular model based techniques for failure prediction and diagnostics. The advantage of Kalman filter resides in their recursive update form which makes it computationally and memory efficient. In fact, Kalman filter relies on the last measurement in order to create an estimate of the state of the system and does not consequently require storing the whole historical data of the process.

Residual based methods (Gertler and Monajemy 1995) resort to the concepts of residual generation which represents the difference between various functions of the outputs and the expected values of these functions under normal (no-fault) conditions. The procedures for residual generation vary from hardware redundancy (voting schemes) to complex state and parameter estimation methods.

The following industrial cases are presented in the sections to follow; Condition monitoring of pipes, Condition monitoring of valves, Condition monitoring of heat exchanger tubes, Water leakage detection & control and Failure prediction of hydraulic systems.

2. CONDITION MONITORING OF PIPES

One of the industrial cases reported in this paper is about condition monitoring of pipes belonging to a plant in the process industry. The objective is twofold: to perform cost-effective condition based monitoring, and to fulfill government imposed rules about health, safety and environment.

In the case studied, the company has a large number of pipes in use, with an accumulated length of many kilometers. For pipes carrying hazardous substances, condition monitoring is mandatory. The owner of the plant is responsible for a safe operation, and has to come up with adequate procedures. This chapter describes an effort to meet these requirements. The majority of the pipes are thin-walled, most of them with 3 to 6 mm wall thickness.

The methods considered, in addition to visual inspection, are ultrasonic wall thickness, x-ray, and pressure testing. For the critical pipes at this plant, acceptance criteria for wall thickness loss after ultrasonic inspection are defined as follows:

- 0-5% is OK within measurement uncertainty.
- 5–10% Increasing attention during subsequent inspections.
- >10 % Physical check-out, disassembly, X-ray inspection, replacement.

A plan for measurement points has been made, limiting the total number of check points to 1000, taking into account the consequences of leakage, and experience with location of problem spots. Locations at the pipes with highest flow velocities have been chosen, such as the outside of pipe bends, and especially bends immediately downstream from pumps, see Figure 2.

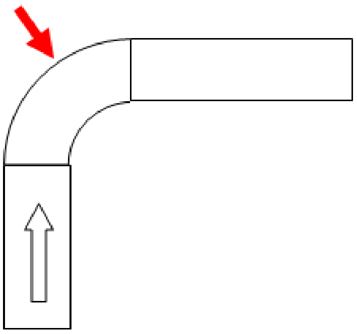


Figure 2: Location of inspection points at the outside of a bend.

Also, points for inspection are selected downstream from reductions in cross-section, near the first weld, where the flow velocity is increased, see Figure 3.

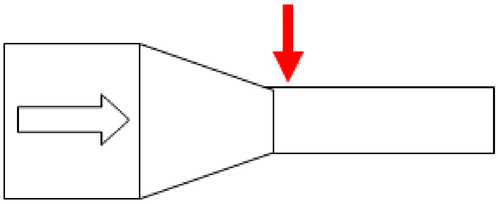


Figure 3: Location of inspection points after reduction in cross-section.

With completely fluid-filled cross-sections, points are selected at the top, bottom and on the sides (at the 3, 6, and 12 o'clock positions), see Figure 4.

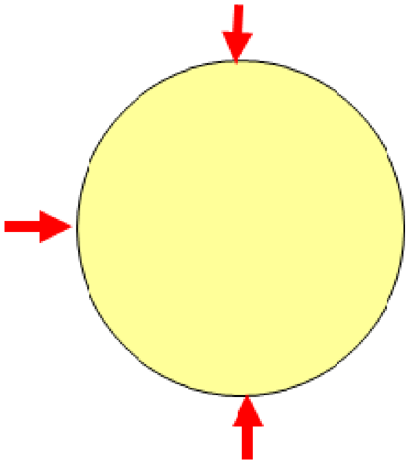


Figure 4: Location of inspection points for completely fluid filled pipes.

With partly filled pipes, inspection is made at the bottom and at the normal liquid level, see Figure 5.

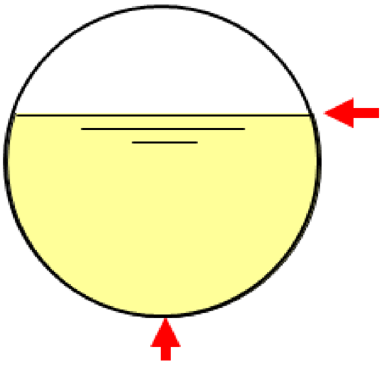


Figure 5: Location of inspection points for partially fluid filled pipes.

At water locks, several points along the axis near the bottom are chosen, and with bends in the vertical plane trapping air, the first point in the flow direction near the liquid to air contact is chosen, see Figure 6.

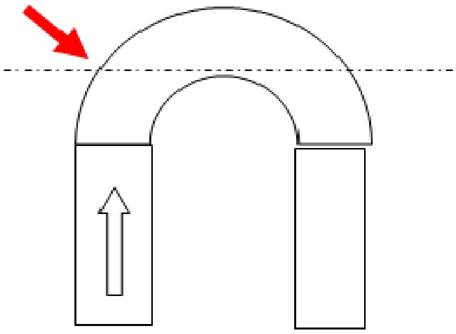


Figure 6: Location of inspection points at air traps.

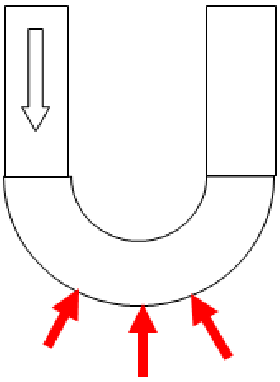


Figure 7: Location of inspection points at water locks.

These points are marked, as shown in Figure 7, after inspection in order to compare subsequent measurements and make trend plots for close checking of material loss in the 5 – 10% range. Even an inspection plan of 1000 points is a formidable undertaking. Results shall be archived, and the measurement locations should be marked very precisely for subsequent inspection and comparison.

Pressure testing should be avoided, since it could lead to leakage of hazardous material and exposure of humans. In cases with harmless fluids it can be used if care is taken to protect the operators well.

2.1. Ultrasonic wall thickness measurement

For the wall thickness measurements, the primary method used is recording of travel times for pulse-echo ultrasonics (NDT Handbook). This means that a short pulse train of ultrasonic energy is generated and transmitted from a piezoelectric transducer, and the instrument switches to listening mode after a short dead time. An echo of the transmitted pulse is then detected when the input level rises above a manually or automatically set threshold, or the detection can be based upon the time difference between multiple echoes travelling back and forth between the front and back wall several times.

The length of the pulse train and the minimum dead time determines the minimum wall thickness that can be measured. The time difference between multiple echoes gives a better estimate of wall thickness than the time to the first echo, provided that the subsequent echoes are strong enough.

The above case, with wall thickness 3 to 6 mm, and a goal of detecting a 5% material loss, requires fairly high-frequent and wide-band ultrasonic signals. Ultrasonic transducers should be selected for having well damped, short pulse trains. The center frequency should preferably be above 5 MHz with a pulse signature consisting of less than 2.5 oscillations before the signal is reduced to below 20% of the maximum amplitude. This corresponds to a 0.5 microsecond pulse length, which with good margin ensures a minimum detectable wall thickness of 1.5 mm, given an approximate sound velocity of 6 mm per microsecond.

Other methods, such as autocorrelation and inversion methods can also be used to find the desired time differences, and may give more accurate results than estimating ultrasonic travel times from simple thresholding of the received pulse amplitudes. These methods require more skilled signal analysis, but will give more precise results, especially with repeated surveys after a certain time interval, e.g. 6 or 12 months. To gain the most from repeated measurements, the same probe should then be located at exactly the same position as the first measurement, not only close to a point marked by a pencil, but preferably guided by firm side supports glued to the pipe.

For location of cracks or bad welds, there are other ways to configure ultrasonic inspection, using two or more transducers, or using non-normal angles of incidence. One of the principles used is then to direct the ultrasound away from the probe position, and if the material to be inspected is homogeneous and free from cracks, nothing will be reflected or scattered back to the probe. Reflected energy detected within a certain time window will therefore indicate defects located at a distance corresponding to this travel time.

Another powerful technique is called TOFD, Time Of Flight Diffraction. This method makes use of the full waveform of the ultrasonic pulses transmitted and received. The set-up is a separate transmitter and receiver, e.g. on each side of a weld to be inspected. The transmitter-receiver pair is moved along the weld and closely spaced recordings are made. The waveforms are plotted in grey-scale and interpreted in terms of diffraction theory, similar to wave optics. This makes visualization of small defects possible.

3. CONDITION MONITORING OF VALVES

When a process plant has a very high number of valves, it becomes a challenge to monitor them. Especially when valves are used to control critical fluids and even a small leakage in a valve cannot be tolerated, it becomes even more challenging. The following are the challenges in order to have online condition monitoring of valves:

- Too many valves of different types from different suppliers
- Different types of valves carrying different types of fluids
- Offline testing may be an option but it will lead to down time for the plant
- Predictive maintenance is preferred over periodic maintenance

To develop a condition monitoring system for valves, various aspects have to be taken into account. Suitable sensors are required depending on the type of valve, purpose of valve and criticality of valve for plant operation. Once the sensors are identified, an optimal wireless data logging system is required to collect data from the valve sensors, and to process and analyze a the data. Decision support system is to be designed to help operators identify a faulty valve with sufficient information; location of valve, type of fault, severity of fault, manufacturer with contact details etc.

4. CONDITION MONITORING OF HEAT EXCHANGERS

There are several commercial Non-Destructive Testing (NDT) condition monitoring methods available for heat exchangers; Ultrasonic testing, Visual inspection, Magnetic particle inspection, Helium leak test and Eddy current testing (Melingen 2010). All the techniques are briefly described, with more attention given to the Eddy current testing, as it is being used as the standard method by the company which presented the case.

4.1. Ultrasonic testing

A transducer sends a high frequency ultrasound pulse through the material and based on the reflected wave characteristics, the condition of the heat exchanger can be quantified. The advantages of ultrasonic testing are the following; it is sensitive towards both surface and subsurface discontinuities, the depth of penetration deeper compared to other methods, it can measure corrosion through thick walls, and furthermore it can also quantify the size of pits. However there some challenges to use this method; reliability of the method can be affected by poor surface finish, thick paint, and temperature. Only skilled personnel can perform the testing.

4.2. Visual inspection

Rigid/flexible fiber-optic boroscopes can give information about the condition of tubes/plates in heat exchangers. This method can be less expensive but it has its own limitations related to quantifying different types of faults.

4.3. Magnetic particle inspection

This method utilizes the magnetic properties of the heat exchanger material to detect a crack. The method is easy to apply and gives quick results. This method can be used for detecting cracks on the surface.

4.4. Helium leak test

Helium is used as trace gas in this method. The gas is pumped to the heat exchanger and a spectrometer is used to detect any leakage. It gives accurate results regarding the leak but it cannot be used to detect other faults than the leakage.

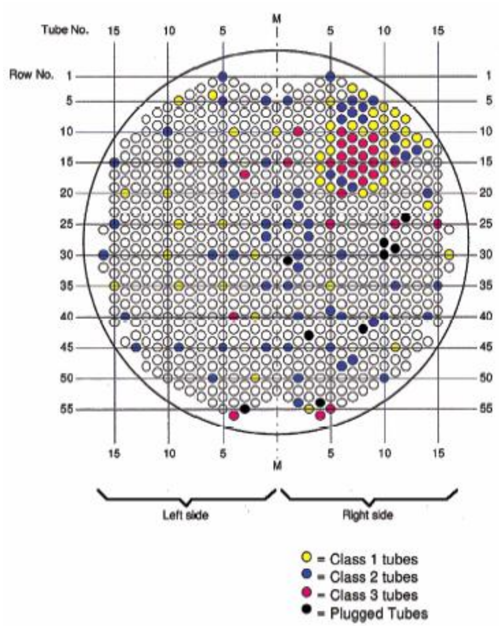


Figure 8: Decision support system from Force Technology

4.5. Eddy current testing

Eddy current based technology is being used to detect faults in thin heat exchanger tubes. The condition monitoring using eddy current based technology has several advantages, some of them are listed below:

- Very effective to detect cracks
- Gives immediate results
- Lot of data is available from test which can be analyzed to predict time for the next maintenance/replacement
- Portable equipment and contact-free
- The user interface system, for example from Force Technology (Force Technology) is very intuitive to find the severity of the fault and location of faulty tubes, see Figure 8.

However there are some challenges which are listed below:

- Calibration of the equipment is very specific to the material. Hence the equipment can only be used for inspection of particular equipment. If shall be used for other equipment, lot of calibration is necessary
- The results of inspection are sensitive to the thickness of the material.

Figure 9: Manifolds for cooling system at Eramet furnace. Ultrasound flow meters are mounted on the inlets and outlets, and water leakages are detected by investigating flow differences.

5. WATER LEAKAGE DETECTION AND CONTROL

In metallurgical plants, for instance Eramet Norway, it is necessary to handle molten materials in different unit operations. As the operating temperature can be as high as 2000°C, it is necessary to have a cooling system for the operating equipment and it is common to have water as the cooling liquid. In this circumstance, it is very critical not have any water leakage from the cooling system onto the molten material as the leakage would lead to explosions. Such explosions are of great concern for the safety of operators. Explosions caused by water leakage will also lead to equipment loss and possibly long term shutdown of the plant (even if the equipment is working, the plant may be shut down while the accident is investigated). The Norwegian government requires the plant to have a reliable system to identify, quantify and handle water leakages. The regulations also state that the water leakage alarms should be handled separately from other technical alarms due to the potential for serious dangers.

As the cooling system supplies cooling for different types of equipment, there can be a large number of water circuits. As a first measure to avoid leakages, it is important to use only clean water. This will minimize clogging and corrosion of pipes, and also ensure good cooling of the furnace hood. Monitoring of individual circuits for a leakage requires heavy instrumentation.

One system which was considered, but found unsuitable, was based on infrared imaging. IR images would be able to tell where a leakage occurs. However, automation of such a system is complicated, and since a large area has to be covered, many expensive cameras would have to be installed. Instead, a system has been implemented with two ultrasound flow instruments for each water circuit – one at inlet and one at outlet, see Figure 9. The difference in the flows at the inlet and outlet is compared to an estimate of the standard deviation over a specified time, to detect a fault. This statistical analysis is required to distinguish leak detection from measurement noise in a robust manner. Leaks down to 40 l/h can be determined using this method without or very few false alarms.

A good decision support system is designed to make sure that the leakage circuit is controlled/closed once a water leak is detected in a particular circuit. The flow instruments provide an opportunity to quantify leakages, and thus differentiate between critical alarms and alarms indicating small leakages, which do not require immediate action. These alarms are indicated by different colors in the HMI. The display observed by the operator also gives a clear message about which circuit is failing. It is important to train operators with operational procedures, what to do in the event of a water leakage in the cooling system corresponding to a critical equipment.

Overall the following are identified as the key challenges to have an optimal condition monitoring system for a process/equipment in the plant:

- Measuring key process/performance variables due to challenging process operating conditions
- Enough motivation for developing and implementing models for the application of fault detection and diagnosis
- Dedicating human resources for data analysis, and a challenge for a transformation from a more human control to less human control in process plants
- Optimal and intuitive decision support system design

6. FAILURE PREDICTION OF HYDRAULIC SYSTEMS

Off-shore industry relies heavily on hydraulic drilling equipment. Failure Prediction of hydraulic systems is an area of research which has recently attracted a considerable amount of research (Angeli and Atherton 2001). These systems are prone to wear over time leading to performance degradation and, ultimately, to failure. Due to the high cost of downtime in off-shore industry, detecting the wearing out of a component should take place at an early stage in order to avoid prominent failure.

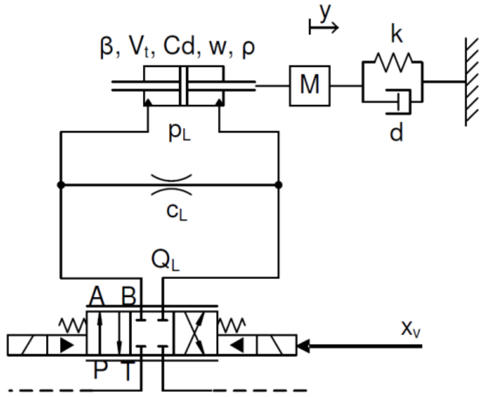


Figure 10: Nonlinear hydraulic-mechanical system with control valve and Nonlinear hydraulic-mechanical system with control valve and and hydraulic cylinder exerting forces on the object to be handled. Total load mass M, equivalent spring coefficient k and damping.

Too low forces in drilling operations might result in loss of grip while too high forces might jeopardize the drilling pipe. In a recent study pertinent to hydraulic drilling pipes (Choux and Blanke 2011), the authors devised an approach for prognosis that involves monitoring changes in two key parameters of the hydraulic system, namely leakage coefficient between the cylinder chambers and friction coefficient against the piston displacement. This stems from the fact that an increase of friction or the leakage coefficient lead to loss of grip. The nonlinearity of the hydraulic drilling equipment constitutes an inherent difficulty, which requires proper modeling techniques. In order to reduce the complexity of the nonlinear model, a general approach to decompose a nonlinear model into a linear hydraulic model connected to a mass-spring-damper system was introduced in (Choux et al 2009). Figure 10 is borrowed from reference (Choux and Blanke 2011) and depicts a nonlinear hydraulic-mechanical drilling pipe that is decomposed according the approach presented in (Choux et al 2009). Most of the legacy research resort to on-line model based fault detection techniques. In this sense, the current techniques are concerned with detecting deviation between the physical model of the hydraulic system (Choux and

Blanke 2011) and the real time data. To achieve this task, different online change detection techniques were employed including artificial neural networks (Muenchhof 2007), statistical abrupt change detection (Choux and Blanke 2011), Kalman filters (Chinniah et al 2008) and expert systems (Angeli and Atherton 2001).

CONCLUSIONS

Several industrial cases are presented in the paper with the focus of condition monitoring of equipment, failure prediction techniques and decision support systems. Further work will be focused on developing techniques for these specific cases with the cooperation of leading universities and R&D institutions.

ACKNOWLEDGMENTS

Authors acknowledge the participated industries for their support and contribution in this work. Further the financial support from Competence Development Fund of Southern Norway, Vest-Agder and Aust-Agder counties is acknowledged.

REFERENCES

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference*, pp. 487-499.

Angeli, C., Atherton, D., 2001. A Model-based Method for an Online Diagnostic Knowledge-based System, *Expert Systems*, 18(3), 150-158.

Becraft, W., Lee, P., 1993. An integrated neural network/expert system approach for fault diagnosis. *Computers and Chemical Engineering* 17 (10), 1001-1014.

Chinniah,Y., Burton, R., Habibi, S., Sampson, E., 2008. Identification of the nonlinear friction characteristics in a hydraulic actuator using the extended Kalman filter, *Transactions of the Canadian Society for Mechanical Engineering,* 32 (2), 121-136.

Choux, M., Blanke, M., 2011. Fault diagnosis for nonlinear hydraulic-mechanical drilling pipe handling system, *Proceedings of CDC-ECE,* pp. 1620-1626.

Choux M., Karimi, H.R., Hovland, G., Hansen, M.R., Ottestad, M., Blanke, M., 2009. Robust Adaptive Backstepping Control Design for a Nonlinear Hydraulic-Mechanical System, Proceedings of *48th IEEE Conference on Decision and Control, Shanghai*, pp. 2460-2467.

Dash, S., Venkatasubramanian, V., 2000. Challenges in the industrial applications of fault diagnostic systems. *Computers and Chemical Engineering*, 24(2), 785–791.

Force Technology, Tube Testing, *Eddy Current Inspection of Heat Exchanger Tubing*, Force Technology.

Gertler, J., Monajemy, R., 1995. Generating directional residuals with dynamic parity relations. *Automatica*, 31, 627-635.

Frank, P. M., Ding, S. X., Marcu, T., 2000. Model-based fault diagnosis in technical processes. *Transactions of the Institution of Measurement and Control,* 22 (1), 57-101.

Guiu, G. I., Golobardes, J. M., Bernado, E., Llora, E., 1999. Automatic diagnosis with genetic algorithms and case-based reasoning. *Artificial Intelligence in Engineering*, 13(4), 367–372.

Hoskins, J.C., Himmelblau, D.M., 1988. Artificial neural network models of knowledge representation in chemical engineering. *Computers and Chemical Engineering* 12, 881-890.

Lapp, S.A., Powers, G.A., 1977. Computer-aided synthesis of fault-trees, IEEE Trans. Reliability, 37, 2-13.

Melingen, D., 2010. Life Cycle Cost Model for Condition Monitoring of heat exchanger, *Master Thesis,* Faculty of Engineering Science, and Technology, NTNU, Trondheim.

Muenchhof, M., 2007. Semiphysical models of a hydraulic servo axis for fault detection, *Proceedings of the American Control Conference ACC '07*, pp. 1834–1839.

NDT Handbook, *Nondestructive testing Handbook, Volume 7, Third edition*, The American Society for Nondestructive Testing, USA.

Przytula, K.W., Thompson, D., 2000. Construction of Bayesian Networks for Diagnostics*, Proceedings of 2000 IEEE Aerospace Conference*, March 18-24.

Pöyhönen, S., Negrea, M., Arkkio, A., Hyötyniemi, H., Koivo, H., 2002. Support Vector Classification for Fault Diagnostics of an Electrical Machine, *Proceedings of the 6th Int. Conf. On Signal Processing ICSP.02*, pp. 1719-1722, Beijing-China.

Rasmussen, C.E., Williams, C.K.I, 2006. Gaussian Processes for Machine Learning. *MIT Press*.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S. N., 2003. A review of process fault detection and diagnosis. Part I. Quantitative model based methods. *Computers and Chemical Engineering*, 27(3), 293–311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003. A review of process fault detection and diagnosis. Part II. Qualitative models and search strategies. *Computers and Chemical Engineering*, 27(3), 313–326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003. A review of process fault detection and diagnosis Part III: Process history based methods. *Computers & Chemical Engineering*, 27 (3), pp. 327-346.

Wold, S., Ruhe, A., Wold, H., Dunn, W., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal of Science Statistical Computer,* 5, 735-743.

AIR-HEATING SET CONTROL VIA DIRECT SEARCH METHOD AND STRUCTURED SINGULAR VALUE

Marek Dlapa[(a)]

[(a)] Tomas Bata University in Zlin, Faculty of Applied Informatics, nám. T.G.Masaryka 5555, 760 01 Zlín, Czech Republic

[(a)]dlapa@fai.utb.cz

ABSTRACT

The contribution describes an application of the algebraic μ-synthesis to the control of a real plant with a nonlinear characteristic. The controller design is considered as a problem of minimization of the peak of the structured singular value denoted μ. The algebraic approach consists of the pole placement principle based on the polynomial Diophantine equations and Differential Migration procedure used for optimization. The results are compared with other controllers designed via the *D-K* iteration, synthesis in the ring of proper and stable functions and the Naslin method.

Keywords: robust control, evolutionary computation, genetic algorithms, algebraic approaches, structured singular value, PID control, time delay systems

1. INTRODUCTION

Intensive research activity performed in the robust control theory during the recent years has brought new methods considering the parametric, dynamic and mixed structured uncertainties. Some of the methods are based on the $\mathbf{H}_\infty$ approach in the ring of stable and proper transfer functions denoted R_{PS}. These methods provide a measure that indicates the robustness of designed controller. However, this measure evaluates only the robust stability. On the other hand, methods based on the Zames' small gain theorem (Zames 1981) yield both the robust stability and performance conditions. One of them is the structured singular value denoted μ (Doyle 1982, Packard and Doyle 1993) treating the robust stability and performance objectives simultaneously. Two methods for the μ-synthesis were derived: the *D-K* iteration (Doyle 1985) and μ-*K* iteration (Lin et al. 1993). The *D-K* iteration yields a suboptimal controller minimizing the peak of the upper bound for μ-function. However, the controller has usually a high order transfer function due to the scaling matrices D, D^{-1} and for further application it is simplified via some kind of approximation. If the simplification is too substantial it can cause degradation of the frequency properties of the controller and the whole feedback loop. In some cases, the scaling matrices cannot be approximated with the desired precision and the resulting controller can be far from the optimality. Moreover, the state-space formulae for the $\mathbf{H}_\infty$ suboptimal controller require the stability of the performance weighting function (Doyle et al. 1989). These problems can be resolved using the algebraic μ-synthesis (Dlapa et al. 2009, Dlapa and Prokop 2010) presented in this contribution, which overcomes both the approximation of the scaling matrices D, D^{-1} and the impossibility of integrating behaviour of the performance weighting function. In this method the controller is designed through the algebraic pole placement principle applied to the nominal plant and the position of the nominal closed-loop poles is tuned through an evolutionary algorithm with evaluation of the upper bound for μ. The problem of instability of performance weighting function is treated by setting the nominal closed-loop poles to the real axis in the left half-plane.

2. PLANT DESCRIPTION AND IDENTIFICATION

The control of heating systems has been an important field in the control theory for decades. There is a number of applications of temperature control involving nonlinear and time-delay systems present in the electrical and heating industry as well as in technology processes (e.g. Fiser 2002 or Liu 2003).

The problem of nonlinear control can be treated by adaptation to parameters changes or by using a robust or nonlinear controller. The usage of the adaptive control is limited by recursive identification, which has not satisfactory results when the input to the controller is subject to noise or if there are other factors, such as time delay or external disturbances causing inaccuracy of measured signals. The algebraic μ-synthesis presented in this contribution is more versatile than common methods for the robust control and can consider the effects of noise, nonlinearity and time delay as well as the influence of external disturbances. However, the usage of this method is limited to the models fitting in the linear fractional transformation (LFT) framework. In order to describe nonlinearity of controlled plant the parametric uncertainty is used and transformed to LFT interconnection which does not increase conservatism of the plant model.

The air-heating set considered in this contribution has three input and seven measured quantities. The input signals are the voltage on bulb and the main and adjacent fan. The circuit was controlled by a standard IBM PC computer, which communicates via serial link (RS232) with the CTRL unit (see Figure 1). The CTRL unit converts the digital data to unified analogue signals. In the transformation and unification unit the unified analogue signals are transformed to the voltage on a particular actuator. Similarly, the measured signals are transformed to the unified voltage 0-10 V (Table 1).

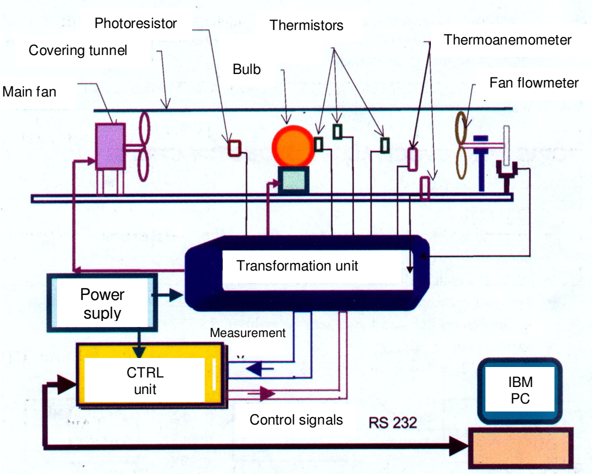


Figure 1: Plant Scheme

Table 1: Input and Output Channels of CTRL Unit

Inputs	Sensor	Outputs	Actuator
$y1$	sensor of bulb radiance	$u1$	voltage of bulb
$y2$	sensor of temperature near bulb T_2	$u2$	voltage of main fan (speed control)
$y3$	temperature of envelope of the bulb T_3	$u3$	voltage of adjacent fan (speed control)
$y4$	temperature at output of tunnel T_4		
$y6$	thermoanemometer TA_6		
$y7$	fan flowmeter		

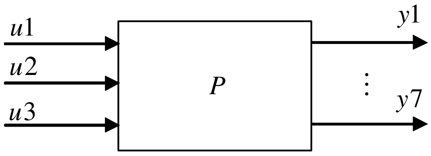


Figure 2: Inputs and Outputs of Plant

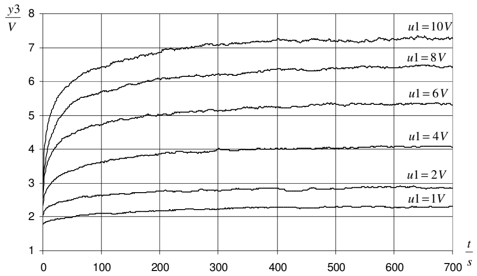


Figure 3: Step Responses of Bulb Temperature

The step responses of the measured quantity $y3$ (bulb temperature) for the step of $u1$ (bulb voltage) and constant speed of the main fan ($u2 = 2$V) are depicted in Figure 3. It is clear from the figure that the step responses have a nonlinear behaviour. The plant acts as if it has a short time constant at the beginning and it slows down at the end of history. Hence it should be taken into account that the time constant will vary in a large range. It follows from the steady-state load characteristic that the gain varies in the range of 0.42 to 0.54. The family of transfer function from $u1$ to $y3$ at $u2 = 2$V is:

$$\mathbf{P}_{31} = \left\{ \frac{k}{Ts+1} : k \in (0.42; 0.54), T \in (5; 20) \right\} \tag{1}$$

This means that both the numerator and denominator in (1) are interval polynomials.

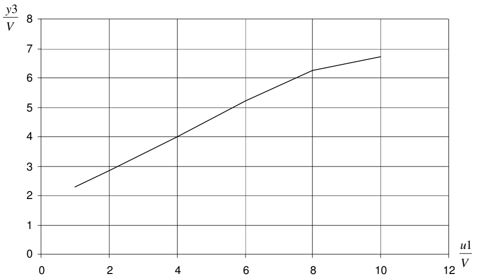


Figure 4: Steady-State Load Characteristic

3. OUTLINE OF POLE PLACEMENT DESIGN

The pole placement principle is one of the well-known methods for the controller design (e.g. Kucera 1991, Kucera 1993, Prokop and Corriou 1997, Prokop et al. 1992, Vidyasagar 1985) which is simple for derivation and tuning. Consider a simple feedback loop (1DOF) structure depicted in Figure 5 with two external inputs – the reference w and disturbance v, respectively. The output and tracking error is according to Figure 5 in the form

$$y(s) = \frac{bq}{d} w(s) + \frac{cfp}{d} v(s) \tag{2}$$

$$e(s) = w(s) - y(s) = \frac{afp}{d} \cdot \frac{h_w}{f_w} - \frac{cfp}{d} \cdot \frac{h_v}{f_v} \tag{3}$$

where

$$afp + bq = d \tag{4}$$

is the characteristic polynomial of the closed-loop system in Figure 5.

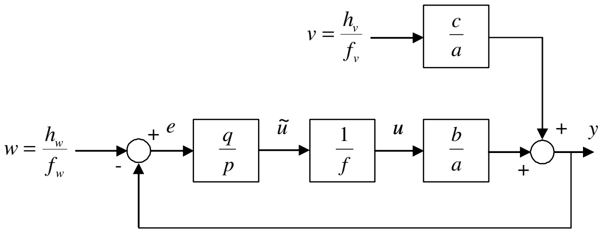


Figure 5: Structure of 1DOF system

It can be proven that the asymptotic tracking of the reference is achieved if and only if the polynomial *pfa* is divisible by the unstable part of f_w and v is rejected if *pfa* is divisible by the unstable part of f_v so that the result is a finite polynomial. As a consequence, the polynomials p, q are the solutions to Diophantine equation (4). It is also desirable that the transfer function $\frac{q}{fp}$ is proper. Analysis of the polynomial degrees in (4) for the most frequent case $f_w = f = s$ (the stepwise reference) gives

$$\deg d = 2\deg a \tag{5}$$

A standard choice for the polynomial d is

$$d(s) = \prod_{i=1}^{\deg d}(s+\alpha_i). \tag{6}$$

where $\alpha_i > 0$ are the tuning parameters of the controller and d is a stable polynomial which ensures the internal stability of the nominal system.

With respect to (1) a nominal plant transfer function can be expressed by the transfer function

$$P(s) = \frac{b(s)}{a(s)} = \frac{b_0}{s+a_0} \tag{7}$$

and

$$w = \frac{h_w}{f_w} = \frac{1}{s} \tag{8}$$

$$v = \frac{h_v}{f_v} = \frac{1}{s} \tag{9}$$

Then equation (4) has the form

$$(s+a_0)s + b_0(q_1 s + q_0) = s^2 + d_1 s + d_0 \tag{10}$$

and by simple equating the coefficients at the like power of s at the left and right of (10) it can be obtained

$$\begin{aligned} q_1 &= (d_1 - a_0)/b_0 \\ q_0 &= d_0/b_0 \end{aligned} \tag{11}$$

Then the resulting controller is proper and has the traditional PI structure in the form

$$Q = \frac{q_1 s + q_0}{s} \tag{12}$$

4. PRINCIPLES OF μ-SYNTHESIS

The parametric uncertainty in $\tilde{P}_{31}$ can be treated via the LFT framework by using the additive and quotient uncertainty. Define the nominal plant

$$P_{31} = \frac{0.48}{12.5s+1} \tag{13}$$

Plant family $\tilde{P}_{31}$ is then equivalent to

$$\tilde{P}_{31}(s,\Delta) = \frac{0.48 + 0.06\delta_2}{12.5s+1}\,\frac{1}{1+\dfrac{7.5\delta_3 s}{12.5s+1}}$$

$$\Delta = \begin{bmatrix} \delta_2 & 0 \\ 0 & \delta_3 \end{bmatrix},\ \delta_2, \delta_3 \in (-1;+1) \tag{14}$$

Let

$$W_2(s) = \frac{0.06}{12.5s+1},\ W_3(s) = \frac{7.5s}{12.5s+1} \tag{15}$$

then

$$\tilde{P}_{31}(s,\Delta) = (P_{31} + \delta_2 W_2)\frac{1}{1+\delta_3 W_3} \tag{16}$$

Expression (16) is represented by the LFT interconnection depicted in Figure 6.

The LFT interconnection for the μ-synthesis which considers the performance objectives and noise suppression is in Figure 7.

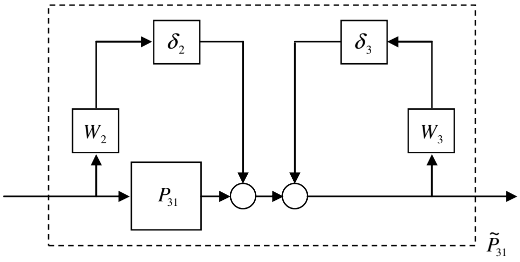


Figure 6: Structure of 1DOF system

Weight W_1 for the performance evaluation was chosen as

$$W_1(s) = \frac{0.01s+1}{s+0.0001}0.09 \text{ and } \frac{0.01s+1}{s}0.09 \tag{17}$$

for the *D-K* iteration and algebraic approach, respectively.

The weight for the *D-K* iteration cannot have integrating behaviour because all weights must be stable. Moreover, it causes uncontrollable states in the closed-loop system. The instability and uncontrollability of the closed-loop in Figure 7 does not make the resulting feedback loop unstable if there is a guarantee that the poles of the nominal feedback loop are in the left half plane. Controllability is a necessary condition for using the state space formulae giving the $\mathbf{H}_\infty$ suboptimal controller (Doyle et al. 1989) as well as stability of all weighting functions. Thus it is impossible to use these formulae in this case. The algebraic approach overcomes the problem by setting the nominal closed-loop poles to the left half-plane. Therefore, it is possible to use performance weights with poles at the imaginary axis, which guarantee the asymptotic tracking.

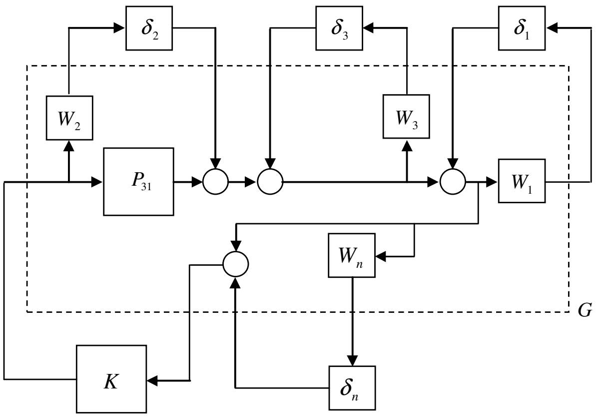


Figure 7: Structure of 1DOF system

The weight of noise is a band-pass filter, which takes into account high frequency noise emerging in sensors

$$W_n(s) = \frac{1/0.001s+1}{1/0.5s+1}0.001 \tag{18}$$

Let S denote a perturbed transfer function from the reference input w to the tracking error e. Let W_1 denote the weighting function and define the performance condition as

$$\|W_1 S\|_\infty \le 1 \tag{19}$$

If the condition (19) holds then behaviour of the closed loop can be changed through W_1.

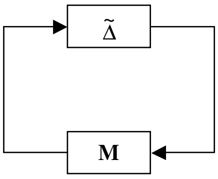


Figure 8: Transformed Closed-Loop System

The closed-loop feedback system in Figure 7 can be transformed to that in Figure 8, where $\mathbf{M}$ is a linear fractional transformation on $\mathbf{G}(s)$ and controller $K(s)$, i.e.,

$$\mathbf{M} = \mathbf{F}_l(\mathbf{G}, K) = \mathbf{G}_{11} + \mathbf{G}_{12}K(1-G_{22}K)^{-1}\mathbf{G}_{21} \tag{20}$$

where $\mathbf{G}(s)$ is the generalized plant including the nominal plant and weighting functions, which can be parted to

$$\mathbf{G}(s) = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & G_{22} \end{bmatrix} \tag{21}$$

where the rows and columns of $\mathbf{G}_{11}$ correspond to dynamic perturbations (22) and G_{22} corresponds to the controller structure. The other element of the system is mixed structured uncertainty which forms the diagonal matrix

$$\tilde{\mathbf{\Delta}} = \{\mathrm{diag}[\delta_1, \delta_2, \delta_3, \delta_n] : \delta_{2,3} \in \mathbf{R}, \delta_{1,n} \in \mathbf{C}\} \tag{22}$$

The algebraic μ-synthesis is applied to the 1DOF system for the interconnection depicted in Figure 7. The *D-K* iteration is applied to the same structure with W_1 without integrating behaviour.

The structured singular value of a matrix $\mathbf{M}$, denoted $\mu_{\tilde{\Delta}}(\mathbf{M})$, is defined as

$$\mu_{\tilde{\Delta}}(\mathbf{M}) = \frac{1}{\min\{\bar{\sigma}(\tilde{\Delta}) : \det(I - \mathbf{M}\tilde{\Delta}) = 0\}} \tag{23}$$

and if no such $\tilde{\Delta}$ exists which makes $\mathbf{I} - \mathbf{M}\tilde{\Delta}$ singular, let $\mu_{\tilde{\Delta}}(\mathbf{M}) = 0$ (Balas and Packard 1996), where $\bar{\sigma}(\tilde{\Delta})$ denotes the maximum singular value. The control objective is to find a stabilizing controller $\mathbf{K}$ minimizing the $\mathbf{H}_\infty$ norm of $\mu_{\tilde{\Delta}}(\mathbf{M})$, i.e.,

$$\min_{\mathbf{K}\text{ stabilizing }\mathbf{G}} \left\|\mu_{\tilde{\Delta}}[\mathbf{F}_l(\mathbf{G}, \mathbf{K})]\right\|_\infty \tag{24}$$

The following result is used for the robust performance test (Balas and Packard 1996):

Theorem 1: The feedback system with $|\tilde{\Delta}| < 1$ has the robust performance, i.e., expression (19) holds and the perturbed feedback loop is internally stable, if and only if

$$\mu_{\tilde{\Delta}}(\mathbf{M}) \le 1 \tag{25}$$

at all frequencies, where $|\cdot|$ denotes the absolute value. ■

In the algebraic approach the nominal closed-loop poles are the tuning parameters and the quality of the controller is evaluated by the upper bound for μ, i.e. as $\inf_{D\in\mathbf{D}} \bar{\sigma}(\mathbf{DMD}^{-1})$ for each frequency. The poles are constraint to the real axis in the left half-plane so that the stability of the nominal feedback loop is guaranteed. As the cost function defined by (24) as well as the upper bound has more than one local minimum an algorithm for global optimization is desirable. In this contribution Differential Migration (Dlapa 2009) was used for the optimization with high efficiency in finding the global minimum. Differential Migration is an evolutionary algorithm based on migration of individuals in the space of tuning parameters giving significantly higher robustness (in the sense of ability of finding the global minimum) than other algorithms of this class. The usage of this algorithm can shorten the computational time and increase the probability of finding the optimal pole placement.

5. CONTROL OF AIR-HEATING PLANT

Experimental studies have been carried out in order to assess the performance of the algebraic μ-synthesis method. The set-point temperature profile provides a reference which comprises 800 iterations. It consists of an initial soak at 3V for 480 iterations followed by a step to 4V which is held constant for 300 iterations. Sampling period is 1 *s* and adjacent fan voltage is ketp zero for all experiments.

The experimental trials are aimed at evaluating the performance of the PI controller obtained via the algebraic μ-synthesis against the *D-K* iteration, synthesis in the R_{PS} with the poles in one point and the Naslin method.

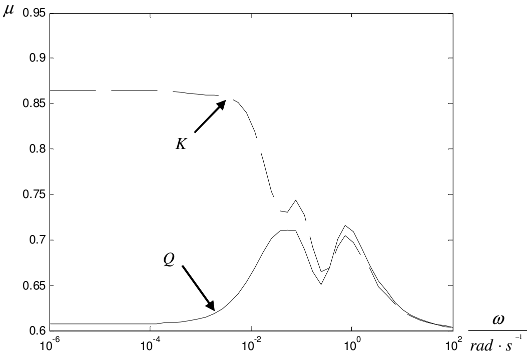


Figure 9: μ-Plot for *D-K* Iteration (dashed) and Algebraic μ-Synthesis (solid)

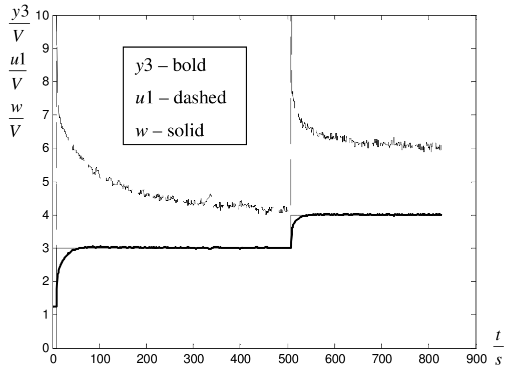


Figure 10: Control of Real Plant for Algebraic Approach

Performance indices. In order to draw comparisons between different control schemes an index or measure of performance is required. The measure of effectiveness which is adopted consists of three factors, these being the amount of energy, the variance of the controlled actuators and the accuracy of set-point tracking. This may be expressed as

$$\in_1 = \frac{\sum u(t)}{\rho} \tag{26}$$

where ρ is the number of iterations.

In cases where there is an increased variance in the control signals to the actuator this can lead to correspondingly increased costs due to maintenance and down time due to failure. The variance of the controlled actuator may be expressed in the form

$$\in_2 = \sum [u(t) - \in_1]^2 \tag{27}$$

The resulting controller quality arising from control action may be expressed in terms of the accuracy of set-point tracking. Using the integral of absolute error the deviation of the system response $y(t)$ from the set-point $r(t)$ is given as

$$\in_3 = \sum |y(t) - r(t)| \tag{28}$$

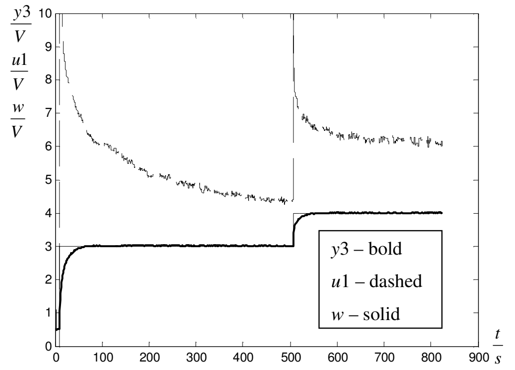


Figure 11: Control of Real Plant for *D-K* Iteration

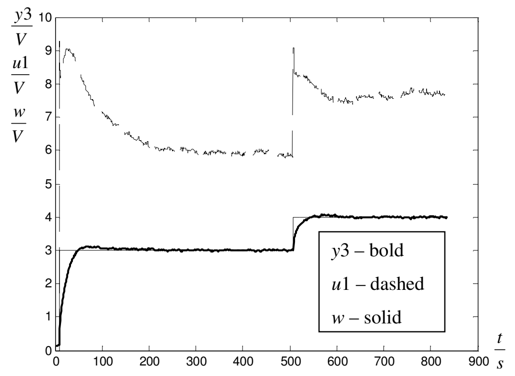


Figure 12: Control of Real Plant for Synthesis in R_{PS}

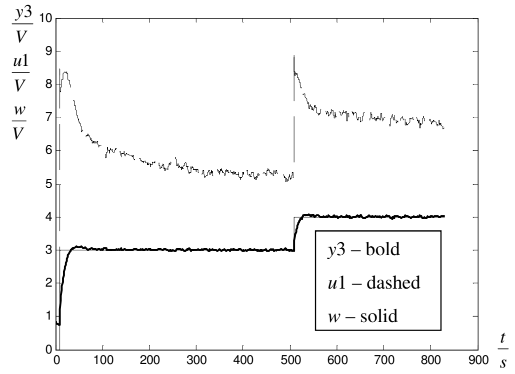


Figure 13: Control of Real Plant for Naslin Method

In order to provide a basis for comparison the controllers were tuned to give satisfactory overall performance across the complete temperature profile. The controllers designed in the ring of proper and stable functions (R_{PS}, see Prokop and Corriou 1997 or Prokop et al. 1992) and by the Naslin method were used as a reference (Figure 12 and 13). It is clear from Figure 9, 10, 11, 12, and 13 showing the system response and control input that the algebraic approach produces more favourable results. An improved accuracy in set-point tracking as well as lower energy consumption of the algebraic μ-synthesis (Figure 10) was achieved via increased control effort to the system resulting in higher variance in the control signal. The set-point, actuator and measured signals are in the unified voltage 0-10V of the CTRL unit. Performance indices are given in Table 2.

Table 2: Comparison of Performance Indices

Method	$\in_1$	$\in_2$	$\in_3$
Algebraic μ-synthesis	6.30	0.24	8.33
D-K iteration	6.37	0.23	9.07
R_{PS}	7.72	0.17	13.42
Naslin method	7.10	0.22	11.02

6. CONCLUSION

The contribution presents an application of the algebraic μ-synthesis to an air-heating set, where the temperature of bulb was controlled by its voltage. The controlled system had a nonlinear behaviour in both the steady-state and dynamic characteristics. The nonlinearity was treated via parametric uncertainty, which was transformed to the LFT framework. In order to achieve asymptotic tracking, performance weighting function with pole at the imaginary axis was used. The instability of the nominal feedback loop was treated by setting its poles to the left half-plane via pole placement technique and by choosing the PI structure of the controller which treats the unstable pole of the general closed-loop interconnection. The algebraic μ-synthesis was applied to the LFT interconnection including performance weight with integrating behaviour and the results were compared with standard methods for robust controller design - the *D-K* iteration, synthesis in R_{PS} and the Naslin method. Finally, it was shown that the algebraic μ-synthesis had better frequency properties in terms of μ-function and faster set-point tracking than the reference methods.

The algebraic μ-synthesis provides exploitable benefits for a wide range of industrial applications. In contrast to the *D-K* iteration, it can tune simple controllers in a more natural way and guarantee asymptotic tracking, which is desirable in most of the control tasks emerging in technology processes. The only drawback is the fact that an algorithm of global optimization is used leading to longer computational times.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund under the Project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

REFERENCES

Balas, G.J. and Packard A., 1996. The Structured Singular Value (μ) Framework. In: *The control handbook*, (Levine, W.S. (Ed)), pp. 671-687, CRC Press, Inc, IEEE Press.

Dlapa, M., 2009. "Differential Migration: Sensitivity Analysis and Comparison Study," *Proceedings of 2009 IEEE Congress on Evolutionary Computation (IEEE CEC 2009)*, pp. 1729-1736, ISBN 978-1-4244-2959-2.

Dlapa, M., Prokop R. and Bakošová M., 2009. "Robust Control of a Two Tank System Using Algebraic Approach," *Proceedings of EUROCAST 2009*, pp. 603-609, ISBN 978-84-691-8502-5.

Dlapa, M. and Prokop R., 2010. Algebraic approach to controller design using structured singular value, *Control Engineering Practice*, **Vol. 18**, No. 4, pp. 358–382, ISSN 0967-0661.

Doyle, J.C., 1982. "Analysis of feedback systems with structured uncertainties," *Proceedings of IEE*, Part-D, 129, pp. 242-250.

Doyle, J.C., 1985. "Structure uncertainty in control system design," *Proceedings of 24th IEEE Conference on decision and control*, pp. 260-265.

Doyle, J.C., Khargonekar P.P. and Francis B.A., 1989. State-space solutions to standard H2 and H∞ control problems, *IEEE Transactions on Automatic Control*, **Vol. 34**, No. 8, pp. 831-847.

Fiser, R., 2002. "Control algorithms of 200 MW production blocks in power plant Chvaletice for stationary and also non-stationary operation, experience, development," *Proceedings of Control of power & heating systems 2002*, Zlin, Czech Republic.

Kucera, V., 1991. *Analysis and design of discrete linear control systems*, Academia.

Kucera, V., 1993. "Diophantine equations in control – a survey," *Automatica*, **Vol. 29**, No. 6, pp. 1361-75.

Lin, J.L., Postlethwaite I. and Gu D. W., 1993. "μ-K iteration: a new algorithm for μ synthesis," *Automatica*, **29**, 219-224.

Liu, H., Liand S. and Chai T., 2003. "Intelligent decoupling control of power plant main steam pressure and power output," *International Journal of Electrical Power & Energy Systems*, **Vol. 25**, No. 10, pp. 809-819.

Packard A. and Doyle, J. C. 1993. "The complex structured singular value," *Automatica*, **Vol. 29(1)**, pp. 71-109.

Prokop, R. and Corriou J. P., 1997. "Design and analysis of simple robust controllers," *Int. J. Control*, **Vol. 66**, No. 6, pp 905-921.

Prokop, R., Dostal P. and Meszaros A., 1992. "Continuous-time control via proper regulators," *Journal A*, **Vol. 33**, No. 4.

Vidyasagar, M., 1985. *Control system synthesis – A factorization approach*, MIT Press Cambridge.

Zames, G., 1981. "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverse," *IEEE Trans. Auto. Control*, **26**, pp. 301-320.

MODELING AND SIMULATION OF SEVERE SLUGGING IN AIR-WATER SYSTEMS INCLUDING INERTIAL EFFECTS

Jorge Luis Baliño

Departmento de Engenharia Mecânica, Escola Politécnica, Universidade de São Paulo
Av. Prof. Mello Moraes, 2231, CEP 05508-900, Cidade Universitária, São Paulo-SP, Brazil

jlbalino@usp.br

ABSTRACT

A mathematical model and numerical simulations corresponding to severe slugging in air-water pipeline-riser systems, are presented. The mathematical model considers continuity equations for liquid and gas phases, with a simplified momentum equation for the mixture. A drift-flux model, evaluated for the local conditions in the riser, is used as a closure law. In many models appearing in the literature, propagation of pressure waves is neglected both in the pipeline and in the riser. Besides, variations of void fraction in the stratified flow in the pipeline are also neglected and the void fraction obtained from the stationary state is used in the simulations. This paper shows an improvement in a model previously published by the author, including inertial effects. In the riser, inertial terms are taken into account by using the rigid water-hammer approximation. In the pipeline, the local acceleration of the water and gas phases are included in the momentum equations for stratified flow, allowing to calculate the instantaneous values of pressure drop and void fraction. The developed model predicts the location of the liquid accumulation front in the pipeline and the liquid level in the riser, so it is possible to determine which type of severe slugging occurs in the system. A comparison is made with experimental results published in literature including a choke valve and gas injection at the bottom of the riser, showing very good results for slugging cycle and stability maps.

Keywords: severe slugging, pipeline-riser system, air-water flow, stability, lumped and distributed parameter systems, switched systems, petroleum production technology

1. INTRODUCTION

Severe slugging is a terrain dominated phenomenon, characterized by the formation and cyclical production of long liquid slugs and fast gas blowdown. Severe slugging may appear for low gas and liquid flow rates when a section with downward inclination angle (pipeline) is followed by another section with an upward inclination (riser). This configuration is common in off-shore petroleum production systems. Main issues related to severe slugging are: a) High average back pressure at well head, causing tremendous production losses, b) High instantaneous flow rates, causing instabilities in the liquid control system of the separators and eventually shutdown, and c) Reservoir flow oscillations.

For steady state and low flow rates, the flow pattern in the pipeline may be stratified, while it may be intermittent in the riser, as shown in Fig. 1(a).

A cycle of severe slugging can be described as taking place according to the following stages (Taitel, 1986). Once the system destabilizes and gas passage is blocked at the bottom of the riser, liquid continues to flow in and gas already in the riser continues to flow out, being possible that the liquid level in the riser falls below the top level at the separator. As a consequence, the riser column becomes heavier and pressure at the bottom of the riser increases, compressing the gas in the pipeline and creating a liquid accumulation region. This stage is known as slug formation (Fig. 1(b)).

As the liquid level reaches the top while the gas passage is kept blocked at the bottom, pressure reaches a maximum and there is only liquid flowing in the riser. This is the slug production stage (Fig. 1(c)).

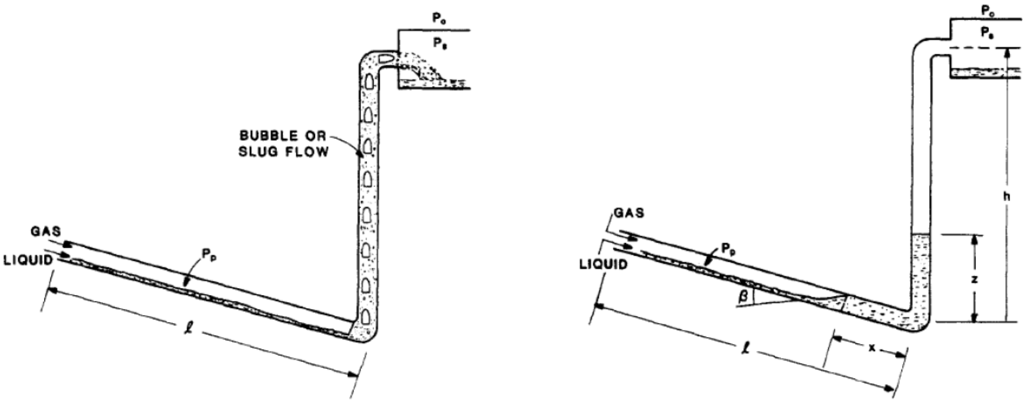
Since gas keeps flowing in the pipeline, the liquid accumulation front is pushed back until it reaches the bottom of the riser, starting the blowout stage (Fig. 1(d)).

As the gas phase penetrates into the riser the column becomes lighter, decreasing the pressure and then rising the gas flow. When gas reaches the top of the riser, gas passage is free through the stratified flow pattern in the pipeline and the intermitent/annular flow pattern in the riser, causing a violent expulsion and a rapid decompression that brings the process to slug formation again. This stage is known as gas blowdown (Fig. 1(e)).

Figure 1(f) shows the different stages in the pressure history at the bottom of the riser corresponding to an experiment under laboratory conditions (Schmidt, 1977).

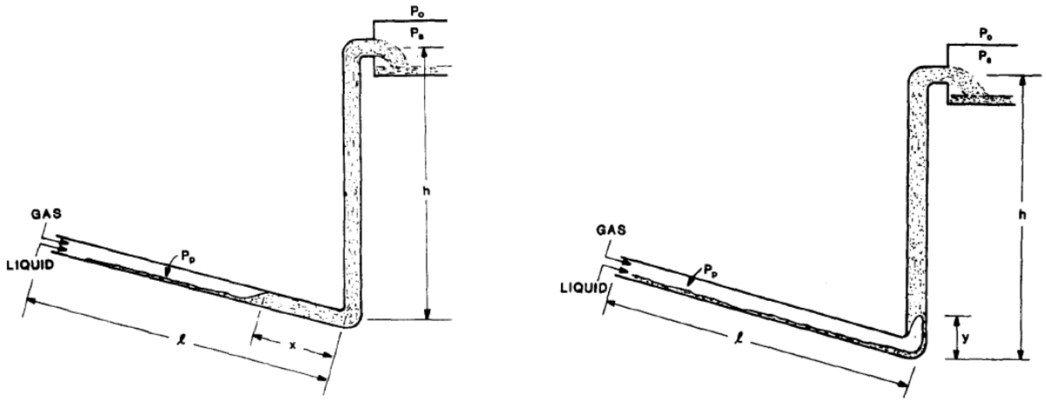
A classification of severe slugging can be made, according to the observed flow regime, as follows:

- Severe Slugging 1 (SS1): the liquid slug length is greater to or equal to one riser length and maximum pipeline pressure is equal to the hydrostatic head of the riser (neglecting friction pressure drop).
- Severe Slugging 2 (SS2): the liquid length is less than one riser length, with intermittent gas penetration at the bottom of the riser.
- Severe Slugging 3 (SS3): there is continuous gas penetration at the bottom of the riser; visually, the flow in the riser resembles normal slug flow, but pressure, slug lengths and frequencies reveal cyclic variations of smaller periods and amplitudes compared to SS1.



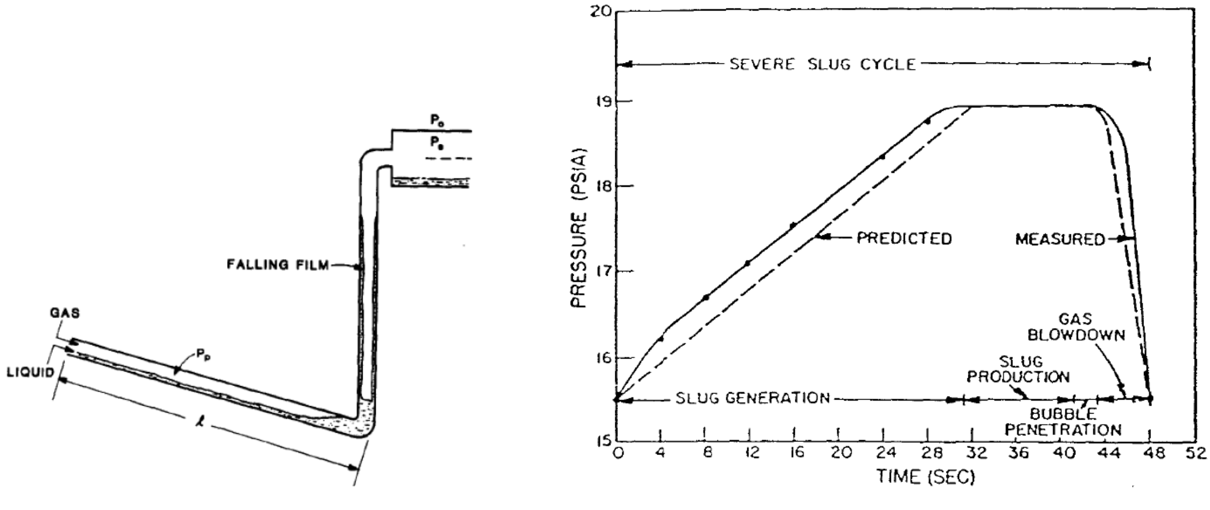
(a) Steady state.

(b) Slug formation.



(c) Slug production.

(d) Gas blowout.



(e) Gas blowdown.

(f) Pressure history.

Figure 1: Stages for severe slugging (from (Taitel, 1986) and (Schmidt, 1977)).

- Oscillation (OSC): there are cyclic pressure fluctuations without the spontaneous vigorous blowdown.

Most of the models for severe slugging were developed for vertical risers and assume one-dimensional, isothermal flow and a mixture momentum equation in which only the gravitational term is important.

In (Taitel *et al.*, 1990) a model was presented considering constant mean values for the gas density and void fraction in the riser, allowing to calculate time variations of pipeline pressure, position of the accumulation region, flow rate into the riser and mean holdup. It was found that as the operation point moves closer to the stability line the numerical procedure did not converge, giving gas mass flows going to infinite as the spatial discretization was decreased. Experimental data were obtained from a facility for different buffer volumes (simulating equivalent pipeline lengths) and a comparison was made with the simulation results, showing good agreement except for the blowout/blowdown stage. Setting apart the non-convergence problems, lumped parameter models seem to work fine for short risers, where the local variations of variables are small, but are not successful in long risers, typical of offshore systems.

In (Sarica and Shoham, 1991) a model with a distributed parameter formulation for the riser was presented. Considering continuity equations for the liquid and gas without phase change and a gravity-dominant mixture momentum equation, the model was capable of handling discontinuities such as liquid accumulation in the piping and liquid level in the riser. The resulting equations were solved by using the method of characteristics. A comparison of simulations with different experimental data showed reasonable agreement, although the model also suffered from non-convergence in the unstable region.

In (Baliño *et al.*, 2010) a model for severe slugging valid for risers with variable inclination was presented. The model was used to simulate numerically the air-water multiphase flow in a catenary riser for the experimental conditions reported in (Wordsworth *et al.*, 1998). Stability and flow regime maps in the system parameter space for the multiphase flow in a catenary pipeline-riser system were built.

In (Nemoto and Baliño, 2012) the model developed in (Baliño *et al.*, 2010) was extended to investigate the dynamics of gas, oil and water flow in a pipeline-riser system. Mass transfer between the oil and gas phases was calculated using the black oil approximation. The properties of fluids were calculated by analytical correlations based on experimental results and field data.

The stationary solution for a given point in the system parameter space is given as initial condition for the numerical simulation; if the numerical solution does not go away from the initial condition with time, the stationary solution is stable and it is the system steady state. If the numerical solution goes away with time, the stationary state is unstable, there is no steady state and an intermittent solution develops with time. By changing the point in the system parameter space and repeating this process, the stability map can be built. For unstable flow, the analysis of the limit cycle leads to the determination of the flow regime map, showing the regions corresponding to the different types of intermittency.

In all the models reviewed above, propagation of pressure waves is neglected both in the pipeline and in the riser; this constitutes the no-pressure-wave (NPW) approximation (Masella *et al.*, 1998). As a result of the NPW approximation, pressure changes are felt instantaneously at any point in the domain (pressure waves travel at infinite speed).

Also in the models reviewed above, a stratified flow pattern is assumed at the pipeline and variations of void fraction are neglected. The void fraction is obtained from a momentum balance in the gas and liquid phases, resulting an algebraic relation between the mean variables (Taitel and Dukler, 1976). The void fraction determined in the stationary state is assumed as constant in the simulations, so the momentum balance equation is not satisfied and variations of the void fraction cannot be calculated in the transients.

This paper shows an improvement in the model previously published by the author (Baliño *et al.*, 2010), including inertial effects.

In the riser, inertial terms are taken into account by using the rigid water-hammer approximation (Chaudhry, 1987). In this approximation, the acceleration terms for the liquid and gas phases are taken into account in the momentum equation, but compressibility of the liquid phase is neglected in the mass conservation equation. In the pipeline, convective acceleration terms are neglected but the local acceleration terms for the water and gas phases are included in the momentum equations for stratified flow, allowing to calculate the instantaneous values of pressure drop and void fraction.

The model includes additional devices, such as a valve located at the top of the riser and a gas injection line at the bottom of the riser. In this way it is possible to evaluate valve closure and gas lift as mitigating actions for severe slugging.

2. MODEL

The model considers one-dimensional flow in both pipeline and riser subsystems. The liquid phase is assumed incompressible, while the gas phase is considered as an ideal gas. Both phases flow in isothermal conditions. The flow pattern in the pipeline is assumed as stratified. The model is capable of handling discontinuities in the flow, such as liquid accumulation in the pipeline, liquid level in the riser and void fraction waves.

2.1 Pipeline

The pipeline, shown in Fig. 2, can be either in a condition of liquid accumulation ($x > 0$) or in a condition of continuous gas penetration ($x = 0$), where x is the position of the liquid accumulation front. The existence of a buffer vessel with volume υ_e is considered in order to simulate an equivalent pipeline length $L_e = \frac{\upsilon_e}{A}$, where A is the flow passage area ($A = \frac{1}{4}\,\pi\,D^2$, where D is the inner diameter).

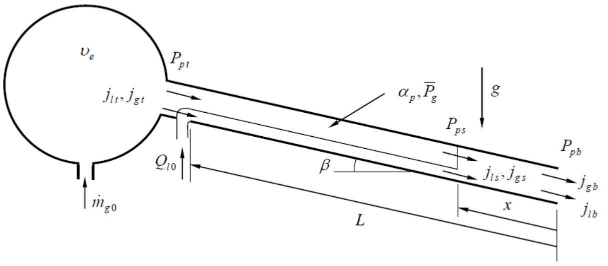


Figure 2: Definition of variables at the pipeline.

Three locations in the pipeline can be identified, with corresponding pressures and superficial velocities for the liquid and gas phases: the bottom ($P_{p\,b}$, $j_{l\,b}$ and $j_{g\,b}$), the stratified region located at the liquid accumulation front ($P_{p\,s}$, $j_{l\,s}$ and $j_{g\,s}$) and the top ($P_{p\,t}$, $j_{l\,t}$ and $j_{g\,t}$).

Mass and momentum conservation equations are applied for the control volumes corresponding to the liquid and gas phases. In the mass equation, a mean pressure $\overline{P}_g(t) = \frac{1}{2}\,(P_{p\,t} + P_{p\,s})$ is assumed. In mass and momentum equations, variations of the void fraction with position are neglected and a mean void fraction $\alpha_p(t)$ is considered. These approximations are made in order to get a lumped parameter model for the pipeline.

The superficial velocities at the top of the pipeline can be written as:

$$j_{l\,t} = \frac{Q_{l\,0}}{A} \tag{1}$$

$$j_{g\,t} = \frac{\dot{m}_{g\,0}\,R_g\,T_g}{P_{p\,t}\,A} \tag{2}$$

where $\dot{m}_{g\,0}$ and $Q_{l\,0}$ are respectively the gas mass flow and the liquid volumetric flow injected in the pipeline and R_g and T_g are respectively the gas constant and temperature.

2.1.1 Condition $x > 0$

For this condition there is no gas flowing out the pipeline, resulting:

$$j_{g\,b} = 0 \tag{3}$$

Applying mass conservation equation for the liquid and gas phases, it is obtained:

$$\frac{d\alpha_p}{dt} = \frac{j_{l\,b} - \frac{Q_{l\,0}}{A} + \alpha_p\,\frac{dx}{dt}}{L - x} \tag{4}$$

$$\frac{d\overline{P}_g}{dt} = \frac{-\overline{P}_g\left(j_{l\,b} - \frac{Q_{l\,0}}{A}\right) + \frac{R_g\,T_g}{A}\,\dot{m}_{g\,0}}{(L-x)\,\alpha_p + L_e} \tag{5}$$

where L is the pipeline length and t is time. A mass balance for the liquid in the stratified region yields:

$$-\,(L-x)\,\frac{d\alpha_p}{dt} + j_{l\,s} - \frac{Q_{l\,0}}{A} = 0 \tag{6}$$

From Eq. (4) and (6), it results:

$$j_{l\,s} = j_{l\,b} + \alpha_p\,\frac{dx}{dt} \tag{7}$$

From the kinematic condition at the liquid penetration front, we get:

$$j_{g\,s} = -\alpha_p\,\frac{dx}{dt} \tag{8}$$

Applying the momentum conservation equation in the control volume of liquid region at the penetration front and including friction and inertial terms, we get:

$$P_{p\,b} = P_{p\,s} + \rho_l\,x\,\left(g\,\sin\beta - \frac{2\,f_{ll}}{D}\,j_{l\,b}\,|j_{l\,b}| - \frac{dj_{l\,b}}{dt}\right) \tag{9}$$

where f_{ll} is the Fanning friction factor (assuming that only liquid is filling the cross sectional area), g is the gravity acceleration constant, ρ_l is the liquid density and β is the pipeline inclination angle (positive when downwards). The Fanning friction factor is calculated by using the correlation from (Chen, 1979).

2.1.2 Condition $x = 0$

For this condition there is no liquid penetration front, resulting:

$$j_{g\,s} = j_{g\,b} \tag{10}$$

$$j_{l\,s} = j_{l\,b} \tag{11}$$

$$P_{p\,s} = P_{p\,b} \tag{12}$$

Applying mass conservation equation for the liquid and gas phases, it is obtained:

$$\frac{d\alpha_p}{dt} = \frac{1}{L}\left(j_{l\,b} - \frac{Q_{l\,0}}{A}\right) \tag{13}$$

$$\frac{d\overline{P}_g}{dt} = \frac{-\overline{P}_g\left(j_{g\,b} + j_{l\,b} - \frac{Q_{l\,0}}{A}\right) + \frac{R_g\,T_g}{A}\,\dot{m}_{g\,0}}{L\,\alpha_p + L_e} \tag{14}$$

2.1.3 Local equilibrium condition for stratified flow

In previous models, the void fraction at the pipeline is determined from an algebraic relationship evaluated from the stationary state, derived from the momentum balance in stratified flow (Taitel and Dukler, 1976). This relationship can be generalized by including the local inertial terms. Assuming stratified flow (see Fig. 3), the generalized relationship can be written as:

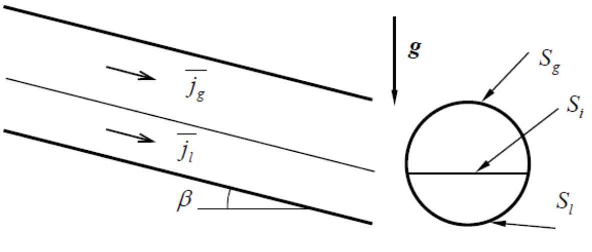


Figure 3: Stratified flow at the pipeline.

$$\tau_{w\,g}\,\frac{S_g}{\alpha_p} - \tau_{w\,l}\,\frac{S_l}{1-\alpha_p} + \tau_i\,S_i\left(\frac{1}{1-\alpha_p}+\frac{1}{\alpha_p}\right)$$
$$+\left(\rho_l-\overline{\rho}_g\right)A\,g\,\sin\beta - A\left(\rho_l\,\frac{d\overline{u}_l}{dt}-\overline{\rho}_g\,\frac{d\overline{u}_g}{dt}\right)=0 \tag{15}$$

$$\overline{u}_g = \frac{\overline{j}_g}{\alpha_p} \tag{16}$$

$$\overline{u}_l = \frac{\overline{j}_l}{1-\alpha_p} \tag{17}$$

$$\overline{\rho}_g = \frac{\overline{P_g}}{R_g\,T_g} \tag{18}$$

$$\overline{j}_g = \frac{1}{2}\left(j_{g\,t}+j_{g\,s}\right) \tag{19}$$

$$\overline{j}_l = \frac{1}{2}\left(\frac{Q_{l\,0}}{A}+j_{l\,s}\right) \tag{20}$$

where $\overline{j}_g$ and $\overline{j}_g$ are respectively the mean superficial velocities for gas and liquid, S_g, S_i and S_l are respectively the gas, interfacial and liquid wetted perimeters, $\tau_{w\,g}$, τ_i and $\tau_{w\,l}$ are respectively the wall-gas, interface and wall-liquid shear stresses, $\overline{u}_g$ and $\overline{u}_g$ are respectively the mean phase velocities for gas and liquid and ρ_g is the gas density.

In Eq. (15) the wetted and interfacial perimeters are determined considering a stratified geometry, while the shear stresses are related to the superficial velocities of the phases through friction factors based on the hydraulic diameters for each phase (Kokal and Stanislav, 1989).

2.2 Riser

Continuity equations for the phases are considered at the riser (see Fig. 4). This results in the following set of equations:

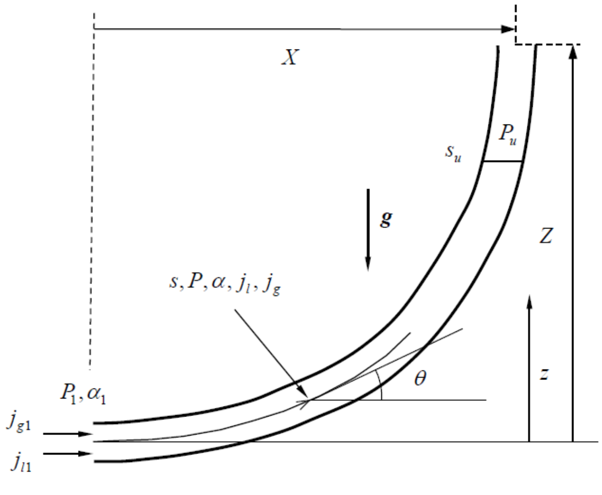


Figure 4: Definition of variables at the riser.

$$-\frac{\partial\alpha}{\partial t}+\frac{\partial j_l}{\partial s}=0 \tag{21}$$

$$\frac{\partial}{\partial t}\left(P\,\alpha\right)+\frac{\partial}{\partial s}\left(P\,j_g\right)=0 \tag{22}$$

where j_g, j_l and j are respectively the gas, liquid and total superficial velocities, P is the pressure, s is the position along the riser, α is the void fraction and u_g and u_l are respectively the gas and liquid phase velocities.

The superficial velocities for the phases are determined by using a drift flux correlation, assumed to be locally valid:

$$j_g = u_g\,\alpha = \alpha\left(C_d\,j+U_d\right) \tag{23}$$

$$j_l = j - j_g = u_l\,\left(1-\alpha\right) = \left(1-\alpha\,C_d\right)j - \alpha\,U_d \tag{24}$$

It will be assumed that the drift parameters C_d and U_d depend at most on the local flow conditions and inclination angle $\theta = \theta\,(s)$, this is, $C_d = C_d\,(\alpha,\,P,\,j,\,\theta)$ and $U_d = U_d\,(\alpha,\,P,\,j,\,\theta)$ (Bendiksen, 1984; Chexal *et al.*, 1992).

The drift coefficients used in the model are (Bendiksen, 1984):

- For $Fr_j < 3.5$:

$$C_d = 1.05 + 0.15\,\sin\,\theta \tag{25}$$

$$U_d = \sqrt{g\,D}\,(0.35\,\sin\,\theta + 0.54\,\cos\,\theta) \tag{26}$$

- For $Fr_j \geq 3.5$:

$$C_d = 1.2 \tag{27}$$

$$U_d = 0.35\,\sqrt{g\,D}\,\sin\,\theta \tag{28}$$

where the Froude number Fr_j is defined as:

$$Fr_j = \frac{j}{\sqrt{g\,D}} \tag{29}$$

Considering as the state variables in the riser the void fraction, pressure and total superficial velocity (functions of position and time), Eq. (21) and (22) can be finally rewritten as:

$$\frac{D_g\alpha}{Dt}+\alpha\,\frac{\partial}{\partial s}\left(C_d\,j+U_d\right)-\frac{\partial j}{\partial s}=0 \tag{30}$$

$$\alpha\,\frac{D_gP}{Dt}+P\,\frac{\partial j}{\partial s}=0 \tag{31}$$

where:

$$\frac{D_g}{Dt}=\frac{\partial}{\partial t}+u_g\,\frac{\partial}{\partial s} \tag{32}$$

A mixture momentum equation is considered, in which the inertial terms corresponding to the liquid and gas phases were included :

$$\frac{\partial P}{\partial s} = -\rho_m\left(g\,\sin\theta + 2\,\frac{f_m}{D}\,j\,|j|\right)$$
$$-\alpha\,\rho_g\,\frac{D_gu_g}{Dt}-\left(1-\alpha\right)\rho_l\,\frac{D_lu_l}{Dt} \tag{33}$$

$$\rho_m = \rho_l\,(1-\alpha)+\frac{P}{R_g\,T_g}\,\alpha \tag{34}$$

$$\frac{D_l}{Dt}=\frac{\partial}{\partial t}+u_l\,\frac{\partial}{\partial s} \tag{35}$$

where f_m is the Fanning friction factor (dependent on the Reynolds number and the relative roughness ϵ/D, where ϵ is the pipe roughness), calculated from (Chen, 1979) using a homogeneous mixture two-phase model and ρ_m is the mixture density.

2.2.1 Gas lift

Gas lift is a process used to artificially lift fluids from wells where there is insufficient reservoir pressure. By injecting gas, it is possible to aerate the liquid column and to reduce the pressure gradient. Due to this effect, it is acknowledged that gas lift can stabilize the flow in severe slugging, although relatively large gas flow rates are necessary (Jansen *et al.*, 1996). In the model it is considered the possibility of injecting gas in a position s_l along the riser. Considering the balance conservation equations with a gas mass source term, it can be shown that the gas injection introduces the following discontinuities in position in the gas superficial velocity, pressure and void fractions:

$$j_g^+ = \frac{\frac{R_g\,T_g}{A}\,\dot{m}_{g\,l} + P^-\,j_g^-}{P^+} \tag{36}$$

$$P^- = \frac{P^+\left[1 - \frac{1}{R_g\,T_g}\left(\frac{j_g^2}{\alpha}\right)^+\right] - \rho_l\,j_l^2\left(\frac{1}{1-\alpha^+} - \frac{1}{1-\alpha^-}\right)}{1 - \frac{1}{R_g\,T_g}\left(\frac{j_g^2}{\alpha}\right)^-} \tag{37}$$

$$\alpha^+ = \frac{j_g^+}{\left(C_d\,j + U_d\right)^+} \tag{38}$$

where $\dot{m}_{g\,l}$ is the injected gas lift mass flow and the superscripts $-$ and $+$ denote evaluation respectively at location upstream and downstream of position s_l.

2.2.2 Gas region

In transients where the liquid level falls below the top of the riser ($s_u < s_t$) a gas region is formed, as shown in Fig. 5. This region is modeled considering a constant mean pressure $\overline{P}_{g\,r} = \frac{1}{2}\left(P_t + P_u\right)$ for the mass balance equation and friction, gravitational and inertial terms in the momentum balance equation; P_t and P_u are respectively the pressure at the top and at the liquid level in the riser. The resulting equations for the mass balance depends on the location of the gas lift injection s_l compared to the liquid level position s_u.

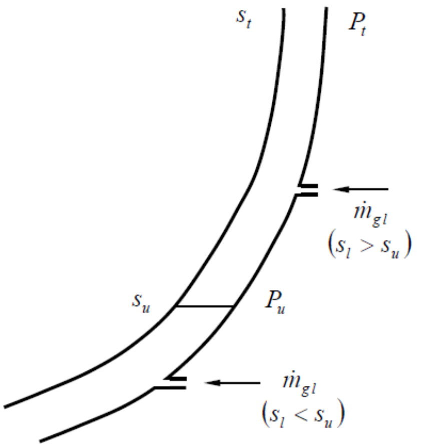


Figure 5: Definition of variables at the gas region.

If $s_u > s_l$, we get:

$$\frac{d\overline{P}_{g\,r}}{dt} = -\frac{\overline{P}_{g\,r}}{s_t - s_u}\left(j_{g\,t} - j_u\right) \tag{39}$$

If $s_u > s_l$, we get:

$$\frac{d\overline{P}_{g\,r}}{dt} = -\frac{\overline{P}_{g\,r}}{s_t - s_u}\left(j_{g\,t} - j_u - \frac{R_g\,T_g}{\overline{P}_{g\,r}\,A}\,\dot{m}_{g\,l}\right) \tag{40}$$

where $j_{g\,t}$ is the gas superficial velocity at the top of the riser, j_u is the total superficial velocity at the liquid level and s_t is the position of the top of the riser.

The momentum conservation equation results, for both cases:

$$\begin{aligned} P_u = P_t + \overline{\rho}_g \Big[& g\left(z_t - z_u\right) \\ & + \left(\frac{2\,f_{gg}}{D}\,\overline{j}_{g\,r}\,\left|\overline{j}_{g\,r}\right| - \frac{d\overline{j}_{g\,r}}{dt}\right)\left(s_t - s_u\right)\Big] \end{aligned} \tag{41}$$

$$\overline{j}_{g\,r} = \frac{1}{2}\left(j_u + j_{g\,t}\right) \tag{42}$$

where f_{gg} is the Fanning friction factor considering only gas flowing, $\overline{j}_{g\,r}$ and $\overline{\rho}_g$ are respectively the mean gas superficial velocity and the mean gas density at the gas region; z_t and z_u are respectively the vertical positions at the top of the riser and at the liquid level.

2.2.3 Riser geometry

The riser geometry is characterized by the coordinates X and Z corresponding to the top of the riser and a set of functions furnishing the ordinate z and local inclination angle θ as a function of the local position s along the riser (see Fig. 4). For a constant angle riser, for instance, it results:

$$\theta = \arctan\left(\frac{Z}{X}\right) \tag{43}$$

$$s_t = \left(X^2 + Z^2\right)^{1/2} \tag{44}$$

$$z = s\,\sin\theta \tag{45}$$

The definition of geometry for a catenary riser can be seen in (Baliño *et al.*, 2010).

2.2.4 Choke valve

In normal operation in petroleum production systems the choke valve controls the flow, allowing a production compatible with the reservoir characteristics. In severe slugging, it is acknowledged that choking can stabilize the flow by increasing the back pressure (Schmidt, 1977; Taitel, 1986). For low pressures, typical of air-water laboratory systems, the valve operates in subcritical condition; in this case, the flow depends on the pressure difference across the valve (see Fig. 6. The model considers a valve characteristic based on the homogeneous flow model:

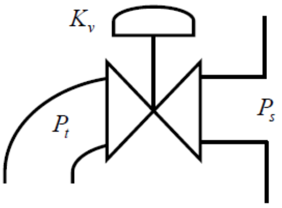


Figure 6: Definition of variables at the choke valve.

$$P_t - P_s = K_v \, \frac{1}{2} \, \rho_{m\,t} \, j_t \, |j_t| \tag{46}$$

$$\rho_{m\,t} = \alpha_t \, \rho_{g\,t} + (1 - \alpha_t) \, \rho_l \tag{47}$$

where j_t, α_t, $\rho_{m\,t}$, $\rho_{g\,t}$ are respectively the total superficial velocity, void fraction, mixture density and gas density, all of them evaluated at the top of the riser, P_s is the separator pressure and K_v is the valve constant.

Another expression for the pressure drop across the valve was used in (Jansen *et al.*, 1996):

$$P_t - P_s = C \, j_{l\,t} \, |j_{l\,t}| \tag{48}$$

where $j_{l\,t}$ is the liquid superficial velocity at the top of the riser and C is a dimensional valve constant. According to this relationship, it is neglected the contribution of the gas phase to pressure drop.

2.3 Coupling between pipeline and riser

Assuming the same flow passage area for the pipeline and riser, the pressure and superficial velocities at the bottom of the riser are continuous:

$$P\,(s = 0,\, t) = P_b\,(t) \tag{49}$$

$$j_g\,(s = 0,\, t) = j_{g\,b}\,(t) \tag{50}$$

$$j_l\,(s = 0,\, t) = j_{l\,b}\,(t) \tag{51}$$

The boundary condition for the void fraction can be obtained from Eq. (23) evaluated at the bottom of the riser:

$$\alpha\,(s = 0,\, t) = \alpha_b\,(t) = \frac{j_{g\,b}}{C_{d\,b}\, j_b + U_{d\,b}} \tag{52}$$

Figure 7 shows the state variables and the coupling between the subsystems. State variables for the pipeline are the average gas pressure, void fraction and position of the liquid accumulation front, while for the riser they are the local pressure, void fraction and total superficial velocity. The pipeline imposes the superficial velocities for the gas and liquid phases at the bottom of the riser, while the riser imposes the pressure to the pipeline; these variables are the boundary conditions for the corresponding subsystems. Additional boundary conditions are the liquid volumetric flow rate and the gas mass flow rate at the pipeline, as well as the gas lift mass flow rate and separation pressure at the riser.

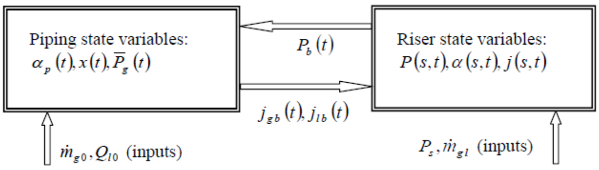


Figure 7: Coupling between subsystems.

As initial conditions, the stationary conditions were chosen, this is, the solution of the system of equations obtained after setting equal to zero the time derivatives.

3. DISCRETIZATION AND NUMERICAL IMPLEMENTATION

The system of equations corresponding to the stationary state, as well as the system of dynamic equations, were discretized and numerically implemented using the software MATLAB (Magrab *et al.*, 2005).

In the riser a moving grid method was adopted (see Fig. 8), in which node i $(1 \leq i \leq N-1)$ moves with the corresponding gas velocity (red lines), in order to calculate the directional derivatives of Eq. (32). Last node N moves with the liquid velocity if the liquid level falls below the top of the riser $(s_u < s_t)$, or remains fixed at position s_t otherwise. The time step Δt^{k+1} is chosen as the time step such that the characteristic propagated from the $N-1$ th node intersects the position s_u at time $t^k + \Delta t^{k+1}$ if the liquid level falls below the top level in the riser, or as the time step such that the characteristic propagated from the $N-1$ th node intersects position s_t otherwise. Values at time t^k are interpolated in order to calculate the directional derivatives of Eq. (35), corresponding to the liquid velocity (blue lines). An implicit scheme was used, with a predictor-corrector method for treatment of the nonlinearities. Details of the procedure can be seen in (Baliño *et al.*, 2010).

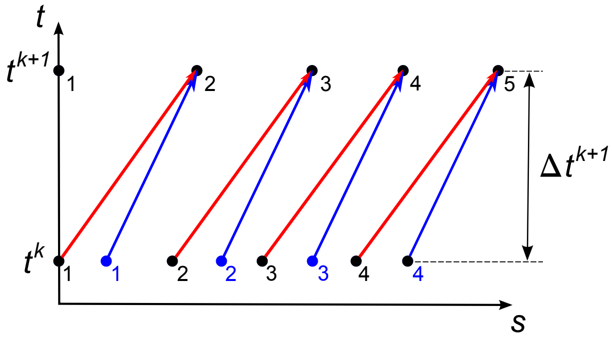


Figure 8: Discretization along the characteristic directions.

4. SIMULATIONS

In this Section, simulations corresponding to experimental data for a vertical riser are shown, considering the effects of the choke valve and gas lift injection at the bottom of the riser.

After a nodalization study, the riser was discretized in $N = 21$ nodes. Input flow variables are defined in terms of the superficial velocities $j_{g\,0}$ and $j_{l\,0}$ at standard conditions (pressure $P_0 = 1.013\,bara$, temperature $T_0 = 293\,K$); these superficial velocities are related to the flows as:

$$j_{g\,0} = \frac{R_g\,T_0\,\dot{m}_{g\,0}}{P_0\,A} \tag{53}$$

$$j_{l\,0} = \frac{Q_{l\,0}}{A} \tag{54}$$

4.1 Data from (Taitel *et al.*, 1990)

The following parameters were chosen for a comparison with experimental data of (Taitel *et al.*, 1990): fluid parameters are $\mu_l = 1. \times 10^{-3}\,kg/m/s$, $\mu_g = 1.8 \times 10^{-5}\,kg/m/s$, $\rho_l = 1. \times 10^{3}\,kg/m^3$, $R_g = 287\,m^2/s^2/K$ and $T_g = 293\,K$; pipeline parameters are $L = 9.1\,m$, $L_e = 1.69\,m$ and $\beta = 5\,^o$; riser height is $Z = 3\,m$; common parameters for pipeline and riser are $D = 2.54 \times 10^{-2}\,m$ and $\epsilon = 1.5 \times 10^{-6}\,m$; separation pressure is $P_s = 1.013\,bara$. No choke valve or gas injection was considered.

Figure 9 shows results of a simulation for $j_{g\,0} = 0.063\,m/s$ and $j_{l\,0} = 0.124\,m/s$ for representative variables: gas (j_{g1} and liquid (j_{l1}) superficial velocities at the bottom of the riser, position (x) of the liquid penetration front, void fraction (α_1) and pressure (P_1) at the bottom of the riser, liquid level at the riser (s_u) and void fraction (α_p) and pressure drop ($P_{pe} - P_{pt}$) in the stratified region at the pipeline.

It can be seen that, in this case, the stationary state used as the initial condition is not stable and the system goes to a limit cycle. It can be observed a large variation of superficial velocities, compared to the initial stationary values. Variations in the void fraction at the pipeline are relatively small, supporting the assumption of constant void fraction made in previous models. On the other hand, variations in pressure drop at the pipeline can be large compared to the stationary values, particularly in the blowdown stage.

Many parameters corresponding to the transient can be calculated from Fig. 9. Considering that the slugging cycle begins when the gas passage at the bottom of the riser is blocked, the severe slugging period and times corresponding to different stages described in Section 1. can be calculated. In this case, it can be seen that the liquid level does not remain at the top of the riser. From the simulations, it is also possible to determine the pressure amplitude at the bottom of the riser, the maximum position of the liquid penetration front in the pipeline and the minimum position of the liquid level in the riser.

Table 1 shows a comparison of experimental and model simulated severe slugging periods for different gas and liquid superficial velocities. The periods calculated with the model are in very good agreement with the experimental ones. There are some cases in which the simulation predicts a unstable condition while the experiment reports a stable condition.

With the simulations, it is also possible to obtain the numeric stability curve by keeping constant a value of liquid or gas flow rate and varying the other in fixed increments until passing from one condition (stable or unstable) to another; when this happens, the procedure is repeated with half the increment until achieving convergence. The procedure is laborious and computationally costly. Stability maps generated for catenary risers can be seen in (Baliño *et al.*, 2010) and (Nemoto and Baliño, 2012).

Figure 10 shows the numerically generated stability map for the conditions corresponding the experimental conditions of (Taitel *et al.*, 1990). In the same figure the experimental data points are shown; these unstable data points were classified as "unstable fall" ($s_u < s_t$ in the transient) or "unstable no fall" ($s_u = s_t$ always in the transient), based on a visual observation. It can be seen that the numerically generated stability curve includes all the unstable data points. Data points for which there is a discrepancy between experiment and stability simulation prediction are located close to the stability curve; for these points pressure amplitudes and liquid penetration lengths are small and liquid level in the riser remains at the top or close to it, making difficult to differentiate between the severe slugging instability condition and the fluctuations associated to the intermittent flow based only on a visualization. Besides, the experimental determination of the stability curve requires a very careful control of variables such as separator pressure and input flows; cases 5 and 7 in Table 1, for instance, show a large discrepancy in experimental period for almost the same values of superficial velocities, indicating that other variables were not kept constant in the experiment.

Table 1: Comparison with experimental results (Taitel *et al.*, 1990).

	Experiment			Simulation	
Case	$j_{g\,0}$ (m/s)	$j_{l\,0}$ (m/s)	T_{exp} (s)	T_{sim} (s)	error (%)
1	0.063	0.124	24	25.4	5.8
2	0.064	0.209	20	19.5	−2.5
3	0.123	0.183	15	15.5	3.3
4	0.124	0.212	14	14.6	4.3
5	0.062	0.679	6	7.36	22.7
6	0.063	0.367	13	13.3	2.3
7	0.063	0.679	9	7.35	−1.8
8	0.064	0.535	10	9.27	−7.3
9	0.065	0.226	19	18.7	−1.6
10	0.122	0.374	11	10.8	−1.8
11	0.123	0.621	8	7.36	−8.0
12	0.126	0.228	13	14.0	7.7
13	0.187	0.226	11	11.9	8.2
14	0.188	0.466	8	8.22	2.8
15	0.188	0.502	7	7.86	12.3
16	0.19	0.312	10	10.1	1.0
17	0.058	0.705	steady	7.19	NA
18	0.063	0.698	steady	7.21	NA
19	0.122	0.730	steady	6.45	NA
20	0.126	0.673	steady	14.0	NA
21	0.126	0.085	steady	18.2	NA
22	0.184	0.127	steady	4.24	NA
23	0.185	0.161	steady	12.3	NA
24	0.187	0.551	steady	7.36	NA
25	0.188	0.755	steady	steady	NA
26	0.19	0.685	steady	6.24	NA
27	0.313	0.433	steady	14.0	NA
28	0.314	0.347	steady	steady	NA
29	0.319	0.614	steady	steady	NA
30	0.321	0.744	steady	steady	NA
31	0.43	0.604	steady	steady	NA
32	0.433	0.701	steady	steady	NA

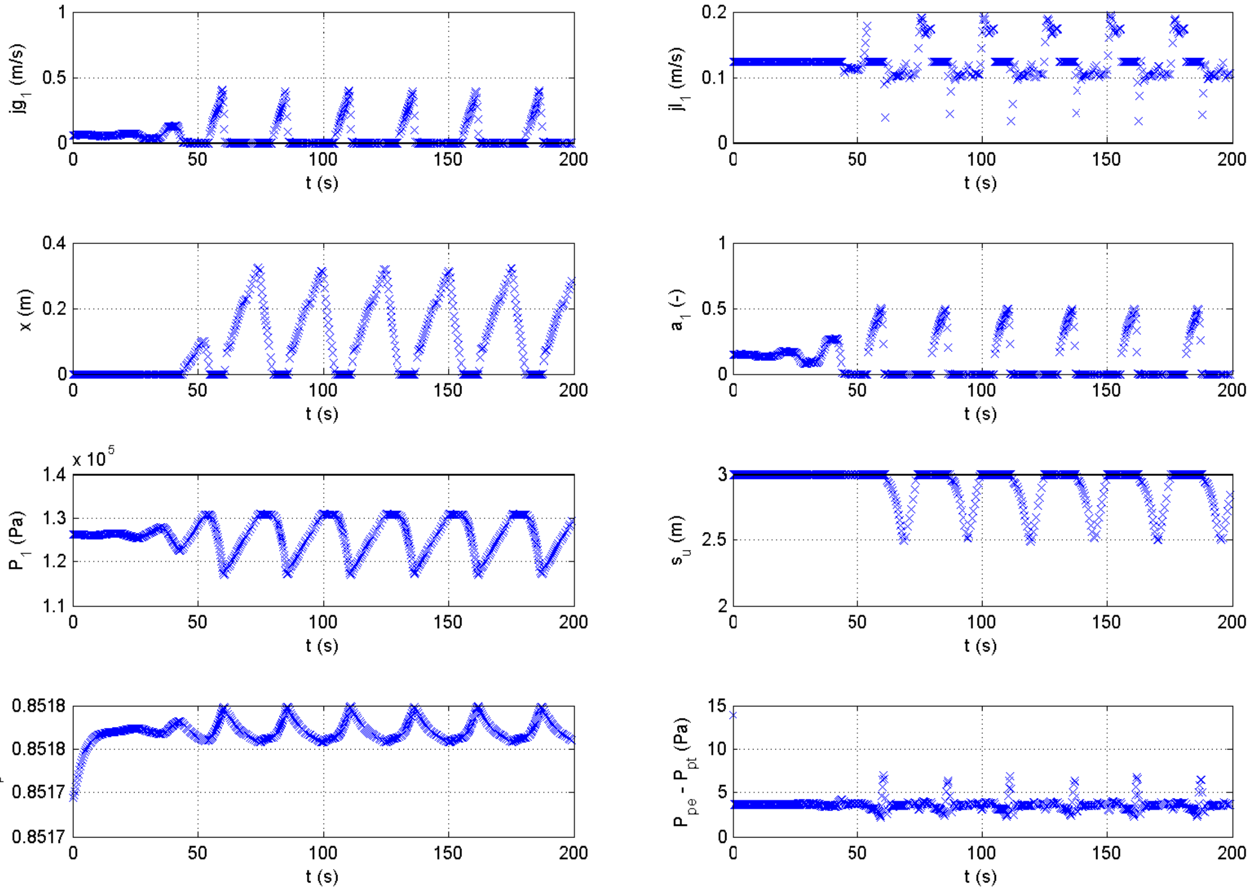


Figure 9: Simulation results for $j_{g\,0} = 0.063\,m/s$ and $j_{l\,0} = 0.124\,m/s$ (case 1, Table 1).

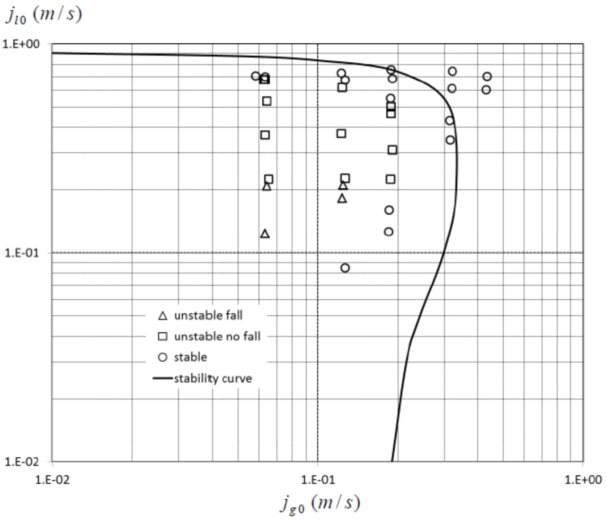


Figure 10: Numerically generated stability map and data from (Taitel *et al.*, 1990).

4.2 Data from (Jansen *et al.*, 1996)

In (Jansen *et al.*, 1996) it was used the same test facility as in (Taitel *et al.*, 1990). The simulations parameters were the same, except for $L_e = 10\,m$ and $\beta = 1\,^o$.

A choke valve with $C = 1.2 \times 10^5\,Pa\,s^2/m^2$, see Eq. (48), was introduced at the top of the riser in order to study the influence of choking on severe slugging; no gas injection was considered.

Table 2 shows a comparison of experimental and model simulated severe slugging periods for different gas and liquid superficial velocities. As in Section 4.1, the experimental and simulated periods are in very good agreement, with some cases with discrepancies in the prediction of the stability condition.

Table 2: Comparison with experimental results, choke valve, $C = 1.2 \times 10^5\,Pa\,s^2/m^2$ (Jansen *et al.*, 1996).

	Experiment			Simulation	
Case	$j_{g\,0}$ (m/s)	$j_{l\,0}$ (m/s)	T_{exp} (s)	T_{sim} (s)	error (%)
1	0.0753	0.0959	46.6	54.9	17.8
2	0.1147	0.0949	37.6	43.7	16.2
3	0.1739	0.0959	31.8	32.8	3.1
4	0.0781	0.0497	47.5	56.2	18.3
5	0.1181	0.0487	38.5	43.7	13.5
6	0.0809	0.1704	45	50.4	12.0
7	0.1209	0.1693	39.9	42.4	6.3
8	0.1713	0.0497	31.5	31.8	1.0
9	0.1209	0.2365	steady	43.0	NA
10	0.1734	0.2354	steady	36.7	NA
11	0.2493	0.2386	steady	30.6	NA
12	0.1698	0.1693	steady	35.5	NA
13	0.2474	0.1704	steady	28.5	NA
14	0.2502	0.0959	steady	25.0	NA
15	0.251	0.0497	steady	22.0	NA

In another experimental campaign, a constant gas superficial velocity at standard condition $j_{g0\,gl} =$

$0.091\,m/s$ was injected at the bottom of the riser in order to study the influence of gas lift on severe slugging; no choke valve was considered. The gas superficial velocity and the corresponding gas mass flow are related by:

$$j_{g0\,gl} = \frac{R_g\,T_0\,\dot{m}_{g\,l}}{P_0\,A} \tag{55}$$

Table 3 shows a comparison of experimental and model simulated severe slugging periods for different gas and liquid superficial velocities. Again, the agreement for experimental unstable data points is very good.

Table 3: Comparison with experimental results, gas injection, $j_{g0\,gl} = 0.091\,m/s$ (Jansen *et al.*, 1996).

	Experiment			Simulation	
Case	$j_{g\,0}$ (m/s)	$j_{l\,0}$ (m/s)	T_{exp} (s)	T_{sim} (s)	error (%)
1	0.0808	0.2528	21.5	29.8	38.6
2	0.1153	0.2549	18.9	23.8	25.9
3	0.1702	0.2571	14.9	18.1	21.5
4	0.2515	0.2582	13.4	13.2	−1.5
5	0.0791	0.152	24.1	36.9	53.1
6	0.1147	0.152	20.5	27.3	33.2
7	0.3125	0.1542	10.8	11.0	1.9
8	0.1695	0.1013	16.8	19.4	15.5
9	0.248	0.0949	13.2	13.9	5.3
10	0.1129	0.0487	27	28.0	3.7
11	0.1737	0.0476	18.8	19.7	4.8
12	0.3215	0.0916	9.5	11.1	16.8
13	0.2489	0.0465	14	13.6	−2.9
14	0.366	0.1552	steady	9.82	NA
15	0.4115	0.1552	steady	8.96	NA
16	0.369	0.0981	steady	9.41	NA
17	0.3141	0.0444	steady	10.4	NA

5. CONCLUSIONS

A mathematical model and numerical simulations corresponding to severe slugging in air-water pipeline-riser systems, are presented. The model is an improvement of the one previously published by the author (Baliño *et al.*, 2010), including inertial effects. Inertial effects are taken into account by using the rigid water-hammer approximation, which was numerically implemented without increasing substantially the complexity of the model.

A comparison is made with experimental results published in literature for vertical risers including the effect of a choke valve at the top and gas injection at the bottom of the riser, showing very good results for slugging cycles and stability maps.

ACKNOWLEDGEMENTS

This work was supported by *Petróleo Brasileiro S. A.* (Petrobras). The financial support of *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) to attend IMAACA 2012 is deeply acknowledged.

REFERENCES

Baliño, J.L., Burr, K.P. and Nemoto, R.H., 2010. "Modeling and simulation of severe slugging in air-water pipeline-riser systems". *International Journal of Multiphase Flow*, Vol. 36, pp. 643–660.

Bendiksen, K.H., 1984. "An experimental investigation of the motion of long bubbles in inclined tubes". *International Journal of Multiphase Flow*, Vol. 10, No. 4, pp. 467–483.

Chaudhry, M.H., 1987. *Applied Hydraulic Transients*. Van Nostrand Reinhold, New York.

Chen, N.H., 1979. "An explicit equation for friction factor in pipe". *Ind. Engng. Chem. Fundam.*, Vol. 18, pp. 296–297.

Chexal, B., Lellouche, G., Horrowitz, J. and Healer, J., 1992. "A void fraction correlation for generalized applications". *Progress in nuclear energy*, Vol. 27, No. 4, pp. 255–295.

Jansen, F.E., Shohan, O. and Taitel, Y., 1996. "The elimination of severe slugging - experiments and modeling". *International Journal of Multiphase Flow*, Vol. 22, No. 6, pp. 1055–1072.

Kokal, S.L. and Stanislav, J.F., 1989. "An experimental study of two-phase flow in slightly inclined pipes - i. flow patterns". *Chem. Engng. Sci.*, Vol. 44, pp. 665–679.

Magrab, E.B., Azarm, S., Belachandran, B., Duncan, J.H., Herold, K.H. and Walsh, G.C., 2005. *An Engineer 's Guide to MATLAB*. Pearson Prentice Hall.

Masella, J.M., Tran, Q.H., Ferre, D. and Pauchon, C., 1998. "Transient simulation of two-phase flows in pipes". *International Journal of Multiphase Flow*, Vol. 24, pp. 739–755.

Nemoto, R.H. and Baliño, J.L., 2012. "Modeling and simulation of severe slugging with mass transfer effects". *International Journal of Multiphase Flow*, Vol. 40, pp. 144–157.

Sarica, C. and Shoham, O., 1991. "A simplified transient model for pipeline-riser systems". *Chemical Engineering Science*, Vol. 46, No. 9, pp. 2167–2179.

Schmidt, Z., 1977. *Experimental study of two-phase slug flow in a pipeline-riser system.* Ph.D. thesis, The University of Tulsa, Tulsa.

Taitel, Y., 1986. "Stability of severe slugging". *International Journal of Multiphase Flow*, Vol. 12, No. 2, pp. 203–217.

Taitel, Y. and Dukler, A.E., 1976. "A model for predicting flow regime transitions in horizontal and near horizontal gás-liquid flow". *AIChE Journal*, Vol. 22, No. 1, pp. 47–55.

Taitel, Y., Vierkand, S., Shoham, O. and Brill, J.P., 1990. "Severe slugging in a riser system: experiments and modeling". *International Journal of Multiphase Flow*, Vol. 16, No. 1, pp. 57–68.

Wordsworth, C., Das, I., Loh, W.L., McNulty, G., Lima, P.C. and Barbuto, F., 1998. *Multiphase Flow Behavior in a Catenary Shaped Riser.* CALtec Report No.: CR 6820, California.

AUTHOR BIOGRAPHY

Jorge Luis Baliño was born in Buenos Aires, Argentina. He graduated from *Instituto Balseiro*, Argentina (Nuclear Engineering, 1983, PhD in Nuclear Engineering, 1991). He worked for Techint S.A. (1983-1984), *Centro Atómico Bariloche* and *Instituto Balseiro* (1985-2000) in Argentina, *Instituto de Pesquisas Energéticas e Nucleares* (2001-2003) at São Paulo, Brazil. Since 2004 he is Professor at *Universidade de São Paulo*. His research interests are Fluid Dynamics, Heat Transfer and Multiphase Flow.

ROBUST MODEL MATCHING CONTROL APPLIED TO A CRANE

Eduardo Luiz Lozano de Campos[(a)], **Fabrizio Leonardi**[(b)]

[(a)(b)] Centro Universitário da FEI, São Bernardo do Campo, Brazil

[(a)] llcampos@uol.com.br, [(b)]fabrizio@fei.edu.br

ABSTRACT

This paper discusses the robust closed loop control design subject to parametric uncertainties applied to a crane during a maneuver. Usually crane trajectories are generated by formulating a minimum time optimal control in open loop. However, the optimality of the solution is not maintained due to variations in the plant over time. This work proposes the use of a model matching structure to reduce the problems related to model uncertainties thus trying to preserve the trajectory optimality. The robust compensator minimizes explicitly the matching error between the real plant and the reference plant. In this application the main uncertain parameter is the pendulum length and plays the role of the load lifting. To illustrate the application experiments were done using a lab scale equipment. The results observed are very close to those obtained from numerical simulation.

Keywords: Robust Control, Optimal Control, Model Matching

1. INTRODUCTION

Cranes are common in various industry segments for transporting loads. These systems are considered efficient for their safety in transportation, mechanical robustness and reliability of loading and unloading tasks.

Even a crane manually operated need to be controlled accordingly but one should not expect great performance since human action can be inaccurate and not always the best path will be produced, which can generate, for example, swings that may endanger the operation, products, equipment, etc. When an automated system is used, it is reasonable to expect for better performance. In this case, the system is responsible for controlling all the variables subjected to the physical constraints, seeking minimum time, lower power consumption and minimal oscillation. (Puglia, Leonardi and Ackermann, 2011; Da Cruz and Leonardi, 2012).

As proposed by Sorensen, Singhose and Dickerson (2007), the control schemes of cranes may be grouped into three categories: time-optimal control, command shaping and feedback control. His own publication can be considered in the category of command shaping with a forward action used to determine the appropriate command signal in order to reduce the swing during a maneuver. The command shaping approach was also used by Lee and Choi (2001), who developed a way to determine the trajectory of the crane based on Lyapunov stability theorem. Another example is the work of Chen, Hein and Wörn (2007), where it is proposed an open-loop control with the trajectory defined based on the principle of acceleration compensation. In the work of Lau and Pao (2001), he discusses about the equivalence of minimum time optimal control and command shaping for flexible systems.

Normally the minimum time problem is treated in open loop, but due to modeling errors and disturbances is necessary to use a closed loop control strategy to maintain the optimal trajectory with a certain precision. An example of this is the work of Hičár and Ritók (2006) which uses the pole allocation method by means of the Ackermann formulae to provide a robust control to a crane.

Different closed loop control structures, such as IMC, multiloop, model matching, etc., are somehow equivalent to each other in the sense that it is usually possible to represent the same control law on different topologies. However, the choice of a particular structure can make easy the analysis, design or even its implementation. This paper proposes the use of a model matching structure to perform the closed loop control law for a crane which must robustly maintain the optimal trajectory of both minimum time and minimum control effort. The optimal maneuver begins with the crane at rest and carries the load at a fixed distance, reaching the destination also at rest. The optimal trajectory does not consider the lift during the maneuver, and therefore, if this occurs, the trajectory is no longer optimal. The purpose of using the model matching is precisely for overcome that problem. The controller must make a closed loop to behave as the plant that was used to obtain the optimal trajectory, that is, without lifting. Thus, the controller must be robust to variations in crane cable length.

The optimal trajectory was generated in the similar way as in Puglia, Leonardi and Ackermann (2011). It takes into account physical constraints of the system, such as the maximum control effort. Notice that the design of the robust model matching controller must also take into consideration those constraints.

To easily incorporate constraints in the robust controller design, we have chosen to conduct the project in the time domain by means of a parametric optimization of the controller coefficients.

A cart-pendulum lab system was used to illustrate the proposed approach.

2. METHOD

The proposed methodology is based on a model matching control structure and its design is discussed in the sequence. This structure is used to robustly maintain the optimal trajectory of a crane during a maneuver.

2.1. Model Matching

Consider the standard closed loop control system diagram in the Fig. 1, where $x(\text{t})$ is the reference signal, $y(\text{t})$ is the controlled output, and $d(\text{t})$ an auxiliary exogenous signal. Respectively, $F(\text{s})$ and $P(\text{s})$ represent the transfer functions of the controller and the plant nominal model.

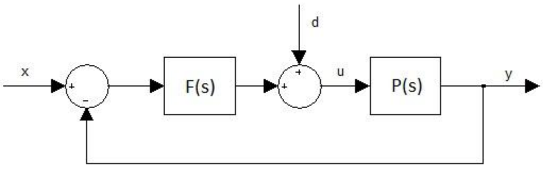


Figure 1: Closed Loop Control.

In general, the goal of the control system is to have $y(\text{t})$ closely following $x(\text{t})$. If the nominal plant model is known and the inverse of its transfer function $M(s)=1/P(s)$ exists, one can use the input $d(\text{t})$ as a feed forward action, as shown in Fig. 2.

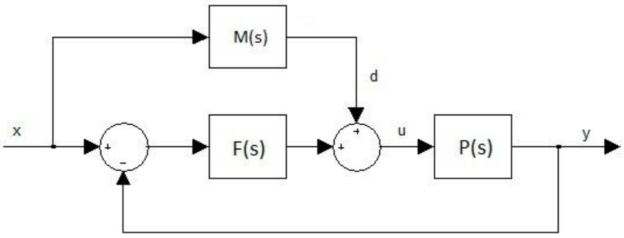


Figure 2: Closed Loop with Feed Forward Action.

In the absence of plant modeling errors the control system is reduced to an open loop, since $y(\text{t}) = x(\text{t})$. In the presence of modeling errors the controller $F(\text{s})$ need to compensate for the difference $x(\text{t})\text{-}y(\text{t})$ by correcting the value of $u(\text{t})$. Even disregarding the modeling errors, often the inverse of the plant model may have issues for a practical application. For those cases $M(\text{s})$ can be taken as an approximation of the $1/P(\text{s})$, and thus the controller $F(\text{s})$ will probably be required to compensate for higher deviations. When the plant exhibits constant gain at low frequency it is common to use a static $N(\text{s})$.

In an optimal control problem, typically, both the optimal trajectory $y = y^*(\text{t})$ and the optimal control $u = u^*(\text{t})$, are available. If it is necessary to keep the optimal trajectory with a closed loop control we can use the structure of the Fig. 2. However, one can achieve feed forward compensation without the use of $M(\text{s})$ explicitly, by simply using $d(\text{t}) = u^*(\text{t})$ and $x(\text{t}) = y^*(\text{t})$. In this scenario, the feed forward control action diagram can be redrawn in the equivalent form of the Fig. 3 where the plant model appears explicitly $N(s)=P(s)$.

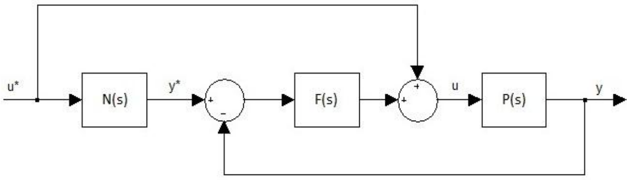


Figure 3: Closed Loop Optimal Control.

In a more general scenario the diagram of Fig. 3 can be used even when $N(s)\neq P(s)$ and may be applied in an attempt to make the closed loop transfer function $T(s)=y(s)/x(s)$ aproach $N(\text{s})$. In fact, this appeal is more easily seen by drawing the diagram in its equivalent form of Fig. 4.

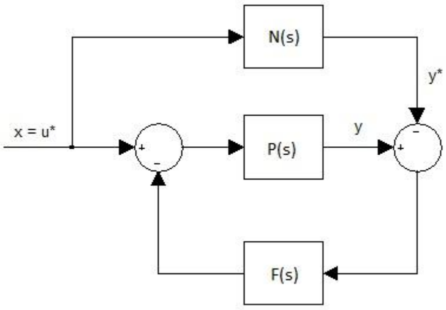


Figure 4: Model Matching Structure.

The structure in Fig. 4 is known in the technical literature and has been used in some applications like the one by Jonckheere (1999) for controlling a crippled aircraft. Note that the error signal in this diagram is, in fact, the difference between the response of the plant $P(\text{s})$ and the response of the reference model $N(\text{s})$. If the controller $F(\text{s})$ can make this error small enough, then the response of the plant is about the same response of the reference model. This effect has been called approximate model matching.

The model matching control structure is used in this work in order to make the plant response following an optimal trajectory. The reference model $N(\text{s})$ is the model used to generate the optimal trajectory and $P(\text{s})$ represents all real plants. The differences between $N(\text{s})$ and $P(\text{s})$ may be due to variations in the plant over time which may even be deliberate. In this work this feature is used to compensate eventual load hoisting and lowering during a maneuver.

2.2. Mechanical Model

A scheme of the cart-pendulum system used is shown in Fig. 5, where m is the load mass, x the cart position and φ the load angle.

The equations of motion describing the dynamics of the cart-pendulum model can be, for instance, derived using the Newton-Euler formalism as described in Schiehlen (1997).

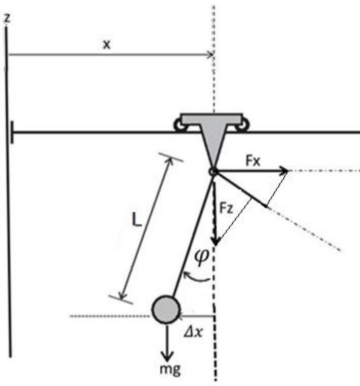


Figure 5: Cart-pendulum Scheme.

The resulting nonlinear model can be presented in the form of the following differential equation.

$$L\,\frac{d^2(\varphi(t))}{dt^2} + g\,sen(\varphi(t)) = \frac{d^2(x(t))}{dt^2}\cos(\varphi(t)) \qquad (1)$$

where g is the gravity acceleration and L de pendulum length.

Considering that the kinematics of the cart can be imposed arbitrarily, we define the manipulated variable

$$u(t) = x(t) \qquad (2)$$

That allows also defining

$$a(t) = \frac{d^2(x(t))}{dt^2}\ . \qquad (3)$$

In handling anti-oscillatory problems, it is expected that the maximum oscillation angle be small. This condition leads to the approximations $sen\varphi \approx \varphi$ and $\cos\varphi \approx 1$. These approximations simplify the equations of motion to

$$L\,\frac{d^2(\varphi(t))}{dt^2} + g\ \varphi(t) = \frac{d^2(x(t))}{dt^2} \qquad (4)$$

Mapping it to the Laplace domain and taking null initial conditions, one obtains the transfer function

$$\frac{Y(s)}{X(s)} = \frac{s^2}{Ls^2 + g} \qquad (5)$$

where

$$y(t) = \varphi(t) \qquad (6)$$

to be consistent with the notation used in section 2.1.

The model does not incorporate the Coulomb friction. However, it can be easily included as an additional torque in the equation (1). In such cases, the design of optimal control signal should take this into account or the closed loop control must be robust in the presence of this modeling error.

2.3. Optimality

The model matching control system proposed in this work should be able to closely keep the solution of the optimization problem proposed by Puglia, Leonardi and Ackermann (2011). This problem is defined by the objective function (7) and the constraints (8). That is, should minimize the sum of the absolute control a (acceleration) in each sampling time (1,.., n), subjected to the plant dynamics N(s), initial state w(t_0) and final state w(t_f) of the maneuver, and the limits $max|v|$ of the control effort. Besides, the optimal control a^*(t) and the optimal trajectory y^*(t) generated by Puglia, Leonardi and Ackermann (2011) also includes the time minimizing in the same optimization problem. Notice that the overall acceleration is minimized but the designer could add penalties related to each sampling time or even limit each value as an explicit constraint.

The optimal signal is used in the model matching control structure of Fig. 4 which is supposed to maintain y(t) close to the y^*(t).

$$\min_a J_1 = |a_1| + |a_2| + \cdots + |a_n| \qquad (7)$$

$$N(s),\ w(t_0),\ w(t_f),\ \max|a| \qquad (8)$$

We define here internal optimality of the model matching problem as the property of a(t), the input signal of the real Plant P(s), be an optimum control signal in the sense of the equations (7) and (8). That is, the value of J_2 obtained by the solution of the optimization problem of equations (9) and (10) must be equal or less than J_1. Besides, since the problem also includes time minimizing, the control signal must also be optimal in this sense.

$$\min_a J_2 = |a_1| + |a_2| + \cdots + |a_n| \qquad (9)$$

$$P(s),\ w(t_0),\ w(t_f),\ \max|a| \qquad (10)$$

Note that typically P(s) ≠ N(s) and in general a(t) may not meet the requirements of internal optimality. That is, from the viewpoint of the real plant, the model matching structure can not preserve the optimality produced by Puglia, Leonardi and Ackermann (2011) since he applied the control to a plant P(s) = N(s).

Thus, it is defined here what we call external optimality. Since the transfer function y(s) to x(s) can match N(s) with a prescribed precision, so if we apply the optimal control signal

$$u^*(t) = \int\int^{t} a^*(t)\,dt \qquad (11)$$

to x(t), it sees a plant very close to N(s). Thus, the optimality of the original solution is preserved in an approximate way. With this external point of view, both kinematics constraints and control minimizing are aproximately preserved.

2.4. Control Effort

In the frequency domain it can be stated that the model matching problem is to find a compensator $F(s)$ such that the absolute value of transfer function $x(j\omega)$ to $e(j\omega) = y(j\omega)-y^*(j\omega)$ is below a certain prescribed value in the largest possible range of frequencies (Leonardi, 2006).

For the system shown in Fig. 4, the following equations apply

$$y(s) = P(s)\left(1+F(s)P(s)\right)^{-1}\left(1+F(s)N(s)\right)x(s) \tag{12}$$

and

$$u(s) = \left(1+F(s)P(s)\right)^{-1}\left(1+F(s)N(s)\right)x(s) \tag{13}$$

Since it is considered here that $N(s)$ is stable, stability of the system is determined solely by the closed loop which contains $P(s)$ and $F(s)$, wich is implicit assured once the performance is achieved.

Consider $\alpha > 0$ (typically $\alpha << 1$), a given number that expresses the desired precision associated to the model matching error in a certain range of frequencies, so that

$$|e(j\omega)| \,/\, |x(j\omega)| \leq \alpha \,. \tag{14}$$

To ensure model matching we have the following sufficient condition,

$$|F(j\omega)P(j\omega)| \geq \frac{|P(j\omega)-N(j\omega)|}{\alpha} \tag{15}$$

obtained from (12) to the typical case in which the loop gain and precision are respectively large, that is, for $|F(j\omega)\ P(j\omega)| >> 1$ and $\alpha << 1$. This condition shows that the loop gain increases with either increasing the distance between P and N as the inverse of α.

From equation (13) is immediate that

$$u(s)-x(s) = F(s)\left[1+P(s)F(s)\right]^{-1}\left[N(s)-P(s)\right]x(s) \tag{16}$$

Under approximate conditions given by $\alpha << 1$ and $|F(j\omega)\ P(j\omega)| >> 1$, then (16) leads to

$$u(j\omega)-x(j\omega) \cong P(j\omega)^{-1}\left[N(j\omega)-P(j\omega)\right]x(j\omega)\ , \tag{17}$$

From equation (17) it follows immediately that

$$\frac{|u(j\omega)-x(j\omega)|}{|x(j\omega)|} \cong \left| P^{-1}(j\omega)\left[N(j\omega)-P(j\omega)\right]\right| \tag{18}$$

This last equation shows that the relative increase in control effort is approximately the same as the relative difference between the plant and the reference model. Therefore, reference models that are distant from the plant model requires a high control effort to be followed. This is consistent with the condition (15) wich shows that the greater the distance between the plant and the reference model, the greater is the loop gain to ensure model matching.

2.5. Robustness

The modeling errors may be uncertainties in transfer function of the plant. However, classical margins of stability alone are unable to reveal the degree of robustness of a system because, even systems with favorable margins as [90º, ∞ dB], may have its corresponding Nyquist diagram close to -1 + 0j, and therefore, are not robust (Da Cruz, 1996).

Model uncertainties can be classified as structured and unstructured. Unstructured uncertainties are usually associated to unmodeled parts of the plant that are frequency dependent such as negletec dynamics. The structured uncertainties are associated with parametric uncertainties such as the one in this work.

The main parametric uncertainty of the plant model (5) is the distance L from the load to the cart. in fact this uncertainty is intentional and represents the changes of L over time required during the maneuver. If the performance of the control system is robust to this variation $L_{\min} \leq L \leq L_{\max}$, the external optimality is approximately preserved.

This paper proposes to use the parametric optimization of the controller in order to include constraints, beyond the problem of robustness to the variation of L. The optimization is performed on a time range suitable for the maneuver and uses as a reference the optimal signal obtained by Puglia, Leonardi and Ackermann (2011), but using the model matching control structure of Fig. 3 or Fig. 4.

To incorporate the problem of robustness in the formulation of parametric optimization, the objective function

$$J = \int_0^{tf} [e_1^2(t) + e_2^2(t) + \ldots + e_m^2(t)]dt \tag{19}$$

includes the sum of square of the matching error between each of the considered m real plant $P_i(s)$, $i = \{1,.., m\}$, and the reference plant $N(s)$. That is, the transfer function of each $P_i(s)$ is considered here equal to $N(s)$ but with a distinct value of L in the range of $L_{\min} \leq L \leq L_{\max}$.

Since in this parametric optimization problem we can easily add several type of constraints, any physical restriction of the problem are conveniently incorporated. It should be noted that limiting the control effort has been considered in generating the signals $a^*(t)$ and $y^*(t)$, however they were generated for the open loop and in the absence of modeling errors. By using the model matching to keep the trajectory $y^*(t)$, there is no guarantee that the acceleration limit is still respected. Therefore, this restriction should be used again, now in the design of closed loop controller.

2.6. Controller Selection

Since the nominal plant model is open loop stable a wide class of controllers are candidate for the optimization problem. Although a purely proportional controller can stabilize the closed loop, it causes excessive noise amplification since the transfer function of the plant is just proper.

To allow attenuation for high frequency noise, the transfer function of the controller must be strictly proper. Thus, a good choice for the controller to this problem may be one with purely integral action.

The controller design methodology for robust model matching has been applied and the controller obtained has the following transfer function.

$$F(s) = \frac{12}{s} \tag{20}$$

3. RESULTS

The results presented in this section refer to the application of the robust model matching control using the controller of the equation (20) to a lab scale plant. The optimal signal was generated for a maneuver problem in minimum time and minimum control effort in a manner similar to the one that was obtained by Puglia, Leonardi and Ackermann (2011). The optimal control signal takes the pendulum from rest to the other end, away 0.25m from the start, also arriving at rest.

3.1. Testing Apparatus

The pendulum of the testing apparatus (see Fig. 6) consists of a 0.215kg mass connected to the cart by a rod. The mass can be fixed on the rod at different distances from the cart.

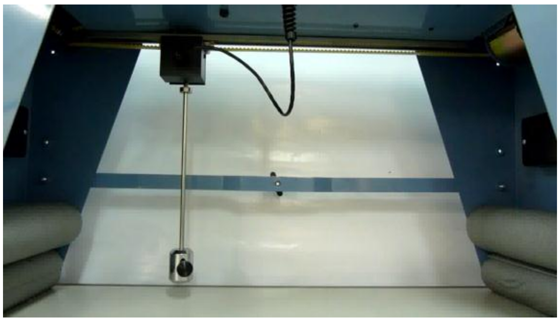


Figure 6: Bytronic Lab Equipment.

The cart driver has a built in position control with tachometer compensation, as indicated in Fig. 7. Since that control system is quite precise over the frequency range that matters in this problem, its dynamics can be reasonably neglected and thus the cart position is considered the manipulated variable as it was admitted in the methodology section.

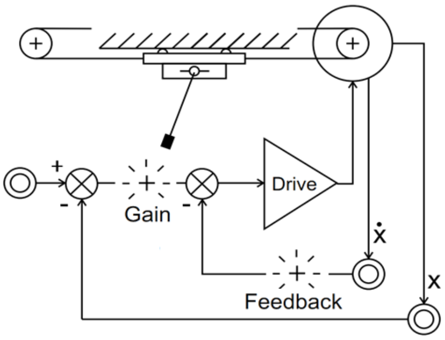


Figure 7: Schematic Diagram of the Equipment.

The robust controller used in this application has only an integral action. He was selected to be extremely simple and yet provide good robustness to parametric design, which in fact can be verified by the experimental results.

To implement the compensator it was used the Real-Time Target Windows™ (Mathworks, 2012) operating at a sampling frequency of 1KHz, the same rate used in the generation of the signals $a^*(t)$ and $y^*(t)$.

3.2. Experimental Results

The controller design was done by considering the variation in the length of the pendulum in the range $0.15 \text{ m} \le \text{L} \le 0.25 \text{ m}$, and being 0.25m the nominal reference value. That is, the reference plant is

$$\frac{Y(s)}{X(s)} = \frac{s^2}{0,25s^2 + 9.81} \tag{21}$$

The plots of Fig. 8 show the performance in time domain obtained with the controller. The maneuver begins at $t = 0$ *s* and ends at $tf = 1.3$ *s*.

The value of the length L in the reference model was kept fixed at $L = 0.25$ m and the L values of real equipment were changed within the range considered. Fig. 8 shows the worst case where the real length $L_{\text{real}} = 0.15$ m is most distant from the nominal $L = 0.25$ m. The figure contains two sets of plots. The first (a) shows the optimum position u^* (red) and the experimental value of u position (blue). Note that since the position is the manipulated variable of the control system, deviations of u from the value u^* represent the extra effort the control system needs to spend to perform the match.

In the second set of plots (b) it is shown the experimental angle (blue) that is expected to be close to the reference angle (red). To complete the analysis, it is also shown the angle behavior if the system is operating without control (black).

It can be seen that the integral controller, in fact, provides a robust performance to the control system. This is true since both the optimal control signal (position) and the optimal trajectory (angle), are close to their reference values even for huge variations of L.

In particular, note that in all plots the final cart position and the final angular position were achieved with small error.

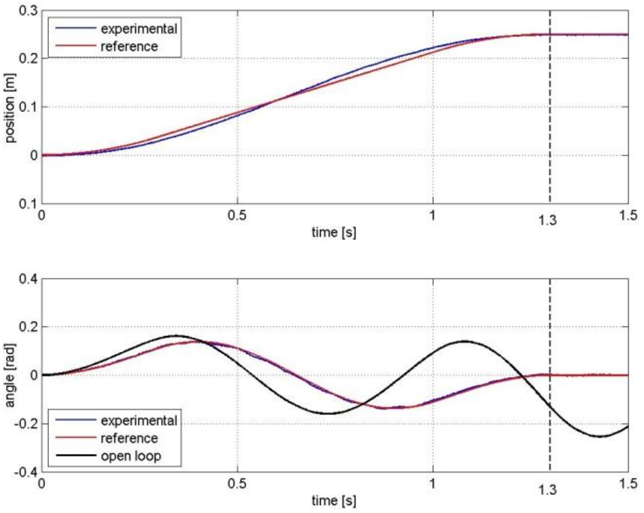


Figure 8: Performance of the Integral Controller for $L_{\text{real}} = 0.15$m.

The system begins at rest and reaches its destination in 1.3 s. It is clear from the plots that system remains at rest after 1.3 s

For the sake of comparison we also designed the controller using the H_2 mixed sensitivity. The performance perceived is similar, but the resulting compensator is of order 5th with two extra resonances to the loop which might be undesirable.

4. CONCLUSIONS

This paper discusses the use of a model matching structure for closed loop control of the optimal trajectory of a crane. We discussed the design of the compensator to reduce the problems related to parametric uncertainties of the plant, thus preserving the optimality of the initial solution. The project was conducted by means of parametric optimization of the compensator and the objective function includes the matching error of a number of plants with different values of the pendulum length.

The practical results were obtained applying the methodology to a cart-pendulum lab scale equipment. It was found that the designed controller gives robust performance even for a large parametric Plant variation as expected during the design.

The overall methodology was developed for a dedicated application. However, it might be applied for linear systems with few uncertain parameters and for modeling errors below 100%.

As a proposal for extending this work the following investigations are suggested.

This study did not investigate the problem of rejecting external disturbances caused, for instance, by wind. Adding a disturbance input to the model matching structure we obtain a control law that has two degrees of freedom. That is, it is possible to tune the robust controller in order to balance between the requirement to follow the reference signal and to reject the disturbance. A current research is examining this issue to propose a design methodology that takes this into consideration.

This study also did not investigate the problem of sensitivity of the response in the face of measurement noise and possible offset in its calibration. The mentioned above research is also investigating how the control structure can be altered to minimize this effect, mainly the one from residual offset. The investigation also includes the definition of artificial measurable variables and how the optimal control trajectory of the crane needs to be modified to do so. Preliminary results show that it is possible to find necessary and sufficient conditions for this mapping.

REFERENCES

Chen, S. J., Hein, B., Wörn, H. 2007. Swing Attenuation of Suspended Objects Transported by Robot Manipulator Using Acceleration Compensation. *Proceedings of the International Conference on Intelligent Robots and Systems*, (San Diego, USA). 2919–2924.

Da Cruz, J. J., Leonardi , F. 2012. Minimum-time anti-swing motion planning of cranes using linear programming, *Optimal Control Applications & Methods*.

Da Cruz, J. J., 2006. Controle Robusto Multivariavel. Editora da USP, 163p

Hičár, M.; Ritók, J., 2006. *Robust Crane Control*, Acta Polytechnica Hungarica. 3(2).

Jonckheere, A. E. 1999. Propulsion Control of Crippled Aircraft by Hoo Model Matching. *IEEE Transactions on Control Systems Technology*, v.7,

Lau, M.A. and Pao, L.Y., 2001. Comparison of Input Shaping and Time-Optimal Control of Flexible Structures. *Proceedings of the American Control Conference*. Arlington, VA, 25-27.

Lee, H.-H. and Choi, S.-G., 2001, A Model-Based Anti-Swing Control of Overhead Cranes with High Hoisting Speeds. *Proceedings of the 2001 IEEE International Conference on Robotics & Automation*. Seoul, Korea. 21-26.

Leonardi, F.; Da Cruz, J. J., Bittar, A., 2006. Multivariable Robust Control with Time Domain Specifications: Servo and Regulator Problems. *Electrical Engineering. Berlin*, (16), 279-289.

Mathworks Inc. 2012. Real-Time Windows Target™ User's Guide.

Puglia, L. V., Leonardi, F., Ackermann, M. 2011. Using Linear Programming for the Optimal Control of a Cart-pendulum System. *Proceedings of the International Conference on Integrated Modeling and Analysis in Applied Control and Automation*. pp. 200–205 (Rome, Italy).

Schiehlen, W. 1997. Multibody System Dynamics: Roots and Perspectives". *Kluwer Academic Publishers*.

Sorensen, K., Singhose, W., Dickerson, S. 2007. A Controller Enabling Precise Positioning and Sway Reduction in Bridge and Gantry Cranes. *Control Engineering Practice*, 15(7), 825–837.

FAULT DIAGNOSIS IN NCS UNDER COMMUNICATION CONSTRAINTS: A QUADROTOR HELICOPTER APPLICATION

K. Chabir, M. A. Sid, D. Sauter

Nancy University, CRAN -CNRS UMR 7039 BP239, 54506 Vandoeuvre Cedex , France.

{karim.chabir, Mohamed-Amine.sid, dominique.sauter}@cran.uhp-nancy.fr

ABSTRACT

In this paper a method for fault diagnosis in quadrotor helicopter is presented. The proposed approach is composed of two stages. The first stage is the modelling of the system attitude dynamics taking into account the induced communication constraints. Then a robust fault detection and evaluation scheme is proposed using a post-filter designed under a particular design objective. This approach is compared with previous results based on the standard Kalman filter and gives better results for sensors fault diagnosis.

Keywords: Networked control systems, Diagnosis, generation residual, evaluation residual, Quadrator helicopter.

1. INTRODUCTION

Unmanned Aerial Vehicles (UAV) are receiving a great deal of attention during the last few years due to their high performance in several applications such as search and critical missions, surveillance tasks, geographic studies and various military and security applications. As an example of UAV systems, the quadrotor helicopter is relatively a simple, affordable and easy to fly system and thus it has been widely used to develop, implement and test-fly methods in control, fault diagnosis, fault tolerant control as well as multi-agent based technologies in formation flight. Navigation and guidance algorithms may be embedded on the onboard flight microcomputer/microcontroller or with the interference by a ground wirelesses/wired controller in others cases. In our setting the quadrotor is controlled over real time communication network with time-varying delays and therefore is considered as a Networked control system (NCS). In general NCS is composed of a large number of interconnected devices (system nodes) that exchange data through communication network. Recent research on NCS has received considerably attention in the automatic control community (Zhang, et al., 01; Tipsuwan and Chow, 03; Huajing et al., 07; Mirkin and Palmor, 05; Hespanha, et al., 07; Richard, 03). The major focus of the research activities are on system performance analysis regarding the technical properties of the network and on the controller design schemes for NCS.

However, the introduction of communication networks in the control loops makes the analysis and synthesis of NCS complex. There are several network-induced effects that arise when dealing with the NCS, such as time-delays (Niculescu, 00; Nilsson, et al., 98; Pan, et al., 06; Schollig, et al., 07; Dritsas, and Tzes, 07; Yi, et al., 06; Zhang, et al., 05; Behrooz, et al.,08), packet losses (Xiong, and Lam, 06; Sahebsara, et al., 07; Yu, et al., 04; Li, et al., 06) and quantization problems (Goodwin , et al., 04; Montestruque and Antsaklis, 07; Frank and Ding, 97). Because of the inherent complexity of such systems, the control issues of NCS have attracted attention of many researchers, particulary taking into account network-induced effects. Typical application of these systems ranges over various fields, such as automotive, mobile robotics, advanced aircraft.

The fault diagnosis has become an important subject in modern control theory (Frank and Ding, 97; Gertler, 98; Isermann, 06; Stoustrup, and Zhou, 08; Basseville, and Nikiforov, 93). The study of fault detection (FD) in NCS is a new research topic, which gained more attention in the past years. For instance, the results in (Sauter and Boukhobza, 06; Sauter, et al., 07 ; Llanos, et al., 07; Chabir, et al., 08; Chabir, et al., 09; Chabir, et al., 10; Al-Salami, et al., 08) are focus on networked-induced delays. The problem studied in (Zhang, et al., 04; Wang, et al., 06) is the analysis and design of FD systems in case of missing measurements. The fault detectability and isolability in NCS have been discussed in (Sauter, et al., 09; Chabir, et al., 09). The fault tolerant structure is studied in (Ding and Zhang, 07 ; Patton, et al., 07; Kambhampati, et al., 06).

Delays are known to degrade drastically the performances of a control systems, for this reason, many works aimed at reducucing the effects of induced network delays on NCS (Tipsuwan and Chow, 03; Yu, et al., 04; Li, et al., 06; Goodwin , et al., 04). In the majority of the studies concerning the stabilization of networked control systems, the delay is considered to be constant (Schollig, et al., 07) or bounded (Dritsas, and Tzes, 07), but the dynamics of the delay corresponding to the characterization of the network is not taken into account in general. Thus, it is interesting to estimate the delay, in order to generate an optimal control, as well as algorithms of faults detection that take into account the network characteristics. One approach is to consider the delay as a Markov chain (Yi, et al., 06; Zhang, et al., 05). In order to predict such a random delay, artificial neural networks can be used (Zhang, et al., 05). However, such a methods are considered to be not suitable for real time implementation (Behrooz, et al.,08).

The objective in this study is diagnosis of quadrotor attitude sensors fault under variable transmission delay. First, attitude dynamics model taking into account the variables transmission delay is presented. Then we propose a robust residual generation and evaluation scheme using a post-filter that verify a particular design objective. This approach is compared with previous results based on the standard Kalman filter and gives better results for sensors fault diagnosis.

The rest of the paper is organized as follows. In section 2, the quadrotor helicopter attitude dynamics is modeled and then controlled using LQR approach. Section 3, presents the first main result of this paper, which is related to the modeling of networked control systems. Finally, section 4 we present our second main result concerned with the residual generation and evaluation using an adaptive threshold. The paper is concluded in Section 5

2. DESCRIPTION OF QUADROTOR HELICOPTER DYNAMICS

The mini-helicopter under study has four fixed-pitch rotors mounted at the four ends of a simple cross frame Figure 1. The attitude is modeled with the Euler-angle representation which provides an easier expression for the linearized model. Moreover the Euler-angle representation is more intuitive. The inertial measurement unit model is given with the quaternion representation of the attitude. This choice is govern by the implementation of the attitude observer that will be easier with the quaternion parameterization of the attitude.

Figure 1: The quadrator mini-helicopte.

2.1. Quadrotor model

The quadrotor is a small aerial vehicle controlled by the rotational speed of four blades, driven by four electric motors (3) A quadrotor is considered a VTOL vehicle (Vertical Take Off and Landing) able to hover. Two frames are considered to describe the dynamic equations: the inertial frame $N(x_n, y_n, z_n)$ and the body frame $B(x_b, y_b, z_b)$ attached to the UAV with its origin at the centre of mass of the vehicle.

The quadrotor orientation can be parameterized by three rotation angles with respect to frame N: yaw (ψ), pitch (θ) and roll (Φ). $\omega \in \Re^3$ is the angular velocity of the quadrotor relative to N expressed in B. The quadrotor is controlled by independently varying the rotational speed ω_{mi}, $i = 1{:}4$, of each electric motor. The force f_i and the relative torque Q_i produced by motor i are proportional to ω_{mi}.

$$f_i = b\,\omega_{mi}^2 \tag{1}$$

$$Q_i = k\,\omega_{mi}^2 \tag{2}$$

where $k > 0$, $b > 0$ are two parameters depending on the density of air, the radius, the shape, the pitch angle of the blade and other factors.

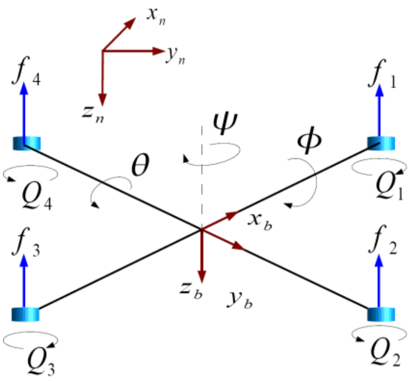


Figure 2: Quadrator mini-helicopte configuration: the inertial frame $N(x_n, y_n, z_n)$ and the body frame $B(x_b, y_b, z_b)$.

The three torques that constitute the control vector for the quadrotor are expressed in frame B as:

$$\tau_a^{\phi} = d\left(f_2 - f_4\right) \tag{3a}$$

$$\tau_a^{\theta} = d\left(f_1 - f_3\right) \tag{3b}$$

$$\tau_a^{\psi} = Q_1 + Q_3 - Q_2 - Q_4 \tag{3c}$$

d represents the distance from one rotor to the centre of mass of the quadrotor. From Newton-Euler approach, the kinematics and dynamic equations of the quadrotor are:

$$\left(\dot{\phi}, \dot{\theta}, \dot{\psi}\right)^T = M\,\omega \tag{4}$$

$$I_f\,\dot{\omega} = -\omega \times I_f\,\omega + \tau_a + G_a \tag{5}$$

where $I_f \in \Re^{3\times3}$ represents the constant inertial matrix expressed in B (supposed to be $I_f = \mathrm{diag}(I_{fx}, I_{fy}, I_{fz})$) and $\times$ in (5) denotes the cross product. Matrix M is defined with

$$M = \begin{pmatrix} 1 & \tan\theta\sin\phi & \tan\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi \\ 0 & \dfrac{\sin\phi}{\cos\theta} & \dfrac{\cos\phi}{\cos\theta} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \tag{6}$$

The gyroscopic torques G_a due to the combination of the rotation of the quadrotor and the four rotors, are modeled as:

$$G_a = \sum_{i=1}^{4} I_r (\omega \times e_z)(-1)^{i+1} \omega_{mi} \quad (7)$$

Ir is the inertia of the so-called rotor (composed of the motor rotor itself, of the shape and of the gears).

A linear control law that stabilizes around hover conditions the system described by the non-linear model (4) and (5) is established. Note that nonlinearities are second order, therefore it is reasonable to consider a linear approximation. From (4) and (5) and for hover condition ($\phi \approx \theta \approx \psi \approx 0$), it comes:

$$(\phi', \theta', \psi')^T = (\omega_1, \omega_2, \omega_3)^T \quad (8)$$

Then the dynamical model is obtained in terms of Euler angles

$$\phi'' = \theta' \psi' \left(\frac{I_{fy} - I_{fz}}{I_{fx}} \right) + \frac{\tau_a^{\phi}}{I_{fx}} \quad (9a)$$

$$\theta'' = \phi' \psi' \left(\frac{I_{fz} - I_{fx}}{I_{fy}} \right) + \frac{\tau_a^{\theta}}{I_{fy}} \quad (9b)$$

$$\psi'' = \phi' \theta' \left(\frac{I_{fx} - I_{fy}}{I_{fz}} \right) + \frac{\tau_a^{\psi}}{I_{fz}} \quad (9c)$$

The gyroscopic torques G_a are not considered for the design of the control law. However, they will be considered in simulations in order to analyze the robustness features.

2.2. Attitude control

In this section, the linearized model of (4) and (5) is first derived. Then a control law is briefly summarized. Note that this paper is not dedicated to the determination of a particular control law (see for instance (Guerrero-Castellanos, *et al.*, 07; Tayebi and McGilvray, 06). Therefore a LQ controller is implemented. In the third subsection, the estimation of the network induced delay with an Extended Kalman Filter is considered. This technique is then applied to the Network controlled quadrotor. Define the state variable:

$$x^T = (\phi, \phi', \theta, \theta', \psi, \psi')^T \quad (10)$$

The system (9) linearization around the hover conditions is:

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (11)$$

where

$$A = \begin{pmatrix} A_0 & 0 & 0 \\ 0 & A_0 & 0 \\ 0 & 0 & A_0 \end{pmatrix}, \quad B = \begin{pmatrix} B_x & 0 & 0 \\ 0 & B_y & 0 \\ 0 & 0 & B_z \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\text{and } B_i = \begin{pmatrix} 0 \\ 1/I_{fi} \end{pmatrix} \quad (12)$$

The attitude stabilization problem is to drive the quadrotor attitude from any initial condition to a desired constant orientation and maintain it thereafter. As a consequence, the angular velocity vector is also brought to zero and remains null once the desired attitude is reached, $x \to 0$, $t \to \infty$. The discrete linear controller is given by

$$u(kh) = -Lx(kh) \quad (13)$$

and the plant is modeled as:

$$x_{k+1} = \Phi x_k + \Gamma u_k \quad (14)$$

satisfy the system dynamics constraints:

$$J = \sum_{k=0}^{N-1} \left[x_k^T Q_d x_k + u_k^T R_d u_k \right] + x_N^T Q_0 x_N \quad (15)$$

where:

$$\Phi = e^{Ah}, \qquad \Gamma = \int_{kh}^{(k+1)h} e^{As} B ds,$$

$$Q_d = \int_{kh}^{(k+1)h} \Phi^T(s) Q \Phi(s) ds, \quad \text{and}$$

$$R_d = \int_{kh}^{(k+1)h} \left(\Gamma^T(s) Q \Gamma(s) + R \right) ds \quad (16)$$

where matrices Q_d, R_d and Q_0 are symmetric and positive definite. Furthermore, the following assumptions are done.

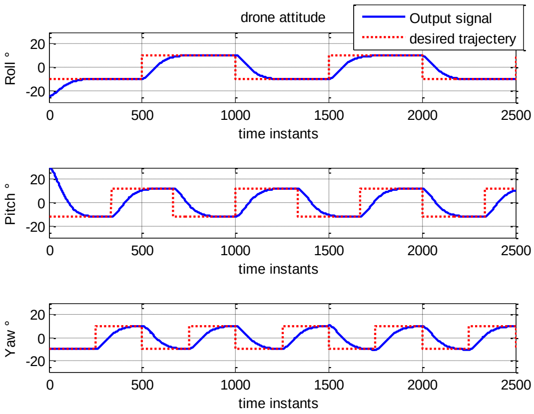


Figure 3: Quadrotor attitude (ϕ, θ, ψ) and reference.

Assumption 1: *The full state vector is available (angles and angular velocities). In practice, these variable states are obtained by merging the measurements of rate gyros, accelerometers and magnetometers using a dedicated attitude observer* (Guerrero-Castellanos, et al., 07).

Assumption 2: *A periodic sampling is used.*

Assumption 3: *The control signals remain constant between two updates.*

Proposition 1: *Consider the quadrotor rotational dynamics described by (9). Then, the discrete control* $u =$ *defined by:*

$$\begin{aligned} u(kh) &= \left[\tau_a^{\phi}(kh) \quad \tau_a^{\theta}(kh) \quad \tau_a^{\psi}(kh) \right]^T \\ &= -Lx(kh) \end{aligned} \quad (17)$$

which satisfies (14) while minimizing (15) locally stabilizes the quadrotor at $x = 0$.

Remark 1: The weighting matrices Q_d and R_d are chosen in order to obtain a suitable transient response, while only feasible control signals are applied to the actuators. Then for a sampling time h = 0.01s the matrix gain is.

$$L = \begin{bmatrix} 0.0352 & 0 & 0 \\ 0.0284 & 0 & 0 \\ 0 & 0.0352 & 0 \\ 0 & 0.0284 & 0 \\ 0 & 0 & 0.0352 \\ 0 & 0 & 0.0284 \end{bmatrix}^T \qquad (18)$$

Here we simply present some results of the drone attitude simulation with a variable step response (Figure 3) and the LQ controller signal (Figure 4)

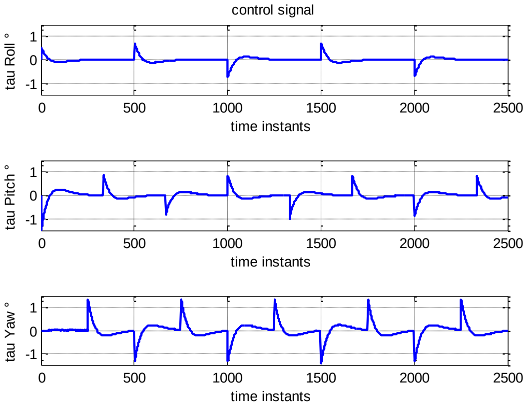


Figure 4: Control signal.

3. NCS MODEL AND TRANSFORMATION

Induced time delays in networked controlled systems can become a source of instability and degradation of control performance (Yi, *et al.*, 06; Zhang, *et al.*, 05; Behrooz, *et al.*, 08; Xiong and Lam, 06; Sahebsara, *el al.*, 07). When the system is controlled over a network, we have to take into account the sensor to controller delays and controller to actuator delays. Note that delays, in general, cannot be considered as constant and known. Network induced delays may vary, depending on the network traffic, medium access protocol and the hardware.

Assumption 4. *For data acquisition it is supposed that the sensor is time-driven and the sampling period is denoted by h. Both the controller and the actuator are event-driven. We mean that calculation of the new control or actuator signal is started as soon as the new control or actuator information arrives as illustrated in Fig. 5*

Assumption 5. *The unknown time-varying network induced delay at time step k is denoted by* τ_k *and* $\tau_k = \tau_k^{sc} + \tau_k^{ca}$ *is smaller than one sampling period* $\tau_k \le h$, τ_k^{sc} *and* τ_k^{ca} *are the sensor-to-controller delay and the controller-to-actuator delay, respectively. There is no packet dropout in the networks.*

Thus, the control input (zero-order hold assumed) over a sampling interval $[kh, (k+1)h]$ is:

$$u_t \ddot{=} \begin{cases} u_{k-1}, t \in \; kh, \quad kh + \tau_k \\ u_k, t \in \left[kh + \tau_k, \quad k+1 \; h\right] \end{cases} \qquad (19)$$

Let us first assume that the residual generation and evaluation algorithms are executed instantaneously at every sampling period k. Based on this assumption, if the control input is kept constant over each sampling interval h, and if we consider that fault inputs present slow dynamics, the discrete time system can be described by:

$$\begin{cases} x_{k+1} = \Phi x_k + \Gamma_{0,\tau_k} u_k + \Gamma_{1,\tau_k} u_{k-1} \\ y_k = C x_k \end{cases} \qquad (20)$$

where

$$\Gamma_{0,\tau_k} = \int_0^{h-\tau_k} e^{As} B ds, \Gamma_{1,\tau_k} = \int_{h-\tau_k}^{h} e^{As} B ds \qquad (21)$$

Like $\Gamma = \int_0^h e^{As} \mathrm{Bds} = \Gamma_{0,\tau_k} + \Gamma_{1,\tau_k}$ thus

$$\Gamma_{0,\tau_k} = \Gamma - \Gamma_{1,\tau_k} \qquad (22)$$

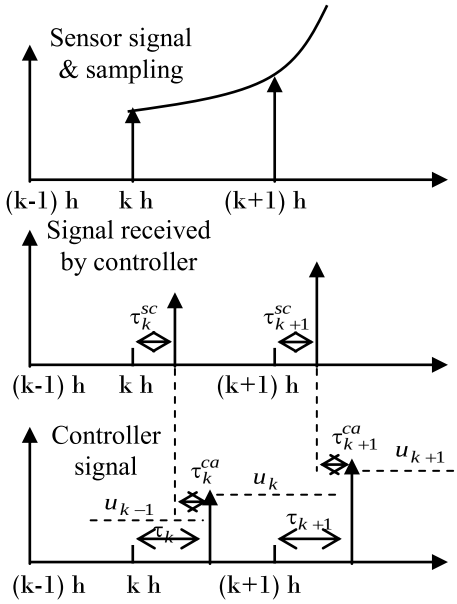


Figure 5: Timing diagram for data communication.

According to the property of definite integral, If we introduce the control increment $\Delta u_k = u_k - u_{k-1}$, let the plant (20) with unknown disturbance vector, d_k and fault vector, f_k which must be detected, is described by:

$$\begin{cases} x_{k+1} = \Phi x_k + \Gamma u_k + \Gamma_{1,\tau_k} \Delta u_k + \Xi_x d_k + \Psi_x f_k \\ y_k = C x_k + \Xi_y d_k + \Psi_y f_k \end{cases} \qquad (23)$$

where $f_k \in \mathbb{R}^q$ the fault vector and $d_t \in \mathbb{R}^q$ the noise vector.

Suppose that the matrix A is called diagonalizable if P is invertible

$$\mathrm{A} = \mathrm{P}\Lambda\mathrm{P}^{-1} = \mathrm{P}diag\left(\lambda_1,\cdots,\lambda_n\right)\mathrm{P}^{-1} \tag{24}$$

where $\lambda_1,\cdots,\lambda_n$ are eigenvalues of matrix A, then there is:

$$\begin{aligned} e^{At} &= I + At + \cdots + \frac{1}{n!}A^n t^n \\ &= PP^{-1} + P\Lambda P^{-1}t + \cdots + \frac{1}{n!}\left(P\Lambda P^{-1}\right)^n t^n \\ &= P\left(I + \Lambda t + \cdots + \frac{1}{n!}\Lambda^n t^n\right)P^{-1} \\ &= Pe^{\Lambda t}P^{-1} \end{aligned} \tag{25}$$

Then, with (23), we have that:

$$\begin{aligned} \Gamma_{1,\tau_k}\Delta u_k &= \int_{h\text{-}\tau_k}^{h} Pe^{\Lambda t}P^{-1}Bds\ \Delta u_k \\ &= P\int_{h\text{-}\tau_k}^{h} e^{\Lambda t}dsP^{-1}B\Delta u_k \end{aligned}$$

$$= P\begin{bmatrix} \int_{h\text{-}\tau_k}^{h} e^{\lambda_1 t}ds & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \int_{h\text{-}\tau_k}^{h} e^{\lambda_n t}ds \end{bmatrix}P^{-1}B\Delta u_k \tag{26}$$

$$\begin{aligned} &= P\begin{bmatrix} \frac{1}{\lambda_1}e^{\lambda_1 h} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n}e^{\lambda_n h} \end{bmatrix}P^{-1}B\Delta u_k \\ &-P\begin{bmatrix} \frac{1}{\lambda_1}e^{\lambda_1(h\text{-}\tau_k)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n}e^{\lambda_n(h\text{-}\tau_k)} \end{bmatrix}P^{-1}B\Delta u_k \end{aligned} \tag{27}$$

$$\begin{aligned} &= \Gamma_\Delta\Delta u_k - Pdiag\left(\frac{1}{\lambda_{11}},\ \frac{1}{\lambda_2},\ \cdots\ \frac{1}{\lambda_n}\right) \\ &\times diag\left(\beta_k\right)\begin{bmatrix} e^{\lambda_1(h\text{-}\tau_k)} \\ e^{\lambda_2(h\text{-}\tau_k)} \\ \vdots \\ e^{\lambda_n(h\text{-}\tau_k)} \end{bmatrix} \end{aligned} \tag{28}$$

$$= \Gamma_\Delta\Delta u_k - \Gamma_{\Delta,k}\begin{bmatrix} e^{\lambda_1(h\text{-}\tau_k)} \\ e^{\lambda_2(h\text{-}\tau_k)} \\ \vdots \\ e^{\lambda_n(h\text{-}\tau_k)} \end{bmatrix} = \Gamma_\Delta\Delta u_k - \Gamma_{\Delta,k}d_{\tau_k} \tag{29}$$

where

$$\beta_k = \begin{bmatrix} \beta_k^1 & \beta_k^2 & \cdots & \beta_k^n \end{bmatrix}^T = P^{-1}B\Delta u_k \in \mathbb{R}^{n,1},$$

$$diag\left(\beta_k\right) = \begin{bmatrix} \beta_k^1 & 0 & \cdots & 0 \\ 0 & \beta_k^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta_k^n \end{bmatrix} = diag\left(P^{-1}B\Delta u_k\right),$$

$$\Gamma_\Delta = P\begin{bmatrix} \frac{1}{\lambda_1}e^{\lambda_1 h} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n}e^{\lambda_n h} \end{bmatrix}P^{-1}B \qquad \text{and}$$

$$\Gamma_{\Delta,k} = Pdiag\left(\frac{1}{\lambda_{11}},\ \frac{1}{\lambda_2},\ \cdots\ \frac{1}{\lambda_n}\right)diag\left(\beta_k\right)$$

According to (29) the model of Eq. (23) can also be rewritten as :

$$\begin{cases} x_{k+1} = \Phi x_k + \Gamma u_k + \Gamma_\Delta\Delta u_k \\ \qquad -\Gamma_{\Delta,k}d_{\tau_k} + \Xi_x d_k + \Psi_x f_k \\ y_k = Cx_k + \Xi_y d_k + \Psi_y f_k \end{cases} \tag{30}$$

By definition, $\Gamma_k^a = \begin{bmatrix}\Gamma & \Gamma_\Delta\end{bmatrix}$, $\mathrm{u}_k^a = \begin{bmatrix} u_k \\ \Delta u_k \end{bmatrix}$, $\Xi_{x,k}^a = \begin{bmatrix}\Xi_x & -\Gamma_{\Delta,k}\end{bmatrix}$ $\Xi_y^a = \begin{bmatrix}\Xi_y & 0\end{bmatrix}$ and $d_k^a = \begin{bmatrix} d_k \\ d_{\tau,k} \end{bmatrix}$ we get:

$$\begin{cases} x_{k+1} = \Phi x_k + \Gamma_k^a \mathrm{u}_k^a + \Xi_{x,k}^a d_k^a + \Psi_x f_k \\ y_k = Cx_k + \Xi_y^a d_k^a + \Psi_y f_k \end{cases} \tag{31}$$

Assuring the robustness of residual generators in practical situations against inevitable unknown input disturbances is commonly recognized as the main design problem for FDI schemes. In the case of structured types of uncertainties, the current literature proposes a variety of solutions for achieving robustness, see for instance (Chen and Patton, 99; Ding, 08). In the next section FDI is revisited, considering network effects.

Model based Fault detection relies on the generation of a residual which must be sensitive to failures and able to distinguish failures from other unknown disturbance inputs. The design must ensure that residuals are closed to zero in fault free situations while clearly deviating from zero in the presence of faults. In a first attempt, the idea is to consider a residual generator based on the state observer.

$$\begin{cases} \hat{x}_{k+1} = \Phi\hat{x}_k + \Gamma_k^a u_k^a + L\left(y_k - \hat{y}_k\right) \\ \hat{y}_k = C\hat{x}_k \end{cases} \tag{32}$$

and the residual generator:

$$r_k = T\left(y_k - \hat{y}_k\right) \tag{33}$$

where T and L are matrices that are designed in order to fulfill fault detection and isolation requirements. From (32) and (33), the estimation error $\varepsilon_k = x_k - \hat{x}_k$ and the output of the filter propagate as:

$$\begin{aligned} \varepsilon_{k+1} = (\Phi - LC)\varepsilon_k + \left(\Xi_{x,k}^a - L\Xi_y^a\right)d_k^a \\ +\left(\Psi_x - L\Psi_y\right)f_k \end{aligned} \tag{34}$$

where $\Phi - LC$ is a stable matrix, and L has to ensure a best estimate of the process states. It results that $\lim\limits_{t\to\infty} \varepsilon_k = 0$, which leads (after z-transformation) to

$$\begin{aligned} r_z = T\left(C\left(zI - \Phi + LC\right)^{-1}\left(\Xi_{x,k}^a - L\Xi_y^a\right) + \Xi_y^a\right)d_k^a \\ +T\left(C\left(zI - \Phi + LC\right)^{-1}\left(\Psi_x - L\Psi_y\right) + \Psi_y\right)f_k \end{aligned} \tag{35}$$

The observer gain matrix L and T are determined such that the following requirements are guaranteed

1. Asymptotic stability under fault free conditions (i.e. $f_k = 0$);
2. Minimization of disturbance effects;
3. Maximization of fault effects;

Perfect fault detection, which means perfect decoupling from unknown inputs with:

$$T\left(C\left(zI - \Phi + LC\right)^{-1}\left(\Xi_{x,k}^a - L\Xi_y^a\right) + \Xi_y^a\right)d_k^a = 0 \tag{36a}$$

$$T\left(C\left(zI - \Phi + LC\right)^{-1}\left(\Psi_x - L\Psi_y\right) + \Psi_y\right)f_k \neq 0 \tag{36b}$$

Actually, there are various approaches (Gertler, 98; Chen and Patton, 99; Frank and Ding, 97; Ding, 08) to determine the gain matrices L and T, but we do not discuss this topic in the paper. If, it is now supposed that the system is controlled over a network, then we have to take into account the sensor to controller delays and controller to actuator delays.

For illustration purpose we consider a simulation of the system described by equations (11). It is supposed that the FD system based on the standard Kalman filtering is connected to the plant via a network.

In the simulations, the network delay is supposed to be Gaussian variable, the fault associated to the first attitude sensor "ϕ : Roll" occurs at time instant k= 1000 and the fault associated to the second attitude sensor "ψ : Yaw" occurs at time k= 1500.

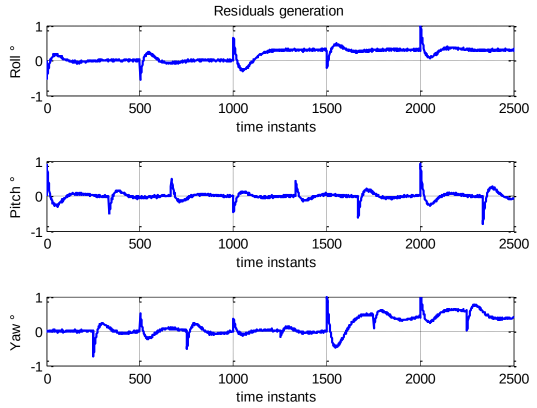


Figure 6: Residuals generation by standard kalman filter (IJAAC).

Result shown before doesn't allow (Fig.6.) to distinguish between the fault and the network variable delay effects. Hence, it appears that the robustness of the fault diagnosis system against network induced delays depend on the amplitude of the unknown term $\Gamma_{\Delta,k} d_{\tau,k}$.

Assuring the robustness of residual generators in practical situations against inevitable unknown input disturbances is commonly recognized as the main design problem for FDI schemes. In the case of structured types of uncertainties, the current literature proposes a variety of solutions for achieving robustness (Chen and Patton, 99; Ding, 08). In the next section FDI is revisited, considering network effects.

4. ROBUST RESIDUAL GENERATION AND EVALUATION

The objective of fault diagnosis is to perform two main decision tasks (Frank and Ding, 97): fault detection, consisting of deciding whether or not a fault has occurred, and fault isolation, consisting of deciding which element of the system has failed. The general procedure comprises the following two steps:

- Residual generation: the process of associating, with the pair model-observation, features that allow evaluating the difference with respect to normal operating conditions.

- Residual evaluation: the process of comparing residuals to some predefined thresholds according to a test and at a stage where symptoms are produced.

This implies designing residuals that are close to zero in fault-free situations while clearly deviating from zero in the presence of faults and that possess the ability to discriminate between all possible modes of faults, which explain the use of the term isolation.

Therefore, the objective here is to design a residual generator similar to the one described by equation (31) which in addition is robust against network delays influence. Several approaches have been proposed in the literature (Wang et al., 06; Sauter and Boukhobza, 06; Chabir, et al., 08)

4.1. Residual generation

A solution of the above mentioned problem towards the design of observer based residual generator will be derived. Under the following assumptions:

$$z_k = \begin{bmatrix} x_k \\ e_k \end{bmatrix} \tag{37}$$

The overall system dynamics, which includes the plant and the residual generator, can be expressed as

$$\begin{cases} z_{k+1} = \bar{A} z_k + \bar{B} \mathrm{u}_k^a + \bar{\Xi}_{k,x} \mathrm{d}_k^a + \bar{\Psi}_x f_k \\ r_k = T\bar{C} z_k + T\,\Xi_y^a d_k^a + T\,\Psi_y f_k \end{cases} \tag{38}$$

where $\bar{A} = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi - K_k C \end{bmatrix}$, $\bar{C} = \begin{bmatrix} 0 & C \end{bmatrix}$, $\bar{B} = \begin{bmatrix} \Gamma_k^a \\ 0 \end{bmatrix}$,

$\bar{\Xi}_{k,x} = \begin{bmatrix} \Xi_{x,k}^a \\ \Xi_{x,k}^a - L\Xi_y^a \end{bmatrix}$, and $\bar{\Psi}_x = \begin{bmatrix} \Psi_x \\ \Psi_x - L\Psi_y \end{bmatrix}$

It is assumed that the plant is mean square stable. Since the observer gain matrix L has no influence on the system in (38). The overall system dynamics (plant + residual generator) is mean square stable.

The post-filter T and the observer gain matrix L are the design parameters for the residual generator. The main objective of the design of the residual generator is to improve the sensitivity of the FD system to faults while keeping robustness against disturbances. Thus, the selection of the design parameters L, T can be formulated as an optimization problem such as:

$$Sup\, J = \sup_{L,T} \frac{\left\| G_z^{rf} \right\|_{-}}{\left\| G_z^{rd} \right\|_{\infty}} \tag{39}$$

where

$$G_z^{rd} = T\bar{C}\left(zI - \bar{A} + L\bar{C}\right)^{-1} \bar{\Xi}_{k,x} + T\,\Xi_y^a \tag{40a}$$

$$G_z^{rf} = T\bar{C}\left(zI - \bar{A} + L\bar{C}\right)^{-1} \bar{\Psi}_x + T\,\Psi_y \tag{40b}$$

4.2. Residual evaluation

The second step of the fault detection procedure is to evaluate the residual. Residual evaluation is an important step of model based FD approach, i.e. see for instance in (Ding, 08). This stage includes a calculation of the residual evaluation function and a determination of detection threshold. The decision for successful fault detection is finally made based on the comparison between the results obtained from the residual evaluation function and the determined threshold.

The following residual evaluation function is proposed :

$$J^e{}_k = \left\| r_k \right\|_{2,N} = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} r_{k-i} \right)^T \left(\frac{1}{N} \sum_{i=1}^{N} r_{k-i} \right)} \tag{31}$$

Where N is the length of the evaluation window. The variance of the residual signal can be expressed as:

$$\sigma_{rk} = \mathrm{E}\left(\left(r_k - \bar{r}_k \right)^T \left(r_k - \bar{r}_k \right) \right) \tag{31}$$

Under the assumption that the unknown input and control input are L_2 - bounded, the following theorem is given:

Theorem 1:

Given system (14) and the constants $\gamma_1 > 0, \gamma_2 > 0$. *The following equation holds true:*

$$\begin{aligned} \sigma_{rk} &= \mathrm{E}\left(\left(r_k - \bar{r}_k \right)^T \left(r_k - \bar{r}_k \right) \right) \\ &< \gamma_1 \sum_{j=0}^{k} \left(v_j^T v_j + \Delta u_j^T \Delta u_j \right) + \gamma_2 \left(v_k^T v_k + \Delta u_k^T \Delta u_k \right) \end{aligned} \tag{31}$$

If there exist $P > 0$ so that:

$$\begin{bmatrix} -P & P\bar{A} & P\bar{B} & \bar{\bar{\Xi}}_{k,x} \\ \bar{A}^T P & -P & 0 & 0 \\ \bar{B}^T P & 0 & -I & 0 \\ \bar{\bar{\Xi}}_{k,x}^T P & 0 & 0 & -I \end{bmatrix} < 0 \tag{31}$$

$$\begin{bmatrix} -P & \bar{C} \\ \bar{C}^T & -\gamma_1 I \end{bmatrix} < 0 \tag{31}$$

$$\begin{bmatrix} -I & \Psi_y \\ \Psi_y^T & -\gamma_2 I \end{bmatrix} < 0 \tag{31}$$

where

$\bar{\bar{\Xi}}_{k,x} = \begin{bmatrix} \bar{\bar{\Xi}}_{x,k}^a \\ \bar{\bar{\Xi}}_{x,k}^a - L\Xi_y^a \end{bmatrix}$, $\bar{\bar{\Xi}}_{x,k}^a = \begin{bmatrix} \Xi_x & -\bar{\Gamma}_{\Delta,k} \end{bmatrix}$, and

$\bar{\Gamma}_{\Delta,k}$ is calculated for $\Delta u = \max\left(\Delta u\right)$.

The proof is similar to the one mentioned in (Al-Salami, et al., 08), hence it is omitted. Note that Δu_k is set to the allowed upper bound of the control input $\max\left(\Delta u\right)$.

The threshold can set as : $J_k^{th} = \sqrt{\alpha_N \beta}$ (31)

Where $\beta = \sup \sigma_{rk}$

$$\beta = \gamma_1 \left(\delta_{d,2} + \sum_{j=0}^{k} \left(\Delta u_j^T \Delta u_j \right) \right) + \gamma_2 \left(\delta_{d,\infty} + \Delta u_k^T \Delta u_k \right)$$

where $\delta_{d,2} \geq \sum_{j=0}^{k} \left(v^T{}_j v_j \right), \delta_{d,\infty} \geq v^T{}_k v_k$.

are the L_2, L_∞ of the unknown input, respectively, and $0 < \alpha_N < 1$ is a constant value depends on the length of the evaluation window N .

The parameters γ_1, γ_2 are some constants which represents the bounds of the variance of the residual signal.

Note that because the residual signal is a white noise process, the threshold will depends on the statistical part of it (which means the variance of residual signal).

After the determination of a threshold, a decision has to performed, if a fault occurs. The Decision logic for the FD system can be defined as follows:

$J_k^e > J_k^{th} \Rightarrow$ fault

$J_k^e \leq J_k^{th} \Rightarrow$ no fault

The threshold $J_{th}(k)$ is adaptive and is influenced from Δu_k , which has to be calculated online.

In the next section simulations are performed in order to validate the results of the proposed residual evaluator.

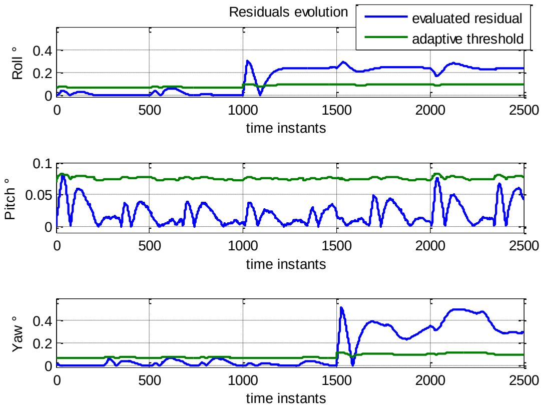


Figure 7: Evaluated residual.

The upper bounds on the unknown inputs are $\delta_{d,2} = 0.15, \delta_{d,\infty} = 0.28$. The length of the evaluation window is set to 50 and α_N is set to 0.3. The parameters of the Threshold (bounds on the variance of residual) are computed as $\gamma_1 = 0.0058, \gamma_2 = 0.05$. The threshold is then to determine (adaptive) on-line during the simulation.

From the result shown (fig. 7.) it is clear that the adaptive threshold allows fault detection and the liklyhood of the false alarm rate is extremely minimized.

5. CONCLUSIONS

In this paper the residual generation and evaluation issue is presented within the framework of networked control systems. The problems, addressed in this paper, are (i) robustness against network delays as well as noise (ii) reducing the false alarm rate. In this context, a quadrotor attitude sensors fault is detected by a post-filter and compared to an adaptive threshold. That considers the variation of control inputs as well as unknown inputs. The problem of threshold design is established in terms of linear matrix inequalities. Validation results show the effectiveness of the obtained results.

REFERENCES

Zhang W, Branicky M. S and Phillips S. M, (2001). Stability of networked control systems. *IEEE Control Systems Magazine*; 21(1):84–99.

Tipsuwan Y. and Chow M.Y., (2003). Control Methodologies in networked control systems. *Control Engineering Practice*, vol. 11, pp. 1099–1111.

Huajing F., Hao Ye, Maiying Z., (2007). Fault diagnosis of networked control systems. *Annual Reviews in Control* 31 55–68.

Mirkin L. and Palmor Z. J.,(2005). Control Issues in Systems with Loop Delays. *in Handbook of Networked and Embedded Control Systems*, 627–648, Birkhauser.

Hespanha J. P., Naghshtabrizi P., and Xu Y.,(2007). A Survey of Recent Results in Networked Control Systems, *Proceedings of IEEE*, vol. 95, pp. 138–162.

Richard J.-P., (2003). Time-delay systems: An overview of some recent advances and open problems", *Automatica*, vol. 39, no. 10, pp. 1667–1694.

Niculescu, S.-I. (2000). Delay Effects on Stability, *A Robust Control Approach*, Springer.

Nilsson J., Bernhardsson B., and Wittenmark B. (1998). Stochastic analysis and control of real-time systems with random time delays, *Automatica*, vol. 34, no. 1, pp. 57-64.

Pan YJ, Marquez HJ, Chen TW (2006). Stabilization of remote control systems with unknown time varying delays by LMI techniques. *International Journal of Control*; 79(9):752–763.

Schollig A., Munz U., Allgower F. (2007). Topology-Dependent Stability of a Network of Dynamical Systems with Communication Delays, *Proceedings of the European Control Conference* ,Kos, Greece, pp. 1197– 1202.

Dritsas L. and Tzes A. (2007). Robust output feedback control of networked systems, *In Proceedings of the European Control Conference*, Kos, Greece, pp. 3939– 3945.

Yi J., Wang Q., Zhao D. and Wen J. T. (2006). BP neural network prediction-based variable-period sampling approach for networked control systems, *Applied Mathematics and Computation*, 2006, pp. 976–988.

Zhang L., Shi Y., Chen T., and Huang B. (2005). A new method for stabilization of networked control systems with random delays, *IEEE Trans. Autom. Control*, Vol. 50, 2005, pp. 1177–1181.

Behrooz R., Amir H.D. M., Naser M. (2008) Real Time Prediction of Time Delays in a Networked Control System ", The 3rd International Symposium on Communications, *Control and Signal Processing (ISCCSP)*, Malta, pp. 1242–1245.

Xiong JL, Lam J. (2006). Stabilization of linear systems over networks with bounded packet loss. *Automatica*; 43(1):80–87.

Sahebsara M, Chen TW, Shah SL. (2007). Optimal H2 filtering in networked control systems with multiple packet dropout. *IEEE Transactions on Automatic Control*; 52(8):1508–1513.

Yu M., Wang L., Chu T., and Hao F. (2004). An LMI approach to networked control systems with data packet dropout and transmission delays", *Journal of Hybrid System*, 3.

Li S., Wang Y., Xia F., Sun Y., and Shou J. (2006). Guaranteed cost control of networked control systems with time-delays and packet losses, *International Journal of wavelets, multiresolution and information processing*.

Goodwin G. C. Haimovich H. Quevedo D. E. and Welsh J. S. (2004). A moving horizon approach to networked control system design, *IEEE Transactions on Automatic Control*, 49(9):1427–1445.

Montestruque LA, Antsaklis PJ. (2007). Static and dynamic quantization in model-based networked control systems. *International Journal of Control*; 80(1):87–101.

Frank P. M. and Ding S. X. (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control*, pages 403-424.

Gertler J. (1998). Fault Detection and Diagnosis in Engineering Systems. *Marcel Dekker, Inc*.

Chen J. and Patton R. J. (1999). Robust Model-based Fault Diagnosis for Dynamic Systems*, Kluwer Academic Publishers*.

Isermann R. (2006). Fault diagnosis systems. *Springer-Verlag*.

Ding S. X. (2008). Model-based fault diagnosis techniques-design schemes, algorithms and tools. *Springer, Berlin*.

Stoustrup J. and Zhou K.(2008). Robustness Issues in Fault Diagnosis and Fault Tolerant Control. *Journal of Control Sciences and Engineering*, Vol. 2008, Article ID 251973, 2 pages.

Basseville M., Nikiforov I.V. (1993). Detection of Abrupt Changes, *Theory and Application*. Prentice-Hall, 1993.

Sauter D, Boukhobza T. (2006). Robustness against unknown networked induced delays of observer based FDI. *Proceedings of IFAC Safeprocess*, China,; 331–336.

Sauter D., Boukhobza T., Hamelin F.(2007). Adaptive thresholding for fault diagnosis of Networked Control Systems robust against communication delays, *In proceeding of Networked Distributed Systems for Intelligent Sensing and Control*, Kalamata Greece, pp. 222–240.

Llanos D, Staroswiecki M, Colomer J, Meléndez J. (2007). Transmission delays in residual computation. *IET Control,Theory and Applications*; 1(5):1471–1476.

Chabir K., Sauter D., Ben Gayed M.K. et Abdelkrim M.N. (2008) " Design of an Adaptive Kalman Filter for Fault Detection of Networked Control Systems ", *16th Mediterranean Conference on Control and Automation Congress Centre*, Ajaccio, France, pp. 1124–1129.

Al-Salami I. M., Ding S. X., Zhang P. (2008). Fault detection system design for networked control system with stochastically varying transmission delays, " *in Proc. of the IFAC* , seoul, Korea.

Zhang P., Ding S. X., Frank P. M. and Sader M. (2004). Fault detection of networked control systems with missing measurements. *Proceedings of the 5th Asian Control Conference*, Melbourne, Australia, 1258-1263.

Wang Y. Q., Ye H., Ding S. X., Wang G. Z., Wang Y. M. (2006). Observer based Residual Generation and Evaluation of Networked Control Systems Subject to Random Packet Dropout. *Proceedings of IFAC Safeprocess*, Barcelona, Spain, pp. 822-827.

Sauter D., Li S., and Aubrun C. (2009). Robust fault diagnosis of networked control systems. *Int. J. Adapt. Control Signal Process*, vol 23, issue 8, pp. 722-732.

Chabir K., Sauter D., Keller J. Y. (2009). Design of Fault Isolation Filter under network induced delay. *in Proc. of IEEE control applications (CCA)*, pp. 25-30, Saint Petersburg, RUSSIA.

Chabir, K., Sauter, D., Abdelkrim, M.N. and Ben Gayed, M.K. (2010). Robust fault diagnosis of networked control systems via Kalman filtering, *Int. J. Automation and Control*, Vol. 4, No. 3, pp.343–356.

Ding S. X., Zhang P. (2007). An observer Based Fault TolerantScheme for Distributed Networked Control Systems. *Proc. of European Contorl Conference*,Kos, Greece.

Patton R. J., Kambhampati C., Casavola A., Zhang P., Ding S., Sauter D. (2007). A Generic Strategy for Fault-tolerance in Control Systems Distributed

Over a Network, *European Journal of Control, Fundamental issues in Control,* Vol 13, Number 2-3,pp 280-296.

Kambhampati C, Patton RJ, Uppal FJ. (2006) Reconfiguration in networked control systems: fault tolerant control and plug-and-play. *Proceedings of IFAC Safeprocess*, China, 151–156.

Guerrero-Castellanos J.F., Lesecq S., Marchand N., and Delamare J. (2005). A low-cost air data attitude heading reference system for the tourism airplane applications. *Conference IEEE sensors 2005*.

Tayebi A. and McGilvray S. (2006). Attitude stabilization of a VTOL quadrotor aircraft, *IEEE Trans. Control Systems Technology*, 14(3), pp. 562-571

AUTHORS BIOGRAPHY

KARIM CHABIR received his Master in Automatic and intelligent technique in 2006 from the National Engineering School of Gabes (Tunisia) and cotuttelle Ph.D. in Automatic Control, from University Henri Poincaré (France) and University of Gabes in 2011. His research works were carried out at the Research Centre for Automatic Control of Nancy (CRAN) and at the Research Unit of Modelling, Analysis and Control systems of the National Engineering School of Gabes, Tunisia. He was a member of Dependability and system diagnosis group (SURFDIAG). His current research interests are focused on model-based fault diagnosis and fault tolerant with emphasis on networked control systems. He was a secondary school teacher of Gabes, Tunisia, from 10/2003 to 09/2007. He was also an Assistant professor in Faculty of Science of Gabes, Tunisia, from 10/2007 to 08/2011. Temporary Teaching and Research (ATER) since 2011 at Faculty of Science and Technology of Nancy, France.

Mohamed Amine SID was born in 1986 in Algeria. He received his master degree in automatic control option from the department of electrical engineering at Sétif University, Algeria. Since 2010, he has been working towards a PhD in Automatic and Signal Processing at the Centre de Recherche en Automatique de Nancy (CRAN, CNRS). His main research interests are in Networked control Systems, Fault Detection.

Dominique Sauter received the D.Sc. degree (1991) from Henri Poincar´e University, Nancy 1 (now the University Lorraine), France. Since 1993 he has been a full professor at this university, where he teaches automatic control. He was the head of the Electrical Engineering Department for four years, and now he is a vicedean of the Faculty of Science and Technology. He is a member of the Research Center in Automatic Control of Nancy (CRAN) associated with the French National Center for Scientific Research (CNRS). He is also a member of the French-German Institute for Automatic Control and Robotics (IAR), where he has chaired a working group on intelligent control and fault diagnosis. His current research interests are focused on model-based fault diagnosis and fault tolerant control with emphasis on networked control systems. The results of his research works are published in over 50 articles in journals and book contributions as well as 150 conference papers.

INTEGRATED MANAGEMENT AND OPTIMIZATION OF THE SANITATION CYCLE BY THE COMBINED USE OF EXPERT SYSTEMS AND SUPERVISORY SYSTEMS IN REAL TIME OPERATION

MARCELO DE SOUZA[(a)], **ALEXANDRE ACÁCIO DE ANDRADE**[(b)], **SERGIO LUIZ PEREIRA** [(c)]

[(a)] Pontifícia Universidade Católica de São Paulo - PUCSP
[(b)] Universidade Federal do ABC - UFABC
[(c)] Escola Politécnica da Universidade de São Paulo – EPUSP

[(a)]souzam@pucsp.br, [(b)]aacacio@ufabc.edu.br, [(c)]sergiope@pea.usp.br

ABSTRACT

The society's acknowledgement of the importance of Earth's water resources has been significantly enhanced over the last decades, in the 1990s and 2000s. The populational growth and the environmental degradation make the efficient management of water resources even more vital for humanity. This article presents a new systematic proposal comprising the use of automation technology in order to allow significant efficiency gains in harnessing, treatment, reservation, distribution, collection and management of water resources and thus to contribute towards a sustainable development. The proposed system is nominated Automated Management Integrated System of Intelligent Sanitation (*Sistema Integrado da Gestão Automatizada de Saneamento Inteligente* - SIGASI). Consists of a hardware and software architecture that integrate the Expert Systems (Artificial Intelligence) and the Supervisory Systems, in actual operation time. The system was virtually validated with tests based on parameters of an actual Water Pumping Station (WPS). The test methodology, the virtual treatment station simulation system and the automation system, the hardware and software architecture and the gains obtained in the tests are also shown in this article.

Keywords: Automation, Management, Water Resources, Sustainability, Expert System, Artificial Intelligence

1. INTRODUCTION

Water is not only an essential element for life, but a factor of major importance in the processes of social and economic development, increasingly taking an essential role in humanity's sustainable development. It is notorious and evident that the concern with water or its shortage has brought, to an increasing extent, challenges to different societies and Government bodies. However, the issue includes more than the existence of water resources, it embraces their use and supply systems, which are required to meet the needs related to people's general health and the economic and sustainable development (Tsutiya 2004).

1.1. Need of a Better Management of Water Resources

The use of water for different purposes has been expanded due to the intensification of human and economic activities and populational growth (Brandão 2003; Tsutiya 2004). Thus, during times in which the development process rate runs faster, there is an increase of the basic resources demand, including water. This was noted in some regions of the world, during the second half of the last century. Therewith, the management of water resources also grew more important and complex, benefiting from systems analysis theories and practices and operational researches since the end of the Second World War (Brandão 2003). Afirming that the sanitation system in Brazil is the context for which this article is headed, it becomes necessary to show a few macrodeficiencies in this segment's infrastructure, which may aid sustainable growth if minimized and help a more efficient management style, for example:

a. The absence of a device establishing the monitoring and encouragement of performance enhancements in the losses reduction in Brazillian water supply systems. According to Tsutiya (2004), the losses measured through the ratio between invoiced volumes and the volumes available for distribution were 40.6% of the national average value. Also in 2001, the State companies recorded an average income loss of 40.4%. (IWA 2002) In 2007 and 2008, this value was respectively 39.1% and 37.4%, according to a National System for Information report about Sanitation (*Sistema Nacional de Informações sobre Saneamento* – SNIS). The physical losses, represented by leakages in the distribution networks, have had serious consequences in the company's business performance. Since the lost water is not billed, they represent greater production costs and greater expenses with electrical energy and chemical products for water treatment, which end up being incorporated into the rates. Internationally, on average, Asia shows an overall loss of about 42% in its water systems, compared with 39% in Africa, 15% in North America, and 42% in Latin America and the Caribbean (WHO-UNICEF 2000). For example, the city of Delhi, India, showed a water loss rate of 53%, while the city of Tokyo, Japan, a country whose cities have the lowest water loss rates in the world, had an average loss of only 4.5%. In case of New York City, in the United States, the water loss indicator in the system suggests a water loss of 30%.

b. Ineffective operationalization of sanitation systems, which must be optimized through the use of more rational techniques and scientific work, possibly through partnerships with Universities. One clear example of optimization would be paying close attention to the electric power consumption issue in the sanitation cycle and making it more efficient, as electric power is the most significant element of the sanitation cost in Brazil, usuallyusually second only to payroll. For example, by using the data released by São Paulo Sanitation Company (*Companhia de Saneamento Básico do Estado de São Paulo* – SABESP) (2010), one can notice that in 2010 the kilowatt-hours/m³ number indicator in the water production process was 0.613 kWh/m³, with a consumption of 1,810,291,074 kWh and a produced volume of 2,952,386,248 m^3, providing a unit cost per kilowatt-hour of 0.13 (US$/kWh). In the sewage treatment process, the indicator came to 0.411 kWh/m³, with a consumption of 320,942,186 kWh, a treated volume of 780,950,895 m³ and a unit cost of 0.14 (US$/kWh).

c. Need for optimization in the use of chemical products used in the water and sewage system, through efficient and automated control, as these products have a direct effect on the water bill.

d. Need for processes automation, using Automation Technology, associated to intelligent system, as a tool to improve the management of water resources – which is the main purpose of this article.

1.2. Justification and Purpose of this Article

Currently, there are three main weak points in the operational models for water resources management. These are:

a. In order to carry out the management and control of sanitation subsystems, most managers and supervisors, known as sanitation managers, use a conventional management model in which the data is treated in sequential manner, which means that they issue a final result and process a certain volume of data in a repetitive way to solve a problem. This model may not perform more efficient and optimized actions such as the automatic systematization of knowledge through rules and procedures, to allow better decision-making, as well as the systematic standardization of knowledge between the different sanitation managers;

b. The sanitation manager action occurs at the third level of the theoretical model in the automation pyramid (Durkin 1994; Webb, J.K 1992). This level operates the Supervisory System, which receives information from controlling devices such as Programmable Logic Controllers (PLC), frequency inverters, control valves, etc. This system executes several vital functions within an automated process and also operates as an interface between Human Machine Interface (HMI), with operators and managers, enabling them to analyze, diagnose, infer and make decisions during the process. The managers and operators spend a lot of their working time operating the Supervisory System. This operation is usually executed by several managers in many subsystems of the sanitation cycle. According to the general automation theory (Andrade 2001), the action of human beings in an automated process is considered a manual action, and it is thus subject to all kinds of occurrences arising from non-automated processes;

c. The procedures involved in the sanitation cycle are usually known to the people responsible, but as a rule, this knowledge tends not to be fully systematized. Hence, the use of knowledge of the managers of the subsystems is not always done in an assertive and efficient manner, in the automation systems. Other factors such as restructuring of the functional workforce and retirements have meant that the knowledge accumulated by operators and sanitation managers tends to be lost, leading to costs resulting from undue operation and new training for replacement labor. Even making an effort for training, this is usually not enough to make sure of the full transmission of knowledge as acquired by the managers over the years.

The shortcomings mentioned in management of Brazilian sanitation system show the need to develop an efficient methodology which is able to ensure the preservation of the knowledge held by those responsible for the function and operation of the sanitation subsystems and also to optimize the operation routines. For this, the use of Knowledge Engineering and Expert Systems at the third level of the theoretical model of the automation pyramid, more specifically at the base of supervisory systems, comes as an alternative to the methods currently used by the managers of the sanitation system in Brazil. Figure 1 shows the proposal made by this methodology.

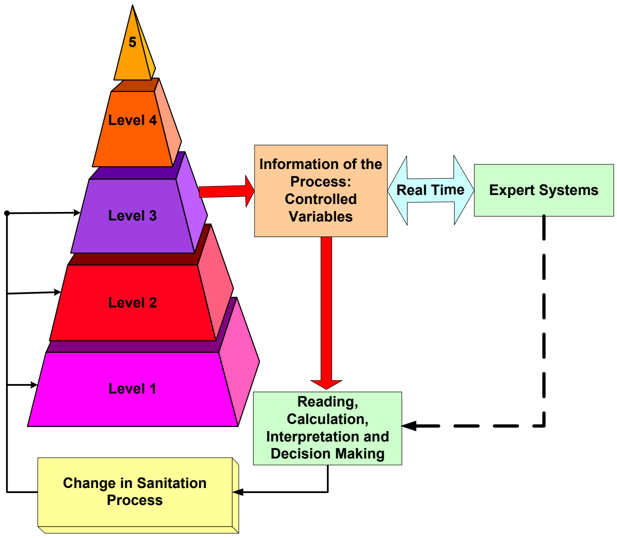


Figure 1. Proposal for Sanitation Cycle Management Methodology

Figure 1 shows the actions of the system operator, as well as the actions of the Expert System, working to promote the integrated management of the sanitation cycle. Here we see that this activity may occur mainly in the three lowest levels of automation period, according to the following cycle:

a. The operator perceives the process information in S.S. through his or her senses;

b. The operator then uses the capacities of his or her brain for data analysis and interpretation, making decisions or not;

c. After the decision, the operator changes the operating parameters, which can range from maintenance help and diagnosis (Level 1), verification of logic of operation (Level 2) or mainly direct operation of the S.S. (Level 3).

d. The Expert System operates in real time with the S.S. base to provide guidance to and/or automate the functions executed by the sanitation system operators, through rules and procedures stored on the Expert System database.

As proposed, this methodology allows two different methods of operation and management:

a. Operate as total or partial substitute, in the sanitation cycles in decisive processes;

b. Operate in real time, as a counselor to the managers and operators, in the sanitation process.

Thus, according to the operational skill of the Supervisory Systems and the capacity to deal with the knowledge of the Expert Systems, the union of architecture of hardware and the architecture of the software packages, in this article, makes it feasible to help or even automate the decision making in the sanitation cycle operation and management. Thus, the objective of this study is to develop an integrated management model, which uses an Expert System and which could give guidance automatically to the operator in the sanitation cycle management, if implemented. This proposed architecture, in this article is known as the Integrated System for the Automated Management of Intelligent Sanitation (*Sistema Integrado da Gestão Automatizada de Saneamento Inteligente* – “SIGASI”).

2. AUTOMATION IN SANITATION SYSTEMS

The automation in systems for water supply and sanitary drains consists of collecting, concentrating and processing of the process information with the use of Information Technology (Souza 2011). Based on the obtained results, the automation systems act in an autonomous manner on the states and the variables in the process to obtain the results as desired. Figure 2 shows the process of data collection and activities in the Sanitation system.

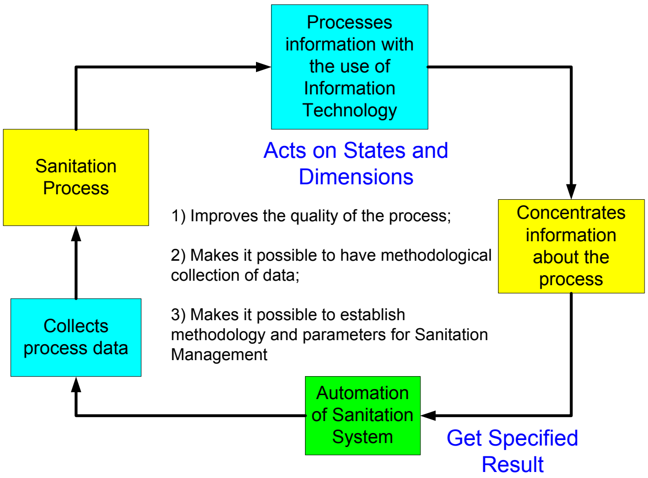


Figure 2. Automatic Process for Data Collection in the Sanitation System

Figure 3 shows the subprocesses of the sanitation cycle.

Figure 3. Sanitation Cycle - Source: Water Supply System Company from State of São Paulo – Sabesp (2005)

Figure 3 shows that the sanitation cycle involves all the water treatment stations, sewage treatment stations, water Pumping stations, reservoirs and all sources of water resources; thus, the number of communication points involved (data exchange between field signals and controllers – tags, as they are known in technical and commercial literature) is highly significant. This requires an enhancement of the automation systems to meet the need for the organizations to carry out a global management of their processes.

The greatest complexity of the automation process then requires that responsible parties have greater effort and skills to manage a high number of communication points. Exemplifying with data from the local water company (SABESP) which serves around 20 million people in São Paulo Metropolitan Area (*Região Metropolitana de São Paulo* - RMSP), including 27 municipalities, in an urban sprawl of over 8,500 km² and which uses an Operational Control Centre (*Centro de Controle Operacional* – CCO), consisting of a supervisory system to monitor and control the water transport system in the RMSP: currently, its set of communication points consists of about fifteen thousand (15,000) tags. To meet the current needs and demand for automation, the same supervisory system is being adapted to control one hundred and fifty thousand (150,000) tags.

The main reasons to justify the investments made in automation in the water supply system refer to the improvement of water treatment and distribution through real time monitoring and control, reduction of operational costs, through the management of electric power consumed and control of physical losses in the system.

F. Mário (2001) explains that the concept of automation in water supply systems is very similar to what happens in the case of electrical system. In the same way that this segment can be divided into generation, transmission and energy distribution, the sanitation segment has water production, transport to reservoirs, distribution to consumers, sewage collection and treatment.

The automation of sanitation in many Brazilian companies is still an one-off occurrence. This is a consequence of the lack of planning and resources of water companies, which has continued for several decades. One obstacle faced for the adoption of automation in this segment is the geographical aspects, which have an influence on communication means. Normally, the remote units of monitoring and control are installed in locations where there is poor infrastructure of telecommunications or electric power, implying the use of alternative structure such as mobile phones, transmission by radio waves, frame relay and others.

3. PROPOSAL OF HARDWARE AND SOFTWARE ARCHITECTURE

The proposal of the development methodology of hardware and software architecture, as well as the operational methodology of SIGASI, are presented in this item. SIGASI is a new proposal for an automated methodology, which proposes an integration between Supervisory Systems and Expert Systems, seeking advances in the sanitation cycle management. Its macroarchitecture is presented in details, as well as the stages of their implementation through simulations in virtual tests, with the due results as obtained.

Supervisory Systems (S.S) and Expert Systems (E.S) have capacities, which can be associated to promote the optimization of water resources management. The Supervisory Services offer easy interpretation, flexibility, process structure, generation of income, scripts, information trackability and easy operation, while the Expert Systems represent a powerful tool to deal with knowledge. Due to this fact, SIGASI, in its architecture, integrates E.S. and S.S, to make the most of the conditions offered by the two systems.

The methodology for development of SIGASI was based on the study by Andrade (2001), which drew up a software solution for operation, in real time (Laplante, P. A. 2004)., of Expert Systems with Supervisory Systems for industrial automation, known as the System for Integration Between Specialist and Supervisory Systems (*Sistema de Integração Sistemas Especialistas e Supervisórios* -(SISES). This new software became one of the models to be part of SIGASI. Figure 4 of this Article shows the macroarchitecture of SIGASI.

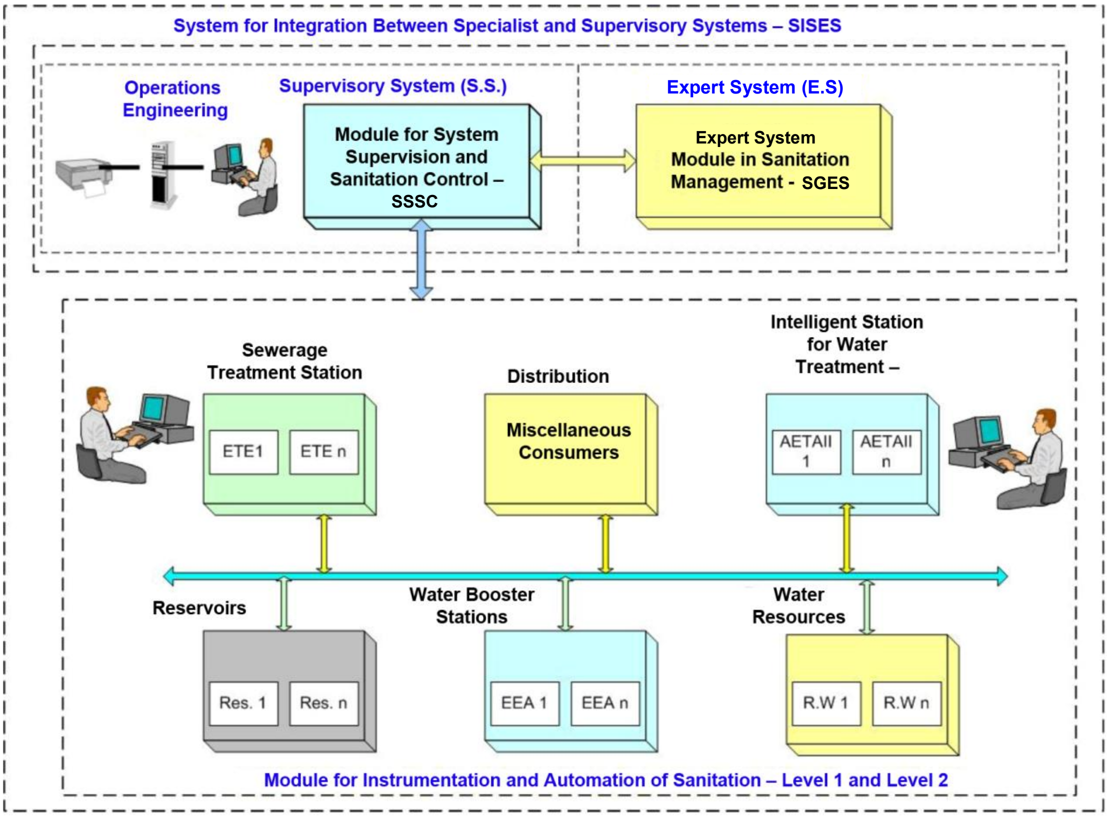


Figure 4 – Macroarchitecture of Hardware e Software of SIGASI

As shown in Figure 4, the SIGASI comprises the following modules:

a. Module for Instrumentation and Automation of Sanitation (*Módulo de Instrumentação e Automação do Saneamento* – MIAS) (levels 1 and 2);

b. System for Sanitation Supervision and Control (*Sistema de Supervisão e Controle do Saneamento* – SSSC) (level 3);

c. System for Integration Between Specialist and Supervisory Systems (*Sistema de Integração Sistema Especialista e Supervisório* – SISES);

d. Expert System of Sanitation Management (*Sistema Especialista e Gestão do Saneamento* – SGES).

The macroblock of the Sanitation Instrumentation and Automation Module (MIAS) conceptually comprises levels 1 and 2 of the automation pyramid module, which is part of Instrumentation and Controllers.

The System for Sanitation Supervision and Control (SSSC) uses electronic supervision and uses data communication technologies and computer systems to concentrate all the information of a system, installation or set of installations in one single operating point. Through electronic supervision, one or more operators operate much equipment, distributed throughout a large area. This module also operates the Supervisory System, which receives information from control devices such as Programmable Logical Controllers, frequency inverters, control valves etc. Supervisory Systems have a tool for storage of the variables of greater relevance, regarding the production system.

The System for Integration between Specialist and Supervisory Systems (SISES) is a computer tool or an environment for development of Expert Systems, which enables communication with the Supervisory Systems in real time. SISES consists of two macroblocks: Module of Interface with Supervisory Systems and the Module for Construction of Expert Systems.

The Expert System of Sanitation Management (SGES) is a Expert System dedicated to operation with S.S. in real time and with the aim of helping and/or automating the taking of decisions in the management of the sanitation cycle. Figure 5 shows the macroarchitecture of the SGES.

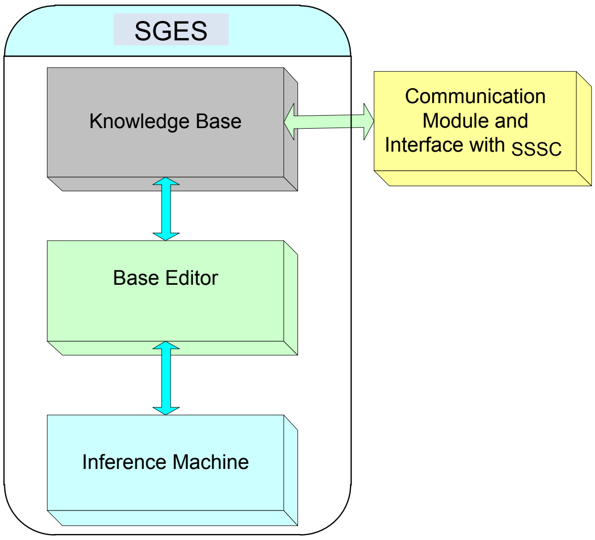


Figure 5. Basic Architecture of the Expert System Generated in the SGES

As shown in figure 5, SGES includes the following features:

a. Knowledge Base: this is the information (facts and rules) used by a specialist, represented in computer form;

b. Base Editor: this is how Shell (specific software packages for drawing up of E.S.) allows the implementation of the desired bases;

c. Inference Machine: this is the part of E.S. responsible for deductions on the knowledge base;

d. Communication Module: this is the part of E.S. that allows communication in real time with System for Sanitation Supervision and Control (SSSC).

3.1. Stages for the Implementation and Validation of SIGASI

As the first stage in a long and permanent project, the phases as defined for this article were the following:

a. Definition of Targets and Subprocesses to be monitored by SIGASI;

b. Development of Expert System:
  - b.1. Knowledge Management activity;
  - b,2. Preparation of Communication Base;
  - b.3. Preparation of Inference Machine;

c. Development and Implementation of Virtual SIGASI;

d. Definition of the methodology to validate SIGASI;

3.2. Target and subprocess monitored by SIGASI

In this article, we present the macrotargets defined to be simulated and monitored with the SIGASI proposal. These are:

a. Reduction of expenses with electrical energy, through water pumping;

b. Reduction of electricity consumption with the pumping of water;

c. Reduction of water loss in the supply system, using intelligent systems;

d. Optimization of the quantifiers of supplies for water production;

e. Optimization and establishment of grounds for maintenance actions of Sanitation Cycle;

f. Generation of database in real time.

To simulate the rules and procedures used in a water supply system, first, a part of the subprocess was selected from a water supply system, where the water Pumping station (WPS) was monitored by SIGASI. According to Tsutiya (2004), Water Pumping Stations are defined as “sets of buildings, installations and equipment, intended to house, protect, operate, control and maintain the lifting sets (motorized pumps) which promote the pumping of the water”.

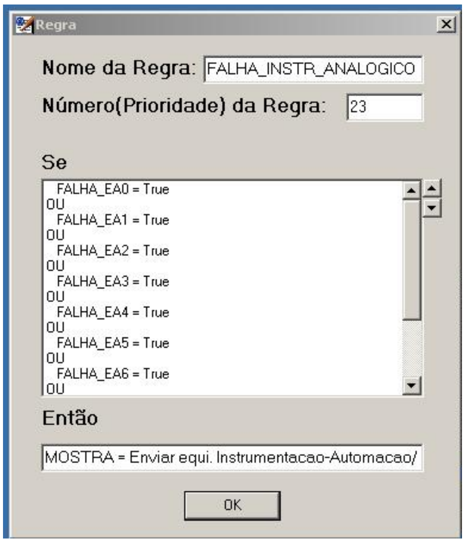
3.3. Development of the Expert System

For the implementation of the Expert System of Sanitation Management (SGES), there was the need to develop three basic phases:

a. Knowledge Management Activities: this stage involved the development of knowledge engineering activities. In this phase, information was collected specialists, end users and operators of the system. The main target here was to acquire relevant information about the sanitation system operation.

b. Knowledge Base Preparation: this stage was the systematization of the information (facts and rules) which had been collected in the previous phase. Such information was used by the sanitation subprocesses specialist to solve a certain process problem, and is computationally represented on the E.S. The information is systematized in this phase so the system may help with decision-making during the sanitation system operation, whether performed by the operator or automatically.

c. Inference Machine Preparation: this stage consisted of preparation of the inference machine, part of E.S. responsible for deductions on the knowledge base, as constructed in the previous stage of this article. Figure 6 shows the rule synthesis, which was prepared to reach the objectives of this article.



Nome da Regra: FALHA_INSTR_ANALOGICO

Número(Prioridade) da Regra: 23

Se

FALHA_EA0 = True
OU
FALHA_EA1 = True
OU
FALHA_EA2 = True
OU
FALHA_EA3 = True
OU
FALHA_EA4 = True
OU
FALHA_EA5 = True
OU
FALHA_EA6 = True
OU

Então

MOSTRA = Enviar equi. Instrumentacao-Automacao/

Figure 6. Example of the Preparation of Rules for the Sanitation Process

3.4. Development and Implementation of Virtual SIGASI

This item presents the development and implementation of the practical application of SIGASI for the real sanitation system. This part of the article shows the development as well as the tests executed, making use of SGES generation tools, an E.S. capable of observing the variation of S.S. parameters in the sanitation system, and have an influence thereon in real time. This E.S. provides messages about the inferences resulting from the interaction with S.S. The SGES Module also allows the inferences lead to actions in S.S., to allow the operationality of SIGASI. For the compliance with this stage, a strategy was prepared, as shown in Figure 7 that follows.

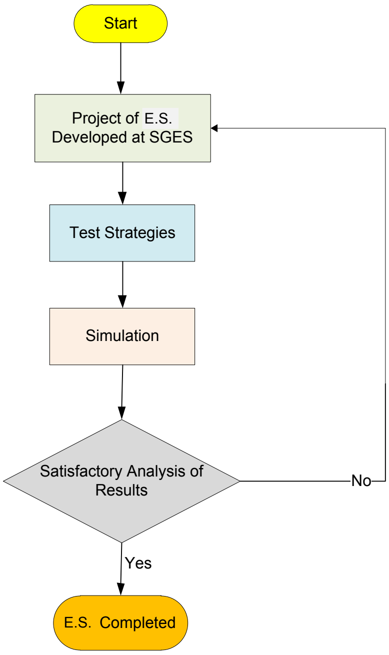


Figure 7. Strategies for the Preparation of E.S. in SGES

3.4.1. Methodology for Validation of Virtual SIGASI

SIGASI was validated through simulation. For this, there was the programming of the SGES module, which operated in a laboratory with a Supervisory System, where the variables of a sanitation process were then treated, according to the need to promote a simulated contingency. For the tests execution, a Programmable Logic Controllers - PLC, emulator was used, a supervisory system developed in a similar way to the current systems of a sanitation company. The main function of the PLC emulator is to feed the supervisory system data which, in turn, communicates with the E.S. in real time. Thus, by simulation, part of a sanitation system is virtually replicated. The macrostructure of this model's virtual hardware is shown in Figure 8.

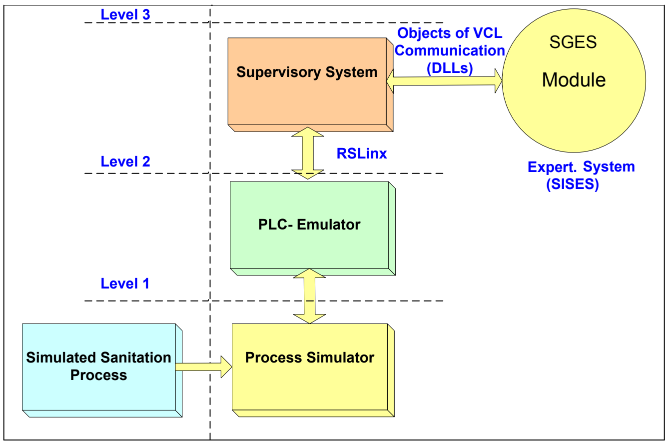


Figure 8. Macroarchitecture of the Virtual Hardware of SIGASI's Validation Model

In Figure 8, it should be noted that the field signals are virtual, which means they are generated by the process simulator. The emulator itself is not enough for simulation because it cannot reproduce the field responses; it needs the process simulator. The software used for the PLC emulator simulation and the process software was the RSLogix Emulate 500 software package made by Rockwell Software (2004), due to its technical features, which allow a simulation of complex processes, truthful to the reality of a process to be simulated. This software makes it possible to emulate PLCs from Allen Bradley, from SLC (Small Logic Controller) family, based on the programs that shall be run on the real PLCs; in other words, it processes the ladder lines (Michel 1990) or the blocks of an SFC (Sequential Flow Chart) (Michel 1990). This emulator, just like the PLCs used by Allen Bradley, uses as data exchange mechanism the RSLinx communication software made by Rockwell (Rockwell 2004), which manages all data exchange, also being responsible for the exchange that occurs between PLCs and S.S. Between SISES and S.S. (RSView), the communication is made through two DLLs known as VCL objects; they are small programs typical of Windows package.

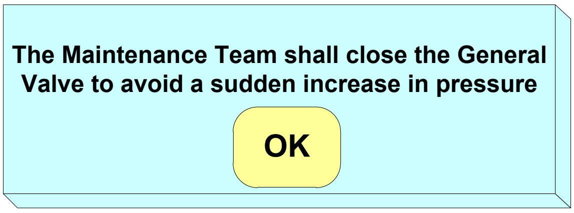
3.4.2. SIGASI Validation Criteria

To validate the SIGASI simulation tests, the following criteria have been adopted:

a. Global performance of the system: the system as a whole shall operate in a satisfactory way;
b. S.S. Performance: the S.S. shall not suffer a significant drop in performance;
c. Execution time: the time for E.S. to reach a conclusion cannot be dimensionally higher than the standardized times that human operators take to reach the same conclusions;
d. Compatibility: the E.S. cannot request such a volume of resources from the operational system (Windows) that has a negative effect on performance;
e. Assertiveness: E.S. has to reach the correct results.

3.4.3. Tests and Performance Analysis of SIGASI

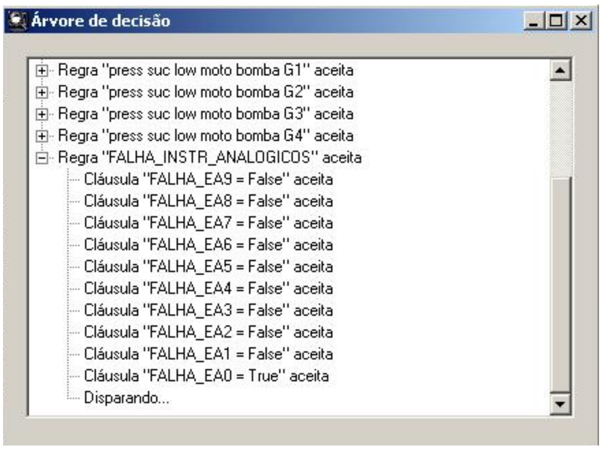
During the tests phase, the correct results, as well as clear and objective messages that were supplied by SGES showed compliance with the objective of this study. SGES presented the conclusions about the actions that the operators should take in seconds, making them much faster than the usual times taken to make decisions, which are often considered in terms of minutes. The performance of the SGES was satisfactory in all testing criteria. Figure 9 shows an example of the message screens generated by SGES, during the tests on the hardware and on the software packages of SIGASI.



The Maintenance Team shall close the General Valve to avoid a sudden increase in pressure

Figure 9. Example of a message screen generated by SGES in the process of automation test of water Pumping station (WPS)

Note that the messages generated by SGES are obtained through heuristics, supplied by specialists in the sanitation process, which have been turned into rules in the SGES (see item 3.3). The message sent to the S.S. shall be recognized by the operator of the subprocess, through the "OK" button, to allow recognition of the contingency informed by SGES, as also for the operationality of E.S. itself. Another relevant confirmation to be stressed during SIGASI tests was the capacity of SGES to map, through a decision tree diagram, the sequence of how the operator came to a solution for a certain contingency of the sanitation process. Figure 10 shows an example of a decision tree, obtained during SIGASI simulation.



```
Regra "press suc low moto bomba G1" aceita
Regra "press suc low moto bomba G2" aceita
Regra "press suc low moto bomba G3" aceita
Regra "press suc low moto bomba G4" aceita
Regra "FALHA_INSTR_ANALOGICOS" aceita
    Cláusula "FALHA_EA9 = False" aceita
    Cláusula "FALHA_EA8 = False" aceita
    Cláusula "FALHA_EA7 = False" aceita
    Cláusula "FALHA_EA6 = False" aceita
    Cláusula "FALHA_EA5 = False" aceita
    Cláusula "FALHA_EA4 = False" aceita
    Cláusula "FALHA_EA3 = False" aceita
    Cláusula "FALHA_EA2 = False" aceita
    Cláusula "FALHA_EA1 = False" aceita
    Cláusula "FALHA_EA0 = True" aceita
    Disparando...
```

Figure 10. Details of the decision tree as generated by SGES

Notice that in Figure 10 the decision tree supplies the sequence of events (rules) and shows the way in which SGES came to the solution of a certain contingency in the sanitation process. This function of the SGES is very useful and important in the system operation, especially in processes with a significant number of tags, such as the sanitation cycle (see item 2). Surely, the human specialist would have great difficulty to describe, in systematic form, how he or she reached the solution of a certain problem. Thus, we can conclude that through this function of the SGES it is possible to systematize, over time, the taking of decisions in a certain contingency situation within the sanitation process, through procedures and rules. Thus, it is able to transfer the knowledge systematically to different operators, some of whom have even less information about sanitation subprocesses, as well as the register itself, of what was stored in theory and in practice by experienced professionals who are now close to retirement.

4. CONCLUSIONS AND FUTURE RESEARCH

This article presents the operational integration of a sanitation automation Supervisory System with the Expert System, thus bringing together the dynamic reality of the sanitation automation process and the knowledge acquired in E.S., in order to prepare the resulting system to make inferences on the sanitation cycle set and also execute actions in this set as mentioned.

This methodology, using E.S., has a relative advantage when compared to conventional automation systems: the fact of allowing deliberate actions on decision-making, in real time, about occurrences involving equipment of the sanitation program, which still, for some reason, are not inserted in the automation process. Through programmed rules and messages on E.S, screen, it is possible, for example, for the operator to receive guidance through E.S. method: "close the entrance valve of a certain water supply system, due to excess pressure, caused by power outages in the region. If this valve is not closed, then there can be significant damage to the tubing". Even though the valve is not part of the automation system, it is still possible to establish actions through E.S.

In short, the conclusion is that SIGASI brings significant technical and economic benefits compared to the management system currently used by managers of sanitation companies. This enables, for example, greater grounds in taking maintenance actions, as well as greater response speed to the disturbances of the process, systematization of procedures used for decision-making, greater knowledge of specialists who are active in leading the subprocesses of the sanitation cycle and the automatic taking of decisions, in real time, including sectors as yet not automated. As this is a new methodology, this study can be easily replicated for other sanitation subprocesses. This new management methodology, as proposed for the sanitation segment, helps to improve automation and, thus, the management of water resources.

As for the macrotargets defined (item 3.2), SIGASI has shown that they comply, in qualitative fashion, with all the six points as raised, mainly due to the fast identification of operational problems, which avoids significant waste in the productive sanitation process, both for water and supplies used in the production thereof, as also in the electricity consumption and respective expenses, incurred with the water pumping process.

The SIGASI proposal is an opening for new lines of investigation in the field of sanitation management, water resources and environment, with the use of automation technology associated to intelligent systems. This article, on seeking a more efficient and optimized style of the sanitation cycle management, intends to expand the contribution so that the availability and use of water, one of the most important natural resources on the planet, may be done in a more rational manner, thereby improving the conditions for sustainability of human development.

As possibilities of future enhancements of this work, we could mention:

a. Continuity of research and implementation of SIGASI for the accurate quantification of the indicators as well as the relations mentioned in item 3.2 of this article;

b. Improvement of H.M.I., making the operation even easier for the users;

c. Research in new techniques of software packages to enhance the SISES;

d. Implementation of the proposal of this research work in a sanitation unit, during a test period for validation, before managers and sanitation operators.

REFERENCES

Andrade, Alexandre Acácio de. *Emprego de Sistemas Especialistas em Sistema Supervisório de Automação Industrial.* [Use of Expert Systems in Supervisory Systems of Industrial Automation], São Paulo, 2001. Dissertation for Mastership – Polytechnic School, University of São Paulo.

Brandão, B. L.J. *Modelo para Operação de Sistema de Reservatório com Usos Múltiplos*. [Model for Operation of Reservoir System with Multiple Uses]. São Paulo, 2003. Doctoral Thesis of the Polytechnic School of the University of São Paulo.

Durkin, J.; *Expert Systems Design and Development.* Prentice Hall, 1994 800p.

F. Mário - *Automação no saneamento básico: diferentes necessidades para um mesmo objetivo*. [Automation in Basic Sanitation: Different Needs for one same Objective]. In: *Revista Controle & Instrumentação*, ed. 61, São Paulo, SP, 2001.

Goldberg DE (1989) *Genetic algorithms in search, optimization and machine learning. Addison-Welsey, Reading.*

Global Water Supply and Sanitation Assessment 2000 (WHO-UNICEF).

IWA – *International Water Association em SNIS* 2002.

Laplante, P. A. (2004). Real-Time Systems Design and Analysis, 3 edn, Wiley-IEEE Press.

Mahdi Zarghami-*Urban Water Management Using Fuzzy-Probabilistic Multi-Objective Programming with Dynamic Efficiency*. Water Resour Manage (2010) 24:4491–4504.

Mascarenhas, F. José. A *Infra-Estrutura no Brasil, Conselho de Infra-Estrutura*, [Infra-structure in Brazil, Infra-structure Council], Brasília, 2005.

Mitchell., T.M., *Machine Learning*. McGraw-Hill, 1997, 52p.

Michel G. *Programmable Logic Controllers*, John Wiley & Sons, 1990. 330p.

Prakash Kaini & Kim Artita & John W. Nicklow. *Optimizing Structural Best Management Practices Using SWAT and Genetic Algorithm to Improve Water Quality Goal.* Water Resour Manage (2012) 26:1827–1845.

Rockwell Software Inc.. *RSLogix Emulate 500* – User's Getting Results Guide - USA, 2004. 28p.

Russell S. Norvig P. *Artificial Intelligence: a Modern Approach*. Prentice Hall 2nd edition, 2004 page 284.

SABESP – Companhia de Saneamento Básico do Estado de São Paulo: *Relatório Mensal de Acompanhamento dos Gastos com Energia* Elétrica, [Monthly report on the monitoring of expenses with electrical energy], December 2010.

Souza, Marcelo de. *Gestão e Otimização Integrada de Ciclo do Saneamento pelo Emprego. Combinado de Sistemas Especialistas com Sistemas Especialistas em Tempo Real de Operação* São Paulo, 2011. [Management and Integrated Optimisation of the Sanitation Cycle through the Combined Use of Expert Systems with Expert Systems in Real Time Operation]. Doctoral

Thesis, Polytechnic School, University of São Paulo.

Tsutiya, M.T.; *Abastecimento de Água. São Paulo*, [Water Supply in São Paulo], Provo e Gráfica, 2004.

Webb, J.K.; *Industrial Control Electronics*. Maxwell Macmillan International Editions, 1992 593p.

AUTHORS' BIOGRAPHY

Sergio Luiz Pereira was born in São Paulo, Brazil, and holds a PhD from Escola Politécnica da Universidade de São Paulo (EPUSP), 2009; Doctor in Electrical Engineering, also from EPUSP (1995); MSc in Robotics Systems and Applications fromby Coventry University (United Kingdom) (1988) and Engineer graduated on Faculdade de Engenharia de São Paulo – FESP, 1982). He teaches at graduate and post-graduate level and also tutors master's and doctor's students at Escola Politécnica da USP (Polytechnic School at USP). Is is a consultant in Ecoeconomics and Industrial Automation and has a wide range of work published in periodicals and national and international congresses. Sergio is also a professor on the MBA program at Business São Paulo School (BSP) and at Pontifícia Universidade Católica de São Paulo (PUC-SP). He has also been a coordinator of the Production Engineering course at PUC-SP, and was the Head of Department and also Coordinator of Electrical Engineering course and also the Coordinator of the MBA program at Fundação Armando Álvares Penteado (FAAP). He coordinated the partnership between EPUSP and ELETROPAULO for the preparation of the Operational Diagnosis of Public Lighting System of the city of São Paulo. He was also a consultant and coordinator of the project between EPUSP and CODESP for the Automation and Integration of the Road and Rail Weighing System at the port of Santos and also a Research and Development Engineer at Metal Leve Controle Eletrônicos.

Marcelo de Souza was born in São Paulo – SP, Brazil, graduated as Technician in Electrotechnique from Escola Técnica Federal de São Paulo (ETFSP) in 1992. In 2001, he graduated as Electrical Engineer from Faculdade de Engenharia de São Paulo – (FESP). In 2006, he received a Master's Degree in Electrical Engineering from Escola Politécnica da Universidade de São Paulo (EPUSP). In 2011, he completed his PhD studies in Electrical Engineering on this same University. He is a professor at Pontifícia Universidade Católica de São Paulo (PUC-SP), and also works as an engineer at the São Paulo water and sanitation supply company, *Companhia de Saneamento Básico do Estado de São Paulo* - SABESP. His main area of interest includes automation engineering and also power and electronic electricity systems in water supply systems. He has had articles published in periodicals and national and international congresses.

Alexandre Acácio de Andrade was born in São Paulo, SP, Brazil in 1969 and graduated as Electrical Engineer on Escola Politécnica da USP – (EPUSP) in 1998. In 2001, he was declared Master in Electrical Engineering by Escola Politécnica da Universidade de São Paulo (EPUSP) and then, in 2007, completed his PhD on the same University. He is an assistant professor at Universidade Federal do ABC, having previously worked for ten years as an automation engineer. His main areas of interest and research include automation engineering, robotics, artificial intelligence and management. He has had articles published in magazines and also national and international congresses.

VOLTAGE ENVELOPE, NOISE AND HILBERT TRANSFORM

Federico Muiño[(a)], Maximiliano Carabajal[(a)], Marcela Morvidone[(a)(b)], Carlos D'Attellis[(a)(b)], Matías Fabbro[(a)]

[(a)] Facultad Regional Buenos Aires, Universidad Tecnológica Nacional, C. A. de Buenos Aires, Argentina.
[(b)] Centro de Matemática Aplicada, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina

[(a)]maximiliano_carabajal@hotmail.com, [(b)]cdattellis@yahoo.com.ar

ABSTRACT

The fluctuation of voltage is one of the important power quality events. Different methods have been proposed for estimating the voltage envelope, but the presence of noise is, in general, not considered. A method for estimating the envelope in presence of noise, based on the Hilbert transform and a low-pass filter, is presented. The results obtained from a real signal measured from an arc furnace are shown.

Keywords: Hilbert transform, flicker, voltage fluctuation

1. INTRODUCTION

Voltage fluctuations can be described as systematic variations or random variations in the voltage envelope. The fluctuation of voltage is one of the important power quality events due to the effects of electronic and control systems, and in the light flicker. There are several sources of voltage flickers as arc furnaces, fans, pumps, lifts, switching of powers factor capacitors, large motors (IEC 60038; Arrillaga, Watson, and Chen 2000).

Different methods have been proposed for estimating the magnitude and frequency of flicker. The IEC 61000-4-15 and IEEE 1453 standards recommend the square demodulation, a method used in demodulation of AM signals, which consists in tracking the flicker envelope by squaring the input voltage signal. Other methods proposed are Fast Fourier Transform (Schauder 1999), Least Absolute Value (Soliman, and El-Haway 2000), Kalman filters (Girgis, Stephens, and Makram 1995), Wavelet transform (Chen, and Meliopoulos 2000), Teager Energy Operator (Abdel-Galil, El-Saadany, and Salama 2002). Hilbert transform is also used (Abdel-Galil, El-Saadany, and Salama 2004; Su, and Wang 2008; Li, Zhao, and Han 2005; Marei, Abdel-Galil, and El-Saadany 2005), and, in particular, using Prony analysis and Hilbert Transform (Feilat 2006).

Recently, the performance of several flicker detecting methods were compared (Chen, Jia, and Zhao 2009). The core of flicker analysis is to track the envelope of voltage signal, that is, the instantaneous amplitude. Then, an important characteristic of the algorithms proposed is their on-line behavior; the faster is the estimation of the voltage envelope values, the better is the on-line behavior.

Another aspect of the problem is the presence of noise. In Tong, Yuan, Li, and Song 2008 this problem is pointed out and solved using the Hilbert transform for estimating the flicker envelope and the wavelet transform for extracting other noises contained in the voltage flicker. A signal obtained from an arc furnace shows that the high frequency noise can not be ignored. However, the influence of noise in the methods previously cited is not considered.

This paper is organized as follows: the use of the Hilbert transform for an efficient estimation of the voltage envelope is explained in Section 2, in Section 3 we describe the signal used to test the method, and finally Section 4 presents the numerical results.

2. ENVELOPE ESTIMATION USING THE HILBERT TRANSFORM

2.1. Estimating the envelope

In this section we review some results concerning the estimation of the envelope of a discrete signal.

The Hilbert transform is used in signal processing to derive the analytic representation of a signal $x[n]$. The analytic representation of a signal is well known for continuous-time signals (Carmona, Hwang, and Torresani 1998) and it is also defined for discrete signals as

$$z[n] = x[n] + i\,x_h[n]$$

Where $x_h[n] = \mathrm{H}\{x[n]\}$ denotes the discrete Hilbert transform of the sequence $x[n]$ (Li, Li, and Qian 2010). This representation allows a straightforward identification of the envelope of an amplitude modulated signal. An amplitude modulated signal is modeled by:

$$x[n] = a[n]\cos(\omega n) \tag{1}$$

Where the frequency content of $a[n]$ has an upperbound less than ω. In this conditions, Bedrosian theorem for discrete signals (Li, Li, and Qian 2010) states that:

$$\mathrm{H}\{x[n]\} = a[n]\,\mathrm{H}\{\cos(\omega n)\},$$

which turns into $\mathrm{H}\{x[n]\} = a[n]\sin(\omega n)$. Now, the analytic representation of the signal takes the simple form

$$z[n] = a[n]e^{i\omega n},$$

and the amplitude (or the envelope) $a[n]$ is easily obtained from $a[n] = |z[n]|$.

For the sake of completeness, we include Bedrosian theorem as stated in (Li, Li, and Qian 2010):

Theorem 1: Suppose that $z_1[n]$ *and* $z_2[n]$ *are complex sequences with discrete-time Fourier transforms* $Z_1(e^{i\omega})$ *and* $Z_2(e^{i\omega})$. *Then*

$$\mathrm{H}\{z_1[n]z_2[n]\} = z_1[n]\,\mathrm{H}\{z_2[n]\}$$

if there exists a nonnegative number $\sigma < \pi$ *such that*

$$Z_1(e^{i\omega}) = 0, \text{ for } 0 < \sigma < |\omega| < \pi, \text{and}$$
$$Z_2(e^{i\mu}) = 0, \text{ for } 0 < |\mu| \le \sigma < \pi.$$

2.2. Hilbert filter

In this section we describe the Hilbert filter in more detail.

The discrete Hilbert transform $\mathrm{H}\{x[n]\}$ of the sequence $x[n]$ is defined in the frequency domain as (Hahn 1996)

$$X_h(\omega) = F\,\{\mathrm{H}\{x[n]\}\} = -i\,\mathrm{sgn}(\omega)X(\omega) \qquad (2)$$

where $X(\omega)$ is the discrete Fourier transform of x:

$$X(\omega) = F\,\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n}$$

From equation (2), the transfer function of the Hilbert transform for discrete signals is

$$H(\omega) = \begin{cases} -i, & 0 < \omega < \pi \\ 0, & |\omega| = \pi \\ i, & -\pi < \omega < 0 \end{cases}$$

$$H(\omega) = -i\,\mathrm{sgn}[\sin(\omega)] = G(\omega)\,e^{i\frac{\pi}{2}}$$

with $G(\omega) = -\mathrm{sgn}[\sin(\omega)]$. The discrete time representation of the Hilbert filter is easily obtained from this expression. In fact, $G(\omega)$ is an odd function whose Fourier transform reads

$$G(\omega) = \frac{4}{\pi}\sum_{m=0}^{\infty}\frac{1}{2m+1}\sin[(2m+1)\,\omega]$$

Denoting $h(k) = F^{-1}\{H(\omega)\}$ the inverse Fourier transform of $H(\omega)$, we have

$$h(k) = \frac{2}{\pi k}\sin k\frac{\pi}{2}, \; k \ge 0, \text{ and } h(-k) = -h(k)$$

Figure 1 shows the impulse response $h[k]$ of a Hilbert filter of order 38 and Figure 2 shows the magnitude response $|H(\omega)|$.

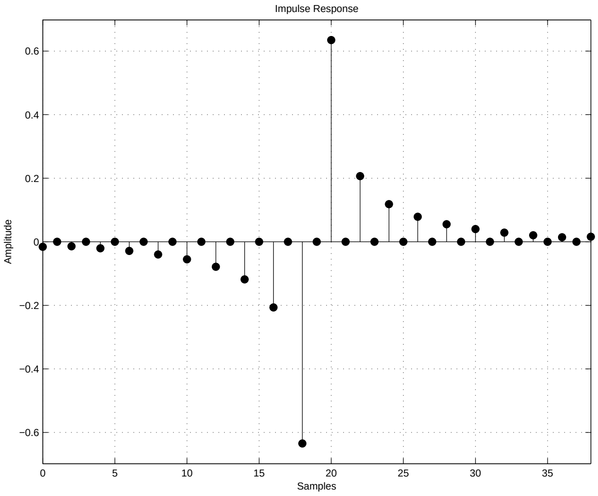


Figure 1: Impulse response

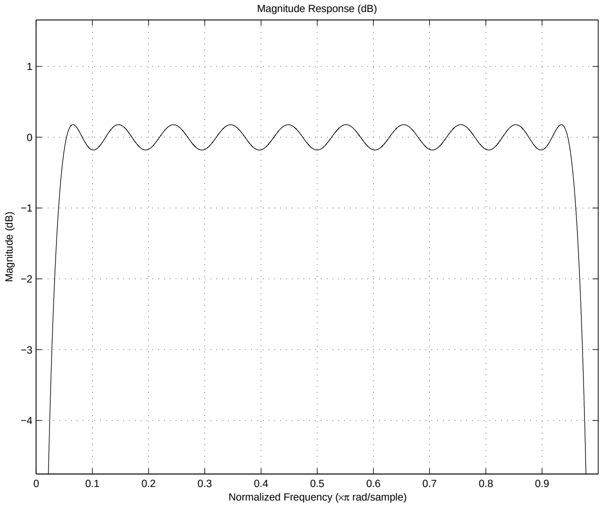


Figure 2: Magnitude response

2.3. Implementation of the Hilbert filter

The Hilbert transform can be easily implemented in the discrete domain by an FIR filter. This kind of discrete filter allows obtaining constant group delay and constant phase over the entire bandwidth. Moreover, the nature of the filter makes unnecessary the stability analysis.

Moreover, if the filter order is even, it behaves as a band pass filter which has zeros at 0 Hz and at the Nyquist frequency. So that its impulse response is similar to that shown in Figure 1, in which the odd coefficients are zero. However, if the filter order is odd, the zero in the Nyquist frequency disappears and the odd coefficients are no longer zero.

For this reason, the Hilbert filter of even order is easier to implement than the filter of odd order. This is because the zero coefficients can be omitted. Therefore fewer multiplications and additions are required (see figure 3).

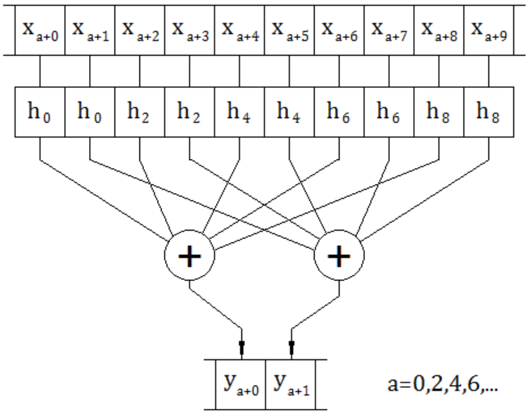


Figure 3: Convolution strategy

In this paper, the calculation of the coefficients is performed with the "Filter Design & Analysis Tool" in MATLAB

3. A SIGNAL FROM AN ARC FURNACE

In this paper, the performance of the proposed model on tracking the voltage flicker signal envelope is examined with a signal coming from real measurements. It is a typical AC arc furnace application in a steel plant. This arc furnace is served from a 13.8 kV bus. The measured signal is one of the phase voltages and it was sampled at a sampling frequency of 1000 Hz.

Since this signal is distorted by noise that comes from making physical measurements, a serious issue is the robustness of the method for estimating flicker. Previous works hardly consider this problem.

As an example, we make some comments on Prony algorithm which has been used in this problematic (Feilat 2006). Prony algorithm is good at system identification provided that the available samples come from a signal completely predictable and free of any randomness. Under these conditions, it is a good alternative for obtaining mathematical models in the form of damped complex exponentials from a small number of samples. This type of representation allows a straightforward calculation of the Hilbert transform of the signal (Feilat 2006). However, when the signals have some degree of randomness like signals immersed in noise, Prony algorithm is very unstable and it requires a large amount of samples to achieve an acceptable approximation. This method has some other drawbacks: it is very difficult to estimate the optimal number of exponentials to use in the approximation, the calculation involves two pseudo-inverse matrices whose systems are poorly conditioned, which increases the instability, and finally it has a high computational cost.

4. NUMERICAL RESULTS

We present numerical results on the estimation of voltage envelope of the signal from an arc furnace which was described in the previous section.

The method is implemented in Simulink from MATLAB (see Figure 4).

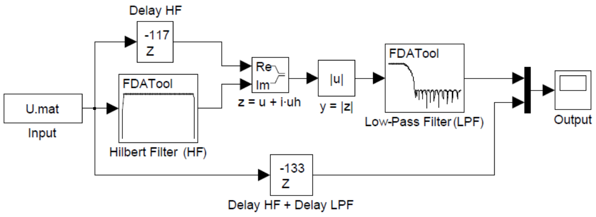


Figure 4: Block diagram of the estimator

From the block diagram Figure 4 we can make the following analysis:

1. The Hilbert filter is a FIR, all zeros filter of order 234, linear phase, whose magnitude response are shown in Figure 5.

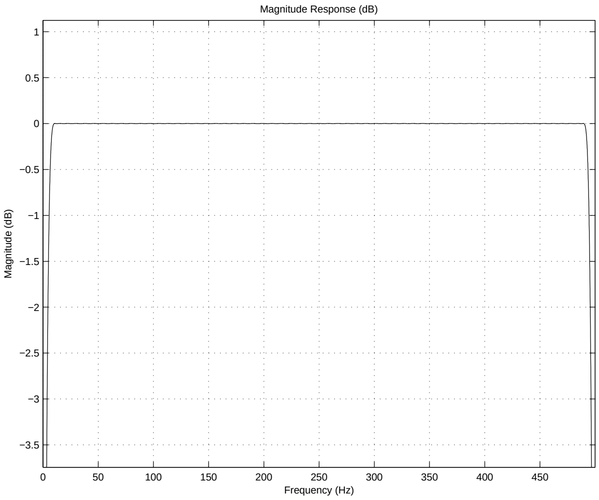


Figure 5: Magnitude response of Hilbert filter

2. The envelope is estimated, then a low-pass filter is applied to minimize the influence of noise. This is a FIR equiripple filter of order 32, with cutoff frequency 120 Hz.
3. Because of the fact that the signal passes through two FIR filters, there is a delay time in the tracking processes. However, as the two filters have a linear phase response, these have a constant group delay response. Therefore, this time delay is constant for all frequencies and it can be calculated. In this case, it resulted in a total delay of 133 samples, and it represents 0.133 sec at a sampling frequency of

1000Hz. It represents approximately 8 cycles of the fundamental frequency (60Hz).

In Figure 6 and 7 are shown estimations of the voltage envelope corresponding to the diagram of Figure 4 using the measurement of the arc furnace voltage. There are two intervals of 0.5 sec: [18.5,19] and [19,19.5].

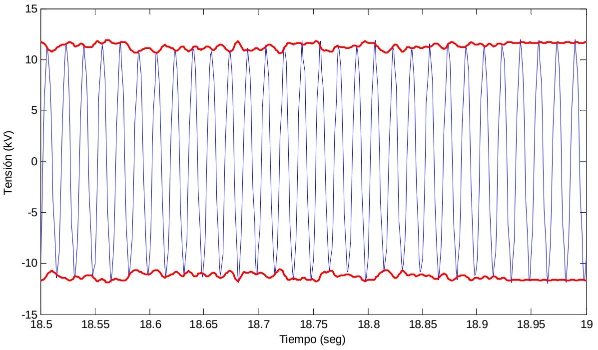


Figure 6: The signal and its estimated envelope.

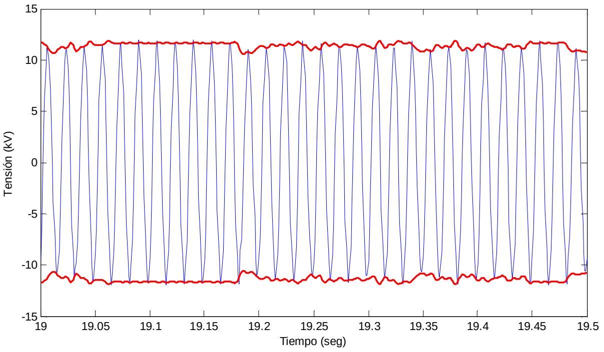


Figure 7: The signal and its estimated envelope.

5. CONCLUSIONS

We have presented a method for estimating signal flicker in presence of noise using the Hilbert transform. The characteristics of both filters (the Hilbert filter and the low pass filter) are described in Section 4. The proposed algorithm was implemented in a DSP Blackfin EZ-537. We tested this technique on a real world signal produced by an arc-furnace. In contrast to the method presented in (Feilat 2006), this algorithm does not have instabilities, as was analyzed in section 4.

ACKNOWLEDGMENTS

We thank Alberto Del Rosso for providing us the measurement of the arc furnace voltage

REFERENCES

International Standard IEC 60038:2009. IEC standard voltages

International Standard IEC 61000-4-15:1999. Flickermeter: Functional and Design Specifications.

IEEE Standard 1453:2004. IEEE Recommended Practice for Measurement and Limits of Voltage Fluctuations and Associated Light Flicker on AC Power Systems.

Abdel-Galil, T.K., El-Saadany, E.F., and Salama, M.M., 2002. Energy operator for on-line tracking of voltage flicker levels. *Proceeding of IEEE/PES Winter Meeting, 3:1153–1157.* January 27-31, New York (New York, USA).

Abdel-Galil, T.K., El-Saadany, E.F., and Salama, M.M., 2004. On-line tracking of voltage flicker utilizing energy operator and Hilbert transform. *IEEE Transactions on Power Delivery* 19:861–867.

Arrillaga, J., Watson, N.R., and Chen, S., 2000. *Power System Quality Assessment*. Ney York: Wiley.

Carmona, R., Hwang, W-L., and Torresani, B., 1998. *Practical Time-Frequency Analysis, Gabor and Wavelet Transforms with an Implementation in S.* Academic Press.

Chen, M.T., and Meliopoulos, A.P.S., 2000. Wavelet-based algorithm for voltage flicker analysis. *Proceedings of the IEEE 9th International Conference on Harmonics and Quality of Power*, 2:732–738. October 1-4, Orlando (Florida, USA).

Chen, Q., Jia, X., and Zhao, C., 2009. Analysis on Measuring Performance of Three Flicker Detecting Methods. *Proceeding of IEEE/PES General Meeting*, 1–7. July 26-30, Calgary (Alberta, Canada)

Feilat, E.A., 2006. Detection of voltage envelope using Prony analysis-Hilbert transform method. *IEEE Transactions on Power Delivery* 21(4):2091–2093

Girgis, A.A., Stephens, J.W., and Makram, E.B., 1995. Measurement and prediction of voltage flicker magnitud and frequency. *IEEE Transactions on Power Delivery* 10(3):1600–1605.

Hahn, S.L., 1996. *Hilbert Transform in Signal Processing*. Artech House.

Li, H., Li, L., and Qian, T., 2010. Discrete-time analytic signals and bedrosian product theorems. *Digital Signal Processing* 20:982–990.

Li, T., Zhao, Y., and Han, Y., 2005. Application of Hilbert-Huang transform method in detection of harmonic and voltage flicker. *Power System Technology* 29:74–77.

Marei, M.I., Abdel-Galil, T.K., and El-Saadany, E.F., 2005. Hilbert transform based control algorithm of the DG interface for voltage flicker mitigation. *IEEE Transactions on Power Delivery* 20(2):1129–1133.

Schauder, C., 1999. STATCOMfor compensation of large electric arc furnace instalations. *Proceeding of IEEE/PES Summer Meeting*, 2:1109–1112. July 18-22, Edmonton (Alberta, Canada).

Soliman, S.A., and El-Haway, M.E., 2000. Measurements of power system voltage and flicker levels for power quality analysis: a static LAV state estimation based algorithm. *International Journal of Electrical Power and Energy Systems* 22:447–450.

Su, H., and Wang, Y., 2008. Voltage flicker detection method based on mathematical morphology filter and Hilbert transform. *Proceeding of Canadian Society for Ecology and Evolution*, 28:111–114. May 11-14, Vancouver (British Columbia, Canada).

Tong, W., Yuan, S., Li, Z., and Song, X., 2008. Detection of Voltage Flicker Based on Hilbert Transform and Wavelet Denoising. *Proceeding of the 3th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, 2286–2289. April 6-9, Nanjing (Jiangsu, China).

QFT CONTROL APPLIED TO A DRIVE BY WIRE (DBW) SYSTEM

Fabio Delatore (a), **Fabrizio Leonardi** (b), **Alexsander T. Carvalho** (c), **Carlos A. Morioka** (d)

(b) Centro Universitario da FEI – Sao Bernardo do Campo/SP
(a), (c), (d) FATEC Santo Andre – Santo Andre/SP

(a) fabio.delatore@gmail.com, (b) fabrizio@fei.edu.br, (c) tressino@gmail.com, (d) carlos.morioka@uol.com.br

ABSTRACT

Tradionally, the throttle valve positioning was performed mechanically by means of a steel cable. Nowadays at the embedded system stage, an electromechanical system named as *Drive by Wire* (DBW) substitutes the direct positioning. The DBW is controlled by the vehicle *Engine Control Unit* (ECU) and is responsible to adjust the mass air flow delivered to the engine and to control the idle engine rotation. The throttle valve control is somehow a challenging task because of nonlinear phenomena caused by the spring and the gearbox. The present work aims to design a robust parametric control for a DBW system, using a plant model identified numerically at different operations points. The results show that the controller is able to deal with the nonlinear phenomena providing a reasonable performance with no steady state error and a consistent setting time.

Keywords: QFT control, Throttle Valve, Engine Control, Drive-by-Wire.

1. INTRODUCTION

Traditionally and for many years, the union between the gas pedal (car accelerator) and the throttle valve was performed mechanically by means of a steel cable, in order to perform the opening/closing procedure of the valve (Deur *et al.* 2005). This valve is responsible to control the air supply to the vehicle engine, keeping the desired engine rotation and torque according with the driver request, by using the gas pedal. Another important item presented on this system, is the engine idle rotation actuator, which is responsible to keep the engine in a specific rotation when there is no driver´s request on the gas pedal (Morioka *et al.* 2011).

With the advent of the electronic embedded systems, mainly on the engine management systems (represented by the electronic fuel injection), the throttle control system was modified and incorporating some others components, such as a potentiometer responsible to inform the valve position (named as *Throttle Position Sensor* – TPS) and a DC motor, responsible to open/close the valve in a combination with a gearbox (Deur *et al.* 2005). This new system is named as *Drive by Wire* – DBW and it is nowadays responsible to adjust the mass air flow delivered to the engine in a similar way of the mechanical system with the steel cable did (Tilli *et al.* 2000).

The DBW is an important evolution on the automotive management systems. Including the DC motor, it not only eliminated the steel cable to perform the opening/closing task, but it also allows eliminate the idle engine actuator. The DBW has two important tasks on the engine management. It is responsible to control the engine rotation at idle and also to supply the engine with the exact air quantity (Tilli *et al.* 2000).

A simplified structure of the DBW system it is shown in the Figure 1. Figure 2 brings a Volkswagen EA-111 engine throttle valve which contains the TPS sensor and the DC motor. Note that the system input is the gas pedal position and the system output is the TPS sensor.

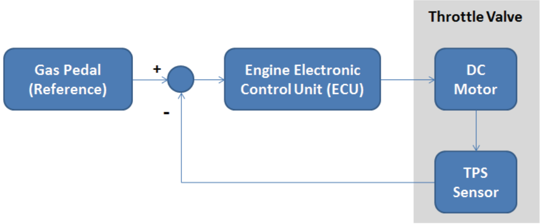


Figure 1: DBW system block diagram.

Figure 2: Volkswagen EA-111 throttle valve (Author).

The throttle valve is one of the most important actuators presents on the moderns' cars engines. As mentioned before, it involves a DC motor which drives the throttle plate through a gearbox unit, and the opening angle is measured by means of a potentiometer integrated into the gearbox. This gearbox has two mechanical stops that define the valve opening range, which is approximately 90º. In case of a failure associated to the DC motor, the valve plate is repositioned into the home position (0º) by a spring mechanism (Reichhartinger and Horn, 2009).

Nonlinear phenomena like stick–slip friction, gear backlash, and discontinuities, mainly caused by the spring mechanism, render the control of this mechatronic system a challenging task (Reichhartinger and Horn, 2009). This fact is documented by a number of publications dedicated to the modeling, identification, and control of electronic throttle devices (Corno *et al.* 2011; Reichhartinger and Horn 2009; Deur *et al.* 2005; Tilli *et al.* 2000; Poggio *et al.*1997).

The authors Gharib *et al.* (2010) proposed to use the QFT technique to design a controller of the engine at idle speed. They used a phenomenological model that relates the inputs, throttle angle and load torque, with the outputs, manifold pressure and engine speed. The model was linearized and they used the second order transfer function from the throttle angle to the motor speed for the robust design. The gain and the damping coefficient were used as uncertainties, but there is no further discussion of this choice. In that paper, the authors considered that the throttle angle is given, i.e., do not take into account the dynamic between the pedal and the throttle valve. This present article addresses the problem of dynamics of this loop between the pedal and the throttle valve, and therefore can be viewed as an internal control loop of the system described by Gharib *et al*. (2010).

The throttle dynamics is highly nonlinear and its performance affects the response of the engine speed control system. As pointed out before, the throttle valve dynamics affects directly the engine rotation. It must be point out that the knowledge of the dynamics of the throttle valve control system is important during the engine calibration process. On this process, different engine operating regimes are evaluated for different temperature and atmospheric pressure. With the DBW system, the valve opening with a correct pre-determined dynamic contributes to the driving comfort.

This work proposes the usage of a QFT controller to be applied on a throttle valve in order to impose an adequate throttle dynamic and robustness to the DBW system.

2. METHODOLOGY

The present work aims to design a robust control system for a DBW system using a plant model identified numerically by applying several steps on a VW EA-111 throttle valve. Different from some works founded on the scientific literature (i.e. Poggio *et al.* 1997) which presents the throttle modeling, this work intends to identify an approximated model by applying several step signals on an open loop structure. Analyzing the parameters variation of the model obtained, the Quantitative Feedback Theory design is used to develop a robust controller to the DBW system with a desired dynamic.

2.1. Robustness

The modeling errors are translated into uncertainty of the plant transfer function. The dynamic model of the plant considered in this work includes three uncertain parameters.

The stability margins (gain and phase) are often used to evaluate the tolerance of the system to modeling errors of the gain and phase of the transfer function. However, in fact, they are fragile to reveal the degree of robustness of the system, because even systems with high stability margins may have its corresponding Nyquist diagram near to the critical point -1+0*j* and therefore are not robust (Da Cruz 1996). The example of Figure 3 illustrates how even high gain and phase margins are unable to represent the robustness of the system.

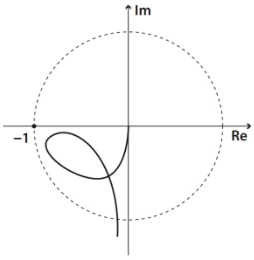


Figure 3: Robustness *x* Stability Margins.

Note that the values of the stability margins are approximately 90° for the phase margin and infinite for the gain margin. These values suggest that the system tolerates large modeling errors, and so has high robustness. Note that due to the apparent vicinity of the curve to point -1+0*j*, a small change in the plant model may cause a change in the number of encirclements of the critical point, causing the system to lose stability.

The uncertainty of a model can be classified as structured and unstructured. The unstructured uncertainties are usually associated to the parts not modeled of the plant and are frequency dependent. Note that normally the neglected characteristics are of high frequency dynamics. The parametric uncertainties are associated to structured uncertainties such as the uncertainties in the model of this paper. The common used techniques for the case of parametric uncertainties are the μ-synthesis and QFT (Houpis 1999).

To characterize the unstructured modeling errors, one may define the multiplicative error representing the relative difference between all the real plants and the nominal plant model in relation to this nominal model. For the design purpose the modeling errors are evaluated by means of the absolute value of a frequency dependent upper limit of the errors. Note that since unstructured modeling error are evaluated without phase information, there is an inherent conservatism associated to the obtained controllers in the sense that the loop gain is higher than minimum needed for the case where the phase information has been taken into account.

For the case of structured modeling errors one way to represent the errors is by means of the frequency response for each transfer function of the real plants, called templates. This frequency response is usually represented in the Nichols chart because it allows evaluating the magnitude and phase in the same plot, and also allows determining directly the values of gain and phase margins.

2.2. Quantitative Feedback Theory (QFT) Design

The QFT design (Quantitative Feedback Theory) is a technique in the frequency domain in the Nichols chart. As pointed by Borghesani (1993) and Yaniv (1999), the first step of design procedure is the determination of templates generated by the parameters uncertainties. A template is defined as the collection of uncertain plant frequency response functions at a given frequency. However, for the design, only the bounds of those templates are important.

Performance specification imposes barriers to the loop gain in the Nichols chart, and these templates should be above barriers in the specified frequency range. Margins of stability or robustness associated to the maximum resonance peak of the closed loop, imposes barriers around (0, -180º) in the Nichols chart.

The region around this point must be reshaped so that the boundary of all the templates does not violate this region. The design starts by selecting a point on the border of the template as the nominal plant and then, based on this point, the curve around (0,-180°) should be reshaped so that, when the nominal point does not violate the new curve, all points of the template are outside the original curve.

Then, the problem of finding a controller that meets the requirements of robust performance and stability should be done, for example, by trial and error, adding poles and zeros to the controller transfer function.

3. EXPERIMENTAL RESULTS

3.1. Overview of the DBW experimental system

The DBW experimental system developed on this work is presented on Figure 4, including some electronic circuits, the throttle valve and the gas pedal (car accelerator).

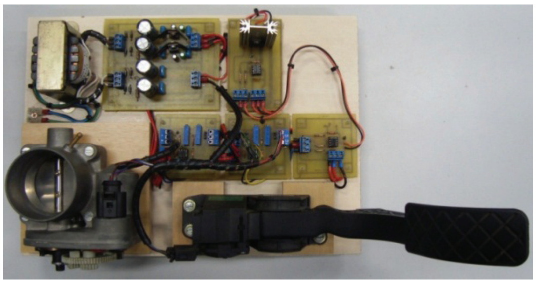


Figure 4: Overview of the experimental system.

The DBW system set point is defined by the gas pedal (desired angle) and the output is the throttle valve angle, measured by the TPS sensor. The gas pedal and the TPS sensor signals are conditioned into a 0-10 Volts output range, turning the plate position totally closed or totally opened, respectively, through the PWM driver (Morioka *et al*. (2011).

Figure 5 shows the connections between the DBW experimental system and the Advantech PCI 1718 DAQ board installed on a PC computer, operating at 30kHz sampling frequency.

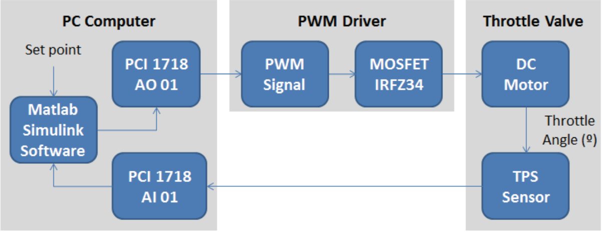


Figure 5: Experimental system and DAQ connections.

3.2. Plant model

In order to reveal the system dynamics it was performed several step tests with different amplitudes in open loop through the PWM. The typical dynamic obtained is presented by Figure 6.

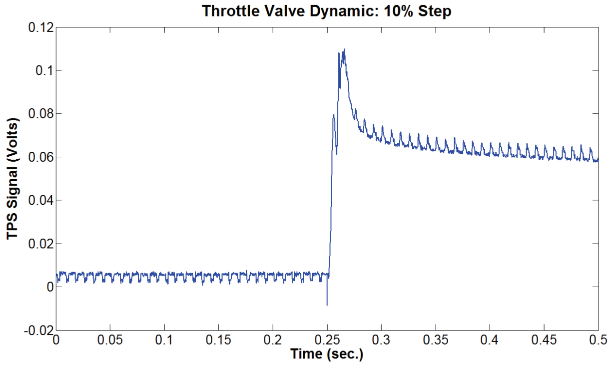


Figure 6: Throttle valve typical dynamic.

Using the information collected by the data acquisition system, several curves have been plotted. The observation of these curves, allow one to elect the transfer function

$$F(s) = K\frac{as+1}{bs+1}, \tag{3.1}$$

as a candidate to represent the plant. Although the amplitude and time constant are not the same for all plots, the shape pattern was approximately the same for all tests. The constants K, a and b were obtained by a numerical identification procedure through the least square method (LSM). The range obtained with these analyses represents the plant uncertainties and are shown on the Table 3.1.

Table 3.1. Range of the model parameters.

Model Parameters	Range	
	Minimum	Maximum
K	0.05	0.09
a	0.12	0.20
b	0.07	0.15

3.3. QFT controller design

As a performance specification, it is desired that the control system output tracks a reference signal with an error below 10% at least up to 0.1 rad/s in the presence of simultaneous parametric uncertainty in the range represented on the Table 3.1.

Based on the plant frequency response with nominal parameters, it was elected the 0.1, 5, 10 and 100 rad/s as the working frequencies for the QFT design. In the sequence, 64 plants have been chosen by the simultaneous variation of the 3 parameters within the specified range, generating the templates presented by Figure 7.

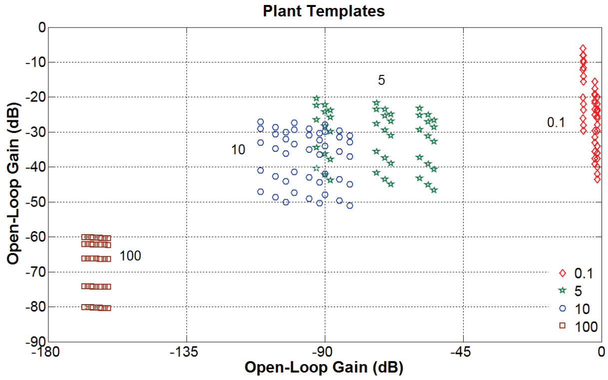


Figure 7: Templates for 0.1, 5, 10 and 100 rad/s.

The choice of the QFT controller structure and its parameters is usually done by trials (Yaniv 1999). However, there are several controllers that do not violate these specific stability bounds in the Nichols chart. To help with this choice, we also used the Root Locus plot. Note that the model plant has a pole and a zero relatively far away from the imaginary axis, making it impossible to achieve the closed loop time constant near to 1s, which is a typical value for this application. To solve this issue, it was chosen a controller with a zero relatively close to the imaginary axis and a pole at the origin, providing also null steady state error, thus resulting in a controller with a PI structure.

In the last step on the QFT design, a stability margin of 1.05 dB was defined and this bound was reshaped by the templates (see Figure 8). The resulting controller is presented by Equation 3.2.

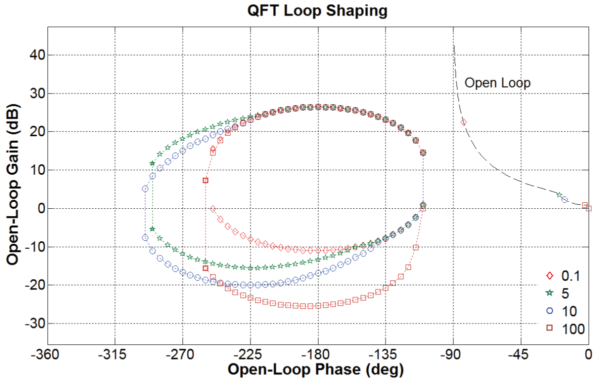


Figure 8: QFT loop shaping.

$$G(s) = 2.2\frac{15s+1}{s}. \tag{3.2}$$

3.4. Time domain performance

The experimental structure used to check the controller performance is the same one presented by Figure 5. The Matlab/Simulink software was used to implement the closed loop, including the controller.

The robustness of the system was tested by operating the valve with different opening angles. Varying the angle it modifies the gain, the pole and the zero position of the $F(s)$ (Equation 3.1), demonstrated by the parameters variation on Table 3.1

The three tests were performed during 14s by applying three different step signals, on t=5s: 11°, 15° and 20° valve opening. The throttle dynamic and the control effort were observed and compared with the desired amplitude defined by the step amplitude. The results obtained are shown by the Figures 9, 10 and 11.

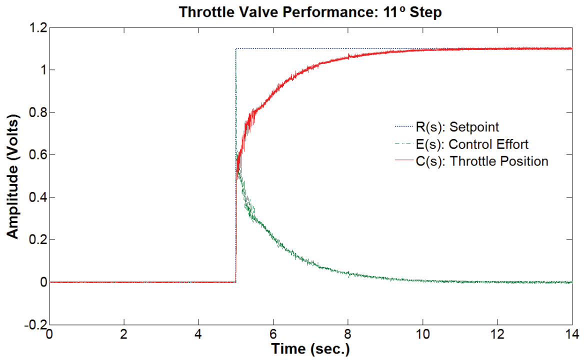


Figure 9: Performance for a 11° step variation.

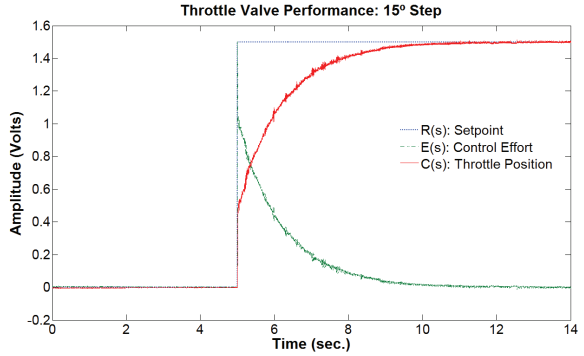


Figure 10: Performance for a 15° step variation.

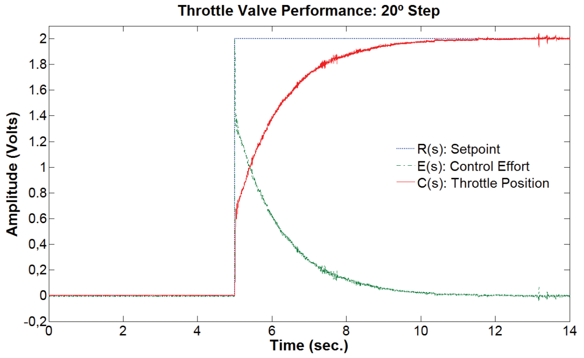


Figure 11: Performance for a 20° step variation.

Analyzing the results presented by the Figures 9, 10 and 11, it is easy to notice that the throttle position achieved the desired set point at the steady state and a reasonable setting time (approximately 0.8 seconds) was obtained on each individual test. The setting time could be modified by changing the controller gains in order to adapt the DBW system + controller to the engine dynamics, i.e. a 2.0L engine has a different dynamic and a demanded air supply of a 1.0L engine.

4. CONCLUSIONS

Due to the nonlinearities of the drive by wire systems, the phenomenological model is not trivial but also implies difficulties in the definition and design of the controller. Alternatively, this article discussed the modeling and experimental parametric robust control applied on the drive by wire system.

An open loop experimental investigation was used in order to provide the insights into the structure of the linear model. The non-linearities have been incorporated through a parametric variation over the three coefficients of the transfer function defined.

Only few robust control methods provide good tolerance to simultaneous variation of the plant parameters, so the robust technique chosen for the controller design was the Quantitative Feedback Theory. Through trial, typically on the QFT design, a PI controller was selected bases on its robustness observed on the Nichols chart and its performance on the real system.

The performance of the system observed on time domain confirmed what was expected by the robust design, since for a wide range of operating conditions, the performance of the system was high. The observed transient dynamic response can be roughly approximated by a 1st order dynamic with a less than 1 sec. time constant.

Our group intends to continue the studies, so we proposed to investigate the ability to adjust the time response for the DBW system in order to be used on different engines, which require different reactions of the DBW system due the different engine dynamics.

REFERENCES

Borghesani, C., Chait, Y. and Yaniv, O., 1993. QFT Frequency Domain Control Design Toolbox. User’s Guide, Terasoft Inc.

Corno, M., Fabbri, L., Savaresi, S. M., Tanelli, M., 2011. Design and Validation of a Gain-Scheduled Controller for the Electronic Throttle Body in Ride-by-Wire Racing Motorcycles. *IEEE Transactions on Control Systems Technology, vol. 19, no. 1, January 2011.*

Da Cruz, J. J., 2006. Robust Multivariable Control. University of Sao Paulo (USP), 163p.

Deur, J., Jansz, M., Pavkovic, D., Peric, N., 2005. Adaptative control of automotive electronic throttle. *Control Engineering Practice 14, March 2006.*

Deur, J., Jansz, M., Pavkovic, D., Peric, N., 2004. An Electronic Throttle Control Strategy Including Compensation of Friction and Limp-Home Effects. *IEEE Transactions on Industry Applications, vol. 40, no. 3, May/June 2004.*

Gharib, A.R., Gharib, M.R., Alireza, S., Mousavi, S., Gorkani, M. R., 2010. Control of Throttle Valve in Idle Speed Condition. *Proceedings of the 2010 International Conference on Modeling, Identification and Control, Okayama, Japan, July 17-19* 2010.

Houpis, C. S.; Rasmussen, S. J. 1999. Quantitative feedback theory. Dekker.

Poggio, L., Mazzucco, M., Scattolini, R., Siviero, C., Ricci, S., Rossi, C., 1997. Modeling and Identification of an Electromechanical Internal Combustion Engine Throttle Body. *Control Engineering Practice, vol 5, no. 9, May, 1997.*

Morioka, C., Albaladejo, F. S., Carvalho, A. T., Lagana, A. A. M., Delatore, F., 2010. Development of a mockup for education in automotive electronics. *Proceedings of XXXVIII Brazilian Conference on Engineering Education, September 12-15, Fortaleza (Ceara, Brazil).*

Morioka, C., Albaladejo, F. S., Carvalho, A. T., Lagana, A. A. M., Delatore, F., 2011. Experimental Teaching Set for ECU Engine Systems. *Proceedings of XXXIX Brazilian Conference on Engineering Education, October 03-06, Blumenau (Santa Catarina, Brazil).*

Reichhartinger, M., Horn M., 2009. Application of Higher Order Sliding-Mode Concepts to a Throttle Actuator for Gasoline Engines. *IEEE Transactions on Industrial Electronics, vol. 56, no. 9, September 2009.*

Tilli, A., Tonielli, A., Rossi, C., 2000. Robust Control of a Throttle Body for Drive by Wire Operation of Automotive Engines. *IEEE Transactions on Control Systems Technology, vol. 8, no. 6, November 2000.*

Yaniv, O. 1999. Quantitative Feedback Design of Linear and Nonlinear Control Systems. *Kluwer Academic Publishers.*

MECHANICAL ANALYSIS OF 2D-BRAZED JOINT USING A NEW HYBRID “MAX-FEM” MODEL

A. IFIS$^{(a)}$, **F. Bilteryst**$^{(b)}$, **M. Nouari**$^{(c)}$

(a) Laboratoire d´Energétique et de Mécanique Théorique et Appliquée, LEMTA CNRS-UMR 7563, GIP-InSIC, 27 rue d’Hellieule, 88100, Saint Dié des Vosges, France
(b) Laboratoire d´Energétique et de Mécanique Théorique et Appliquée, LEMTA CNRS-UMR 7563, GIP-InSIC, 27 rue d’Hellieule, 88100, Saint Dié des Vosges, France
(c) Laboratoire d´Energétique et de Mécanique Théorique et Appliquée, LEMTA CNRS-UMR 7563, GIP-InSIC, 27 rue d’Hellieule, 88100, Saint Dié des Vosges, France

$^{(a)}$abderrazzaq.ifis@gmail.com $^{(b)}$francois.bilteryst@insic.fr $^{(c)}$mohammed.nouari@insic.fr

ABSTRACT

This work deals with the performance of a new approach combining the numerical eXtended Finite Elements Method ‘X-FEM’ and the analytical method of Matched Asymptotic Expansions ‘MAE’. The proposed new “MAX-FEM” model is well adapted for studying and modeling the mechanical behavior of mediums containing singularities such as thin layers or adhesive joints without any required mesh refinement in their vicinity. The methodology consists of the construction of enrichment parameters with the ‘MAE’ technique and their integration into the ‘XFEM’ formulation. Correction matrix of stiffness is then defined and integrated in the FEM computation algorithm. To describe the mechanical behavior of a proposed structure with 2D brazed joints, the “MAX-FEM” hybrid model has been implemented as an UEL subroutine under Abaqus implicit. Compared with the classical FE method, the obtained results in terms of stress field, strains and displacements show a good accuracy without any required mesh refinement.

Keywords: hybrid technique, thin layers, Matched Asymptotic Expansions, X-FEM, correction matrix, UEL ‘MAX-FEM’ subroutine.

1. INTRODUCTION

During the last decade, several methods have been introduced for modelling singular problems such as thin layers, adhesive joints, coating, etc.. Analytical approaches based on asymptotic assumptions such as Matched Asymptotic Expansions (MAE) give at two different-scales an approximation of the main solution in structures containing singularities (M. Van Dyke, 1975; P. Schmidt, 2008). As shown by the work of Leguillon and Abdelmoula (2000), the MAE method has been used to analyse brazed joints in order to describe the crack propagation process at the interface between the joint and the bonded substrates. However the difficulty of the numerical implementation of this method makes its use very complicated and limited to some simple cases.

In the other side, the numerical methods headed by the Finite Element Method (FEM) struggle with singular problems, where a mesh refinement is required in order to take into account singularities. Besides, in order to overcome this limitation, particular numerical methods have been introduced to deal with this difficulty. These methods allow to give a multi-scale analysis in discontinuous mediums as assembly structures, welded mediums, etc.. Two possibilities are given by these approaches: the first procedure consists in making a local analysis and then project the information about the singularity behaviour at the large scale. The homogenization and the Arlequin methods, described by (T.I. Zohdi, J.T. Oden and G.J. Rodin, 1996) and (P.T. Bauman and H. Ben Dhia, 2008) respectively, are based on this principle and are used in several works, especially to analyse composites materials. The second procedure consists in defining a correction of the classical FE method. The information about the singularity is then stored in an added part called “enrichment”. Thereby, the local analysis is not required. X-FEM (N. Moes, J. Dolbow and T. Belytschko, 1999; H. Bayesteh and S. Mohammadi, 2011) and G-FEM (I. Babuska, U. Banerjee and J. Osborn, 2004) are two partition of unity (PUM) (I. Babuska and J.M Melenk, 1999) methods that are used in several works in order to solve various types of problems especially crack problems.

2. STATE OF THE ART: METHODS USED FOR SINGULAR PROBLEMS

2.1. X-FEM approach:

Problems with singularities have been firstly treated using FE method by updating the mesh ‘topology’ in order to match the geometry of the singularity. However, the introduction of X-FEM circumvents this problem by enriching a standard approximation with special additional functions. The form of the enriched approximation follows the partition of unity. The geometry of the singularity is then involved by updating the enrichment scheme; no remeshing of the domain is required. The only interaction between the mesh and the geometry of the singularity involves the construction of the enriched basis functions. The classical X-FEM formulation is given by:

$$\mathbf{u}(\mathbf{x}) = \sum_{i}^{N} N_i(\mathbf{x})\mathbf{u_i} + \sum_{k}^{Nenr} N_k(\mathbf{x})\psi_k(\mathbf{x})\mathbf{b_k} \qquad (1)$$

Where *Nenr* is the enrichment terms number, $\mathbf{u}(\mathbf{x})$ the approximated function, $N_i(\mathbf{x})$ the standard FEM functions for node i, $\mathbf{u_i}$ the unknown of the standard FEM part at node i, ψ_k the enrichment function and $\mathbf{b_k}$ the enrichment parameter.

2.2. MAE approach:

MAE approach has been firstly used for fluid mechanics (M. Van Dyke, 1975). It shows its efficiency in treating perturbed problems and boundary layers. This method based on asymptotic assumptions has been applied for adhesive joint problems as illustrated in Figure 1.

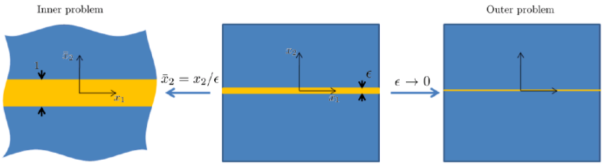


Figure 1. Two-scales partition of 'Matched Asymptotic Expansions' approach.

It consists in solving problems at two different scales (Figure 1) by introducing two different asymptotic expansions, the outer (Equation 2) and the inner (Equation 3) expansions.

$$\mathbf{u}^{\pm}(x_1,x_2) = \mathbf{u}^{0\pm}(x_1,x_2) + \varepsilon\mathbf{u}^{1\pm}(x_1,x_2) + \ldots \qquad (2)$$

$$\mathbf{v}^{\pm}(x_1,\bar{x}_2) = \mathbf{v}^{0\pm}(x_1,\bar{x}_2) + \varepsilon\mathbf{v}^{1\pm}(x_1,\bar{x}_2) + \ldots \qquad (3)$$

With $\bar{x}_2$ is the stretched variable $\bar{x}_2 = x_2/2$ and $(.)^{\pm}$ denotes the displacement values at the both sides of the interface ($\bar{x}_2 = x_2 = 0$).

Each expansion is constituted of two main parts: the leading terms $\mathbf{u}^0$ and $\mathbf{v}^0$ which are the classical solutions obtained by finite element method while the second part is considered as a correction given by a perturbed terms $\varepsilon\mathbf{u}^1$ and $\varepsilon\mathbf{v}^1$ depending on ε; the characteristic thickness of the brazed joint.

In the works of (D. Leguillon and R. Abdelmoulab, 2000; D.H. Nguyen et al., 2008), the MAE has been used to compute the jumps expressions of displacement field across the interface of discontinuity defined by the outer domain. This algorithm is based on simultaneous resolution of classical equations of the model (equilibrium equations, constitutive laws, continuity conditions) for each order, and a matching process (Equation 4) for both expansions at their respective limits (i.e. $x_2 \to 0$) for the outer expansion, and $\bar{x}_2 \to \pm\infty$ for the inner expansion).

$$\begin{cases} \lim\limits_{\bar{x}_2 \to \pm\infty} \left(\mathbf{v}^{0\pm}(x_1,\bar{x}_2) - \mathbf{u}^{0\pm}(x_1,0)\right) = \mathbf{0} \\ \lim\limits_{\bar{x}_2 \to \pm\infty} \left(\mathbf{v}^{1\pm}(x_1,\bar{x}_2) - \bar{x}_2\dfrac{\partial \mathbf{u}^{0\pm}}{\partial x_2}(x_1,0) - \mathbf{u}^{1\pm}(x_1,0)\right) = \mathbf{0} \quad (4) \\ \ldots \end{cases}$$

Using this algorithm, the jump **u** of displacement (D. Leguillona and R. Abdelmoulab, 2000; D.H. Nguyen et al., 2008) is computed and expressed by:

$$\mathbf{u} = \begin{cases} \varepsilon\left[u_1^{1+}(x_1,0) - u_1^{1-}(x_1,0)\right] = \varepsilon\dfrac{\mu_- - \mu_2}{\mu_2}\left[u_{2,1}^{0-}(x_1,0) + u_{1,2}^{0-}(x_1,0)\right] + \\ \qquad \varepsilon\dfrac{\mu_+ - \mu_2}{\mu_2}\left[u_{2,1}^{0+}(x_1,0) + u_{1,2}^{0+}(x_1,0)\right] \\ \varepsilon\left[u_2^{1+}(x_1,0) - u_2^{1-}(x_1,0)\right] = \varepsilon\left[\dfrac{\lambda_- - \lambda_2}{\lambda_2 + 2\mu_2}u_{1,1}^{0-}(x_1,0) + \left[\dfrac{\lambda_- + 2\mu_-}{\lambda_2 + 2\mu_2} - 1\right]u_{2,2}^{0-}(x_1,0)\right] + \\ \qquad \varepsilon\left[\dfrac{\lambda_+ - \lambda_2}{\lambda_2 + 2\mu_2}u_{1,1}^{0+}(x_1,0) + \left[\dfrac{\lambda_+ + 2\mu_+}{\lambda_2 + 2\mu_2} - 1\right]u_{2,2}^{0+}(x_1,0)\right] \end{cases} \qquad (5)$$

As it will be presented in the next section, this value has been used in the setup of the MAX-FEM model.

2.3. Proposed hybrid MAX-FEM Model

In this work, the hybrid MAX-FEM model has been introduced for the thin layers modelling to release mechanical analysis of a 2D-brazed joint. The MAX-FEM model is a Partition of Unity Method (PUM) where the main solution is formulated using two main parts: a classical FEM discretization and the enrichment terms. The proposed PUM formulation exploits the Matched Asymptotic Expansions (MAE) in the definition of the enrichment parameters. This procedure links the two main parts of the PUM formulation to give a corrected form of the standard FEM where correction matrix is introduced to compute the stiffness matrix.

3. *MAX-FEM SETUP*

3.1. *MAX-FEM Principle*

By analysing the shapes of the X-FEM and MAE approaches, it is notable that these two methods share the common subdivision of the unknown solution at two different parts: the classical solution and the correction terms (Figure 2). Thereby the MAX-FEM model feats the similarity between the two methods to define a new procedure describing the brazed joint behaviour. In fact, the joint is assimilated to a discontinuity interface which creates a jump of displacement field. The new configuration containing the introduced singularity is described using a specific MAX-FEM formulation while the enrichment parameters are defined using the jumps computed by the MAE approach (Equation 5).

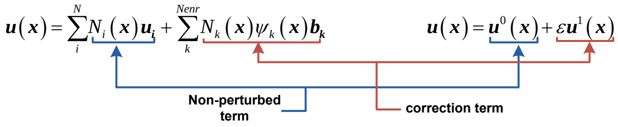


Figure 2. Identification of leading and correction terms using MAX-FEM model.

3.2. Problem position

The main purpose here is the setup of the MAX-FEM model for a 2D-brazed joint under mechanical loading (see Figure 3). The adhesive joint and substrates are considered as elastic domains.

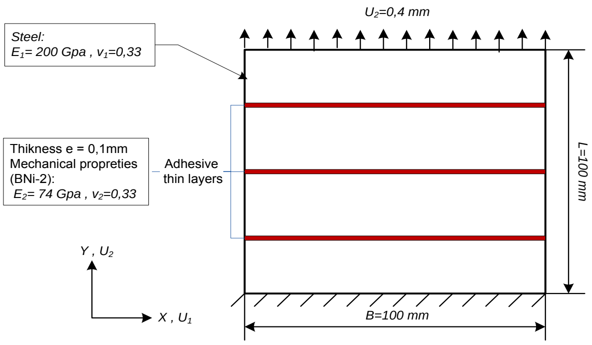


Figure 3. 2D brazed joint under mechanical loading

3.3. MAX-FEM formulation

The MAE outer expansion (Equation 2) crushes the thin layer. The latter is turned to an interface of discontinuity ($\varepsilon \to 0$) which completely crosses the width of the assembly (see Figure 1). As a result, the problem is assimilated to the strong discontinuity problem with presence of displacement jump across the interface. Herein, the MAX-FEM will consider the formulation (Equation 1) and adapts it to the considered application. In this context, the formulation proposed in equation (6) uses the same strategy as the outer expansion of MAE while keeping the basic shape of X-FEM:

$$\mathbf{u}(\mathbf{x})=\sum_i^N N_i(\mathbf{x})\mathbf{u}_i^0+\sum_k^{Nenr} N_k(\mathbf{x})H(\mathbf{x}_k)\mathbf{b}_k \qquad (6)$$

Where $\mathbf{b}_k$ and are the enrichment parameters which must be defined and H is the Heaviside function defined in the point $\mathbf{x}\begin{pmatrix}x_1\\x_2\end{pmatrix}$ by:

$$H(\mathbf{x})=\begin{cases}-1 & if \quad x_2 \prec 0\\ 1 & if \quad x_2 \geq 0\end{cases}$$

Instead of using nodal values of the global field $\mathbf{u}_i$ as used in formulation (Equation 1) and in the work of Nguyen et al. (2008), this approximation (Equation 6) integrates the continuous solution $\mathbf{u}^0$.

The similarity between the outer expansion (Equation 2) and formulation (Equation 6) is employed to identify the enrichment parameters $\mathbf{b}_k\begin{pmatrix}a_k\\b_k\end{pmatrix}$. The algorithm of identification is illustrated for linear structural element containing the joint position (Figure 4):

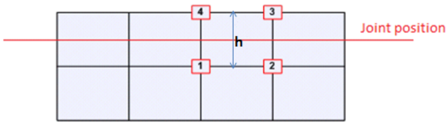


Figure 5. Linear structural enriched element containing the joint position

Equations (6) and (2) are used to compute the nodal values of the global fields (Equation 7) and (Equation 8) as follow:

Firstly, by using the formulation (Equation 6):

$$\begin{cases}u_1(\mathbf{x}_1)=u_1^0(\mathbf{x}_1)-a_1\\u_1(\mathbf{x}_2)=u_1^0(\mathbf{x}_2)-a_2\\u_1(\mathbf{x}_3)=u_1^0(\mathbf{x}_3)+a_3\\u_1(\mathbf{x}_4)=u_1^0(\mathbf{x}_4)+a_4\end{cases}\qquad\begin{cases}u_2(\mathbf{x}_1)=u_2^0(\mathbf{x}_1)-b_1\\u_2(\mathbf{x}_2)=u_2^0(\mathbf{x}_2)-b_2\\u_2(\mathbf{x}_3)=u_2^0(\mathbf{x}_3)+b_3\\u_2(\mathbf{x}_4)=u_2^0(\mathbf{x}_4)+b_4\end{cases}\qquad (7)$$

And secondly, by using MAE outer expansion (Equation 2):

$$\begin{cases}u_1(\mathbf{x}_1)=u_1^0(\mathbf{x}_1)+\varepsilon u_1^1(\mathbf{x}_1)\\u_1(\mathbf{x}_2)=u_1^0(\mathbf{x}_2)+\varepsilon u_1^1(\mathbf{x}_2)\\u_1(\mathbf{x}_3)=u_1^0(\mathbf{x}_3)+\varepsilon u_1^1(\mathbf{x}_3)\\u_1(\mathbf{x}_4)=u_1^0(\mathbf{x}_4)+\varepsilon u_1^1(\mathbf{x}_4)\end{cases}\quad\begin{cases}u_2(\mathbf{x}_1)=u_2^0(\mathbf{x}_1)+\varepsilon u_2^1(\mathbf{x}_1)\\u_2(\mathbf{x}_2)=u_2^0(\mathbf{x}_2)+\varepsilon u_2^1(\mathbf{x}_2)\\u_2(\mathbf{x}_3)=u_2^0(\mathbf{x}_3)+\varepsilon u_2^1(\mathbf{x}_3)\\u_2(\mathbf{x}_4)=u_2^0(\mathbf{x}_4)+\varepsilon u_2^1(\mathbf{x}_4)\end{cases}\qquad (8)$$

Then, the displacement difference between two nodes located on the same ridge and crossing the joint position is computed with the two approximations (Equation 7) and (Equation 8) as follow.

Firstly, by using the formulation (Equation 7):

$$\begin{cases}u_1(\mathbf{x}_4)-u_1(\mathbf{x}_1)=u_1^0(\mathbf{x}_4)-u_1^0(\mathbf{x}_1)+a_4+a_1\\u_1(\mathbf{x}_3)-u_1(\mathbf{x}_2)=u_1^0(\mathbf{x}_3)-u_1^0(\mathbf{x}_2)+a_3+a_2\end{cases}$$
$$\begin{cases}u_2(\mathbf{x}_4)-u_2(\mathbf{x}_1)=u_2^0(\mathbf{x}_4)-u_2^0(\mathbf{x}_1)+b_4+b_1\\u_2(\mathbf{x}_3)-u_2(\mathbf{x}_2)=u_2^0(\mathbf{x}_3)-u_2^0(\mathbf{x}_2)+b_3+b_2\end{cases}\qquad (9)$$

Secondly, by using MAE outer expansion (Equation 8):

$$\begin{cases} u_1(\mathbf{x_4})-u_1(\mathbf{x_1})=u_1^0(\mathbf{x_4})-u_1^0(\mathbf{x_1})+\varepsilon(u_1^1(\mathbf{x_4})-u_1^1(\mathbf{x_1})) \\ u_1(\mathbf{x_3})-u_1(\mathbf{x_2})=u_1^0(\mathbf{x_3})-u_1^0(\mathbf{x_2})+\varepsilon(u_1^1(\mathbf{x_3})-u_1^1(\mathbf{x_2})) \end{cases}$$
$$\begin{cases} u_2(\mathbf{x_4})-u_2(\mathbf{x_1})=u_2^0(\mathbf{x_4})-u_2^0(\mathbf{x_1})+\varepsilon(u_2^1(\mathbf{x_4})-u_2^1(\mathbf{x_1})) \\ u_2(\mathbf{x_3})-u_2(\mathbf{x_2})=u_2^0(\mathbf{x_3})-u_2^0(\mathbf{x_2})+\varepsilon(u_2^1(\mathbf{x_3})-u_2^1(\mathbf{x_2})) \end{cases} \quad (10)$$

The equality between Equations (9) and (10) gives the conditions that the enrichment parameters have to satisfy:

$$\begin{cases} a_4+a_1=\varepsilon(u_1^1(\mathbf{x_4})-u_1^1(\mathbf{x_1})) \\ a_3+a_2=\varepsilon(u_1^1(\mathbf{x_3})-u_1^1(\mathbf{x_2})) \end{cases}$$
$$\begin{cases} b_4+b_1=\varepsilon(u_2^1(\mathbf{x_4})-u_2^1(\mathbf{x_1})) \\ b_3+b_2=\varepsilon(u_2^1(\mathbf{x_3})-u_2^1(\mathbf{x_2})) \end{cases} \quad (11)$$

To compute these parameters values, it is not necessary to have an explicit expression of the perturbed term $\mathbf{u^1}$ in each node. In fact, by using the MAE algorithm introduced in works (D. Leguillona and R. Abdelmoulab, 2000; D.H. Nguyen et al., 2008) the term $\varepsilon(\mathbf{u^1}(\mathbf{x_4})-\mathbf{u^1}(\mathbf{x_1}))$ for nodes 1 and 4 (respectively $\varepsilon(\mathbf{u^1}(\mathbf{x_3})-\mathbf{u^1}(\mathbf{x_2}))$ for nodes 2 and 3) are approximated to $\varepsilon(1-h)\ \mathbf{u^1}\ (x_1^{1-4},0)$ and $\varepsilon(1-h)\ \mathbf{u^1}\ (x_1^{2-3},0)$, respectively. h is the height of the enriched element while $\mathbf{u^1}\ (x_1^{1-4},0)$ (respectively $\mathbf{u^1}\ (x_1^{2-3},0)$) denotes the displacement jump crossing the interface in the common abscissa between nodes 1 and 4 (respectively 2 and 3).

Taking into account the previous approximation, the expression proposed for the parameters $\mathbf{b_k}$ is given bellow:

$$\mathbf{b_k}=\begin{cases} \dfrac{\varepsilon}{2}(1-h)\ \mathbf{u^1}(x_1^{1-4},0) & for\ k=1\ ,4 \\ \dfrac{\varepsilon}{2}(1-h)\ \mathbf{u^1}(x_1^{2-3},0) & for\ k=2\ ,3 \end{cases} \quad (12)$$

3.4. Stiffness matrix

The expression (Equation 5) shows a linear dependence between the jump and the leading derivative terms. From this dependence rises a transfer matrix between the jump and the unperturbed strain vector:

$$\mathbf{u^1}=\begin{Bmatrix} u_1^1 \\ u_2^1 \end{Bmatrix}=\begin{bmatrix} 0 & 0 & A \\ B & S & 0 \end{bmatrix}\begin{Bmatrix} u_{1,1}^0(x_1,0) \\ u_{2,2}^0(x_1,0) \\ u_{1,2}^0(x_1,0)+u_{1,2}^0(x_1,0) \end{Bmatrix}=[\mathbf{C}]\{\varepsilon^0\} \quad (13)$$

$\{\varepsilon^0\}$ is the leading strain vector, $[\mathbf{C}]=\begin{bmatrix} 0 & 0 & A \\ B & S & 0 \end{bmatrix}$, $A=\dfrac{\mu_{..}-\mu_2}{\mu_2}$, $B=\dfrac{\lambda-\lambda_2}{\lambda_2+2\mu_2}$ and $S=\dfrac{\lambda+2\mu}{\lambda_2+2\mu_2}-1$.

By integrating this result in the enrichment parameters expression, the latters are linked to the leading term nodal values as shown in (Equation 14).

$$\mathbf{b_k}=\begin{cases} \dfrac{\varepsilon}{2}(1-h)[\mathbf{C}][\mathbf{B}]\left(x_1^{1-4},0\right)\{u^0\} & for\ k=1\ ,4 \\ =\dfrac{\varepsilon}{2}(1-h)[\mathbf{C}][\mathbf{B}]\left(x_1^{2-3},0\right)\{u^0\} & for\ k=2\ ,3 \end{cases} \quad (14)$$

[C] is the enrichment matrix defined in (Equation 13) and **[B]** is the strain matrix linking the strain vector to the nodal values of the leading term.

The expression (Equation 14) is injected in the X-FEM formulation (Equation 6) creating then a transfer matrix between the global nodal displacement values $\{u\}$ and the leading term one $\{u^0\}$:

$$\{u\}=\left([\mathbf{I}]_{\mathbf{8,8}}+\frac{1}{2}[\mathbf{H}][\mathbf{C}][\mathbf{B}](x_1,0)\right)\{u^0\} \quad (15)$$

$[\mathbf{I}]_{\mathbf{8,8}}$ is the identity matrix and $[\mathbf{H}]$ is called “Heaviside matrix” and expressed below:

$$[\mathbf{H}]={}^t\begin{bmatrix} H(\mathbf{x}_1) & 0 & H(\mathbf{x}_2) & 0 & H(\mathbf{x}_3) & 0 & H(\mathbf{x}_4) & 0 \\ 0 & H(\mathbf{x}_1) & 0 & H(\mathbf{x}_2) & 0 & H(\mathbf{x}_3) & 0 & H(\mathbf{x}_4) \end{bmatrix}$$

By considering that the global problem and the leading outer one are assumed under the same exterior loading ($\{F\}=\{F^0\}$), the global stiffness matrix $[\mathbf{K}]$ is linked to the one of the leading problem $[\mathbf{K^0}]$ as follow:

$$[\mathbf{K}]=\int_{\Omega}{}^t[\mathbf{B}][\mathbf{D}][\mathbf{B}]\left([\mathbf{I}]_{\mathbf{8,8}}+\frac{1}{2}[\mathbf{H}][\mathbf{C}][\mathbf{B}](x_1,0)\right)^{-1}d\Omega \quad (16)$$

Consequently, the global stiffness matrix can be defined as a correction of the standard FEM stiffness matrix using a “correction matrix” $[\mathbf{CM}]$ defined by:

$$[\mathbf{CM}]=\left([\mathbf{I}]_{8,8}+\frac{1}{2}[\mathbf{H}][\mathbf{C}][\mathbf{B}](x_1,0)\right)^{-1} \tag{17}$$

Finally, the linear system to consider is:

$$[\mathbf{K}]\{u\}=\{F\} \tag{18}$$

The enrichment parameters $\mathbf{b_k}$ and the leading term $\mathbf{u^0}$ of the finite elements formulation (Equation 6) do not appear explicitly in the final system (Equation 18). This is due to the MAE results which reveal a linear dependence liking the enrichment parameters to the leading term derivatives. Also, the continuity of the finite elements formulation (Equation 6) allows to define a correction matrix $[\mathbf{CM}]$ (Equation 17) and to link the global stiffness matrix $[\mathbf{K}]$ to the standard one. The computation of the leading term $\mathbf{u^0}$ and the enrichment parameters is then not needed, and the global displacement field $\mathbf{u}$ is directly computed by numerical resolution of the system (Equation 18). It can be noted that only the correction matrix $[\mathbf{CM}]$ is computed and then injected in a standard FEM program.

The obtained solution covers the outer domains. This means that only the influence of the thin layer on the global structure is characterized. In this work, the local analysis of the thin layer is not considered. However, if necessary, the model can take into account the complex behavior in the vicinity of the joint, including crack, delamination or other nonlinear aspects for example. This can be possible by resolving the inner problem (Equation 3) of the MAE. Following this approach, the local analysis is restricted in the enriched elements which are expanded to zoom on the joint and the contact interfaces. The resolution of the inner problem can be performed by using the existing methods (FEM, X-FEM, etc.)

4. *NUMERICAL IMPLEMENTATION OF MAX-FEM MODEL*

This section provides a validation study for the proposed MAX-FEM hybrid model. The obtained results are compared to those given by the standard FEM by meshing finely the thin layer. In this work, ABAQUS code is used to implement the model using an UEL subroutine. The same code is also used to compute the reference solution (FEM) for 2D-brazed joint in figure 3

4.1. *Numerical implementation*

The model developed here has been implemented in ABAQUS as a user element using the UEL subroutine. The procedure of the development of this subroutine is based on a previous work introduced by Giner et al. (2009). The model starts from a combination of X-FEM and MAE, and will give as a final result the corrected form of standard FEM. This correction is presented by the matrix $[\mathbf{CM}]$ (Equation 17) which stores the required information for the joint behavior. As a result, the structure of the program (Figure 4) will be constituted by the standard FEM while the correction matrix will be computed in an internal subroutine depending on the considered application.

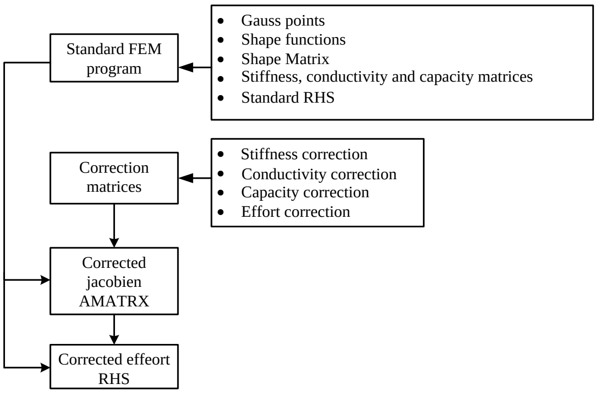


Figure 5. Algorithm structure of the implemented UEL "MAX-FEM" subroutine

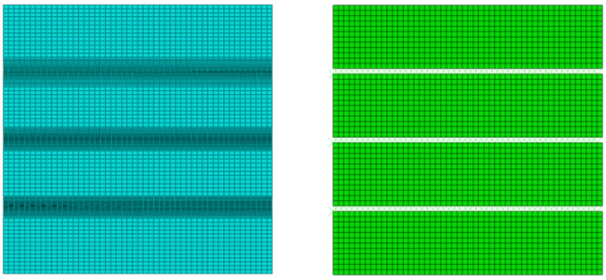
The results given in the setup of the model are implemented in the UEL "MAX-FEM" subroutine following the structure presented in Figure 6. The correction matrix $[\mathbf{CM}]$ (Equation 17) and the corrected stiffness matrix (Equation 16) are computed for the 2D-brazed joint presented in Figure 3.

The computation is limited to 2D plane stress state. Thereby, the enrichment matrix $[\mathbf{C}]$ (Equation 13) needed for the computation of the correction matrix $[\mathbf{CM}]$ (Equation 26) is written as:

$$[\mathbf{C}]=\varepsilon\left(\frac{E_1}{E_2}-1\right)\begin{bmatrix}0 & 0 & 1\\ \nu & 1 & 0\end{bmatrix} \tag{19}$$

By resolving the system (Equation 18), the displacement field is obtained. Then, other mechanical fields are deducted. The results given by the UEL "MAX-FEM" subroutine are compared to ones obtained by standard FEM under the same code. The considered mesh, displacement, stress, strain and error are presented for the two methods.

(i) Mesh



FEM MAX-FEM

Figure 6. Overview of the classical FEM discretization and the MAX-FEM meshing

User elements are not represented by ABAQUS post-processor. However, in order to reach this aim, a post-processing subroutine should be required. This will be performed in forthcoming developments.

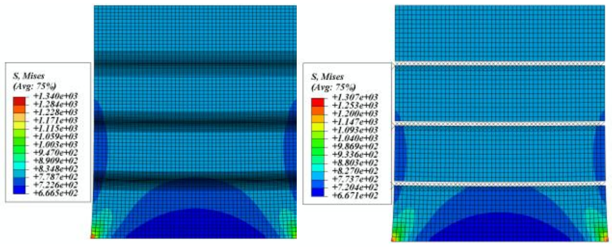
The comparison between the two mesh strategies is summarized in Table 1:

Table 1. FEM and MAX-FEM mesh data

	FEM	MAX-FEM	Ratio (FEM/MAX-FEM)
Nodes number	7242	2652	2,73
Elements number	7052	2550	2,76
Minimum element size	0,1 mm	2 mm	20

From Table 1, it can be noted that the model presents an important mesh optimization. This should be more notable in forthcoming transient analyses where the time increment for such computations depends on the size of elements.

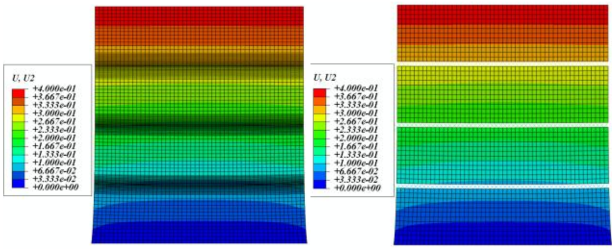
(ii) Stress



FEM MAX-FEM

Figure 7. FEM and MAX-FEM stress distributions

(iii) Displacement



FEM MAX-FEM

Figure 8. FEM and MAX-FEM displacements

By analyzing these results, it can be noted that the model reproduces the outer displacements and stress fields given by the FEM method. In order to give an accurate comparison, a plot of the displacements in the section *x = 20 mm* are presented in Figure 13:

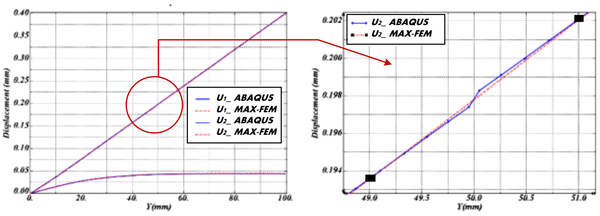


Figure 9. FEM and MAX-FEM displacements in the section x = 20 mm

Figure 8 shows that the curve given by the model fits the one obtained by standard FEM from the first enriched node. Consequently, a single MAX-FEM element reproduces the same solution given by several FEM elements. The line linking the two enriched nodes is a linear interpolation and does not present the solution in the vicinity of the joint. The latter can be obtained by the resolution of the inner problem of MAE.

To analyze the accuracy of the model, three error expressions are computed taking the FEM solution as reference:

- Relative error: $err_{relative} = \dfrac{\left|u^{i}_{FEM} - u^{i}_{MAX-FEM}\right|}{u^{i}_{FEM}}$
- Maximum error: $err_{\max} = \max\left(\left|u^{i}_{FEM} - u^{i}_{MAX-FEM}\right|\right)$
- Norm L^2: $\|err\| = \sqrt{\sum \left(u^{i}_{FEM} - u^{i}_{MAX-FEM}\right)^2}$

These expressions are used to estimate the error into the section *x = 20 mm* by using the previous results of Figure 8

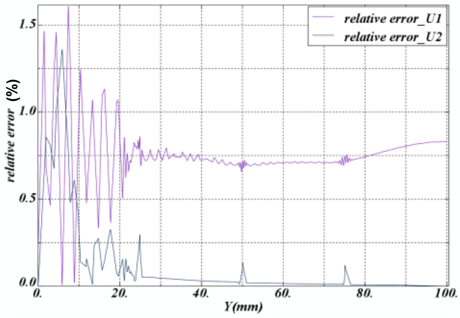


Figure 10. Relative error in the section x = 20 mm

Table 2. error approximation

	$\max(err_{relative})$	$err_{\max}$	$\|err\|$
U1	1,6 %	0,0039	1,339 10-7
U2	1,36 %	0,00128	1,295 10-7

By analyzing these error approximations, it becomes trivial that the hybrid model presents accurate results without meshing the thin layer. However, the accuracy of the model depends on several parameters and the model can have some limitations and critical points.

Among these limitations, there are boundary problems which affect the results accuracy near to boards. These problems are especially due to the MAE method. The rules used to link the outer and inner solutions still available in an infinite inner domain. However, in the boards, the solution has to satisfy the boundary conditions. The enrichment strategy used in this work cannot then reproduce the solution in the vicinity of boards.

In addition to board problems, other parameters may affect the model accuracy. The three points below are the most important:

1. The ratio h/e linking the element height and the joint thickness. Actually, more this ratio is smaller; more the results highlight a good accuracy.
2. The number of enriched elements: in this work, a single element is enriched in the joint thickness direction. However, it is trivial that increasing that number will increase the model accuracy. Besides, a new enrichment strategy has to be developed.

5. *CONCLUSION*

The hybrid model introduced in this work combines the MAE technique which is used to define the enrichment parameters and the X-FEM formulation. This new approach called 'MAX-FEM' model has been established and used to solve the mechanical singular problem of two-dimensional brazed joints. The 'MAX-FEM' approach leads to a corrected form of standard FEM where a correction matrix is used to compute the main solution in the whole structure.

Once set up, the model has been implemented under ABAQUS code using the UEL subroutine. The results given by the model are compared to those obtained using standard FEM for the 2D- brazed joint. From this application, it is notable that "MAX-FEM" model provides accurate results in terms of displacement and temperature fields without any required mesh refinement in the vicinity of the thin layer.

However, the model has its limitations near to boards where the proposed enrichment is not available. New enrichment strategies have to be developed in future work to deal with this limitation. Also, even if it is not developed here, the model can take into account other behaviors of the joint such as damage, cracks and delamination. These behaviors have to be considered in the computation of jumps using MAE approach. Besides, if it is necessary, the local behavior of the joint can be described by the resolution of the inner problem. Thereby the analysis can be developed in two scales instead of the global analysis presented in this work. Finally, it can be said that the model needs to be enhanced and generalized to be used for more complicated applications.

REFERENCES

I. Babuska, J.M Melenk, *The partition of unity finite element method*, Int. J.Numer.Meth.Engrg.40 (1997) 727-758

I. Babuska, U. Banerjee, J. Osborn, *Generalized finite element methods: Main ideas, results, and perspective*, Int. J. Comput. Methods 1 (1) (2004) 1-37

P.T. Bauman, H. Ben Dhia, N. Elkhodja, J.T. Oden, S. Prudhomme, *On the application of the Arlequin method to the coupling of particle and continuum models*, Comput. Mech., 42 (2008) 511–530.

H. Bayesteh, S. Mohammadi, *XFEM fracture analysis of shells: The effect of crack tip enrichments*, Computational Materials Science, 50 (2011) 2793–2813

Giner E, Sukumar N, Tarancon JE, Fuenmayor FJ. *An Abaqus implementation of the extended finite element method*. Eng Fract Mech., 76 (2009) 347-368.

D. Leguillona, R. Abdelmoulab, *Mode III near and far fields for a crack lying in or along a joint*, International Journal of Solids and Structures, 37 (2000) 2651-2672

N. Moes, J. Dolbow, T. Belytschko, *A finite element method for crack growth without remeshing*, Int. J. Numer. Meth. Engrg. 46 (1999) 131–150.

D. H. Nguyen, F. Bilteryst, M. Lazard, P. Lamesle, G. Dour, *Coupling of the eXtended Finite Element Method and the matched asymptotic development in the modelling of brazed assembly*, International Journal of Material Forming, 1, 2008, pp. 1119-1122

P. Schmidt, *Modeling of adhesively bonded joints by an asymptotic method*, International Journal of Engineering Science, 46 (2008) 1291–1324

M. Van Dyke, *perturbation methods in fluid Mechanics* (1975)

T.I. Zohdi, J.T. Oden, G.J. Rodin, *Hierarchical modeling of heterogeneous bodies*. Computer Methods in Applied Mechanics and Engineering, 138 (1996) 273-298,

A STUDY ON THE ECMS PARAMETER ADAPTION FOR THE DRIVER CHARACTERICSTIC VARIATION

Joohee Son[a], Taeho Park[b], Kanghee Won[c], Hyeongcheol Lee*[d]

[a] Department of Electric Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 134-791, Korea
[b] Department of Electric Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 134-791, Korea
[c] Department of Electric Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 134-791, Korea
[d] Division of Electrical and Biomedical Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 134-791, Korea

[a]sonjhsla@hanyang.ac.kr, [b]koreapow@hanyang.ac.kr, [c]wkh@hanyang.ac.kr, [d]*hclee@hanyang.ac.kr

ABSTRACT

This paper proposes a novel supervisory control method for a parallel hybrid electric vehicle (PHEV) with a transmission mounted electric drive (TMED). An equivalent consumption minimization strategy (ECMS) is the supervisory control method and provides real-time sub-optimal energy management decisions by minimizing the “equivalent” fuel consumption of a hybrid electric vehicle (HEV). The equivalent fuel consumption is a combination of the actual fuel consumption an electrical energy use, and an equivalence factor is used to convert electrical power used into an equivalent chemical fuel quantity. In this study, the proposed ECMS parameter adaption focused on the driver characteristic variation. In the stage of development, the longitudinal driver model is developed and the represented driving patterns are defined. Results obtained in this research clarify the causal connection between the driver characteristic and the equivalence factor as the initial step of the adaptive ECMS implementable into microcontroller. The simulation results show that optimizing the control parameter is needed as the driver characteristic variation.

Keywords: Adaptive ECMS, Equivalence factor, Driving pattern, Driver model

1. INTRODUCTION

The advanced researches about the supervisory control algorithm of the HEV are heading to the parameter adaptation algorithm for global optimality. A research adapting the equivalence factor of the ECMS using past driving data has been done where the computational effort is too high. Another research proposed a pattern recognition algorithm to identify the drive cycle. However, that is insufficient to isolate the different driver on almost same drive-cycle. The purpose of this study is to clarify the causal connection between the driver characteristic and the equivalence factor as the initial step of the adaptive ECMS implementable into microcontroller.

Previous adaptive ECMS algorithm requires enormous computational effort to calculate the equivalence factor for stored driving data of past hundred seconds. However, the proposed algorithm calculates the combination of pre-calculated equivalence factor in offline so additional computation effort is only occurred for identifying the driver type and the drive cycle. Moreover, there was no research about analyzing the tendency of the equivalence factor according to the driver characteristic

In this paper, a forward simulator for the TMED type HEV is constructed based on the Cruise® and Simulink® including custom longitudinal driver model to imitate the characteristic of the actual driver. As a simulation group, some representative driver types are defined by the parameter set of the driver model. Each equivalence factor of these driver types for some drive cycles is calculated according to the conventional ECMS researches for the parallel HEV. These results are carefully analyzed in terms of the tendency of the equivalence factors. Operating points of actuators, the engine and the electric motor, are also depicted. Finally, the fuel economy of the conventional ECMS algorithm is simulated for the various combinations of the driving patterns and the equivalence factor to show the effectiveness of proposed adapting algorithm.

This paper introduces the target vehicle for the simulation (Chapter 2), and explains the used HCU (Chapter 3). Continually, the developed longitudinal driver model (Chapter 4) and the simulation result for the fuel economy (Chapter 5) is analyzed. Finally, the last chapter is the conclusion of this paper.

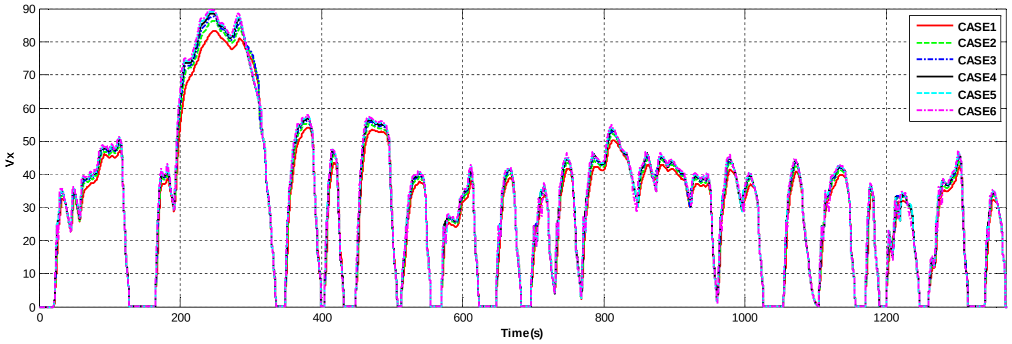
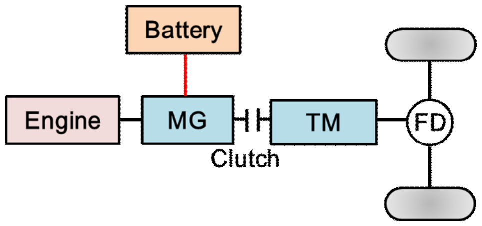


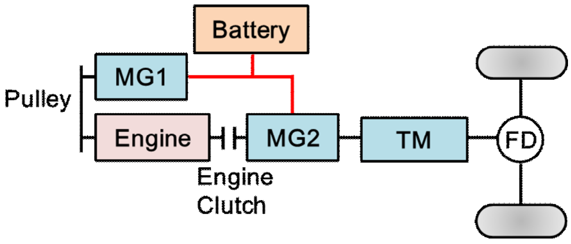
Figure 1: Example of the different 6 drivers driving almost same drive-cycle (FTP72: see the chapter 5.1)

2. TARGET VEHICLE

In this paper, target vehicle is the PHEV with the TMED type. This chapter introduces the configuration of the used PHEV with TMED type. There are many possibilities of configurations in a parallel hybrid powertrain. However, two different configurations are applied to the most of the parallel HEVs in production, which are shown in Fig. 2. The TMED can separate the traction motor (MG2) from the engine by the engine clutch. Therefore, the TMED can provide the electric vehicle mode which the practical parallel hybrid powertrain cannot support due to the engine friction. The TMED also offers a number of advantages such as flexibility to mix and match different sized electric motors and transmissions to suit different vehicles, and utilizing an existing off-the-shelf transmission. However, due to the additional complexity and the increased degree of freedom in the energy flow of the hybrid powertrain, the TMED requires more complicated and subtle supervisory control to take advantage of the advanced configuration. In general, the achievable improvements of fuel economy in HEVs depend strongly on the implemented energy management strategy, which is major part of the supervisory control. The next chapter will explain the HCU applied optimal supervisory control algorithm for a TMED in the sense of minimization of fuel usage.



(a) Practical parallel hybrid powertrain



(b) Transmission mounted electric drive (Target Vehicle configuration)

Figure 2: Parallel hybrid powertrain configuration

3. HYBIRD CONTROL UNIT (HCU)

In this chapter, The HCU and the ECMS algorithm are explained. The HCU is main controller to control the vehicle, in which energy management strategy that is the ECMS minimizing the equivalent fuel usage is implemented.

3.1. HCU

The hybrid control unit (HCU) is the main controller of the HEV which keeps the optimal driving condition as controlling each subsystem by observing the driving states of the vehicle for the optimal driving. Additionally, the subsystems are engine control unit (ECU), motor controller unit (MCU), transmission controller unit (TCU), battery management system (BMS) and voltage DC-DC converter (LDC). To be more concrete, the roles of the HCU are various and about 11s. First, starting a HEV with motor when turning ignition key or auto-stop or operating the starter-motor when no motor admitted, the next, assisting the engine torque by motoring when accelerating a HEV. Third, controlling the ratio of the transmission according to driving states, fourth, storing the electric energy by generation of the motor (regenerative breaking), fifth, stopping the engine when stopping with breaking after driving setting the D-range (auto-stop) or restarting the engine when non-zero accelerator pedal or zero decelerator pedal. Sixth, controlling the sliding at the slope, seventh, permitting the fuel cut and inject or not according to state of charge (SOC) and the ratio of the shift gears. Eighth, preventing the overcharging of the battery and limiting the motor torque. Ninth, control of the booster pressure, tenth, on/off control, generating of the LDC and voltage control. The last, inducing the eco-driving. Figure 3 is the block diagram of the HCU and subsystems.

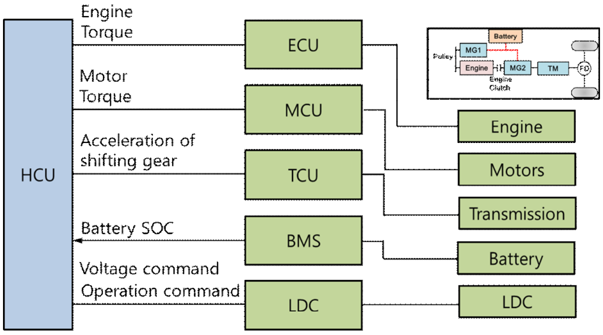


Figure 3: Hybrid control unit (HCU)

3.2. ECMS

A real-time control strategy based on an instantaneous optimization needs a definition of the cost function to be minimized at each instant. Such a function has to depend only upon the system variables at the current time. Since the main control goal is the minimization of the fuel consumption, it is clear that this quantity has to be included in the cost function. However, based on the requirements of electrical self-sustainability, the variations in the stored electrical energy (or state-of-charge, SOC) have to be taken into account as well. To deal with such aspects, various approaches have been proposed in this area. One-approach was used in the recently paper. It consists of evaluating the instantaneous cost function as a sum of the fuel consumption and an equivalent fuel consumption related to the SOC variation equivalent consumption minimization strategy (ECMS). In this case, it is clearly recognized that the electrical energy and the fuel energy are not directly comparable, but an equivalence factor is needed. The equivalence between electrical energy and fuel energy is basically evaluated by considering average energy paths leading from the fuel to the storage of electrical energy. If the overall efficiencies of the electrical and thermal paths were rigorously constant, such an equivalence would be theoretically exact. Since efficiencies vary with the operating point, this approach only allows the use of average values.

In the real-time control strategy, the equivalent fuel consumption is evaluated under the assumption that every variation in the SOC will be compensated in the future by the engine running at the current operating point. The equivalent fuel consumption therefore changes both with the operation point and with the power split control, and its evaluation requires an additional, inner loop in the instantaneous optimization procedure, or the prior storage of the results in a look-up table.

A control of the ECMS is presented at Sciarreta paper. This paper has the similar ECMS method. It is based on a new method for evaluating the equivalence factor between fuel and electrical energy. This method does not require the assumption of the average efficiencies of the parallel paths, and it is based on a coherent definition of system self-sustainability. The ECMS is valid for different system architectures and types of machines involved. The advantage includes good control performance with comparing that obtained with conventional control strategies, the robustness with respect to the variation of control parameters, and the system behavior in case of steady operating point.

4. LONGITUDINAL DRIVER MODEL

The longitudinal driver model is developed for imitating the real driver behavior. This driver model presents three driver tendency, flexibility, sensitivity and competence. Also, PI controller control the accelerator pedal (AP) and braking pedal (BP), and making the shape of AP, BP by driving maneuver mode.

Chapter 4.1 explains the behavior of the driver model and chapter 4.2 introduces the representative driving pattern in more detail.

4.1. The behavior of the driver model

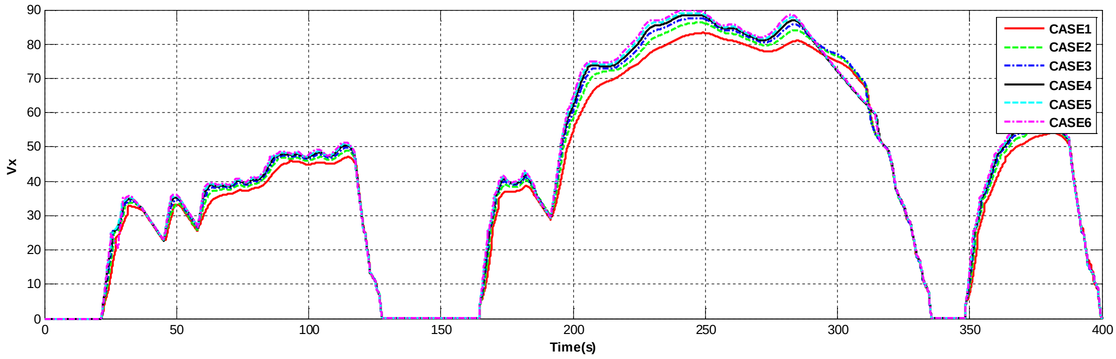
Figure 4 is the developing driver model. The driver model follows the reference driving cycle and makes the shape of AP signal (APS), BP signal (BPS). The red box of the figure 4 is reference driving cycle with FTP-72 speed profile. This model inputs are the vehicle speed, APS, BPS which have the blue line. The main control parameters have green, yellow, blue color box in the figure 4. The green box is making the APS delay (apdel). The figure 4 (g) depicts the vehicle speed graph changing the apdel. The behavior of the driver model is similar to competence of the drivers. The blue boxes are controller gains about AP, BP. Increasing the AP P_Gain (pa) effects the slope of the APS. This time, the behavior of the driver model is similar to flexibility of the drivers. Figure 4 (b) is APS example of the different flexibility. The yellow boxes are Bounds. The bounds related AP, BP control timing and vehicle speed error is reference. Bound1 is hb which decides activating AP, BP. If the hb is large amplitude, speed error is larger. Bound2 is lb which decides deactivating AP and Bound3 is blb which decides deactivating BP. Figure 4 (e) presents the APS graph of changing the lb. Moreover, driving maneuver mode is made with SR_latch and integrator reset designs the anti-winding.

4.2. The representative driving pattern

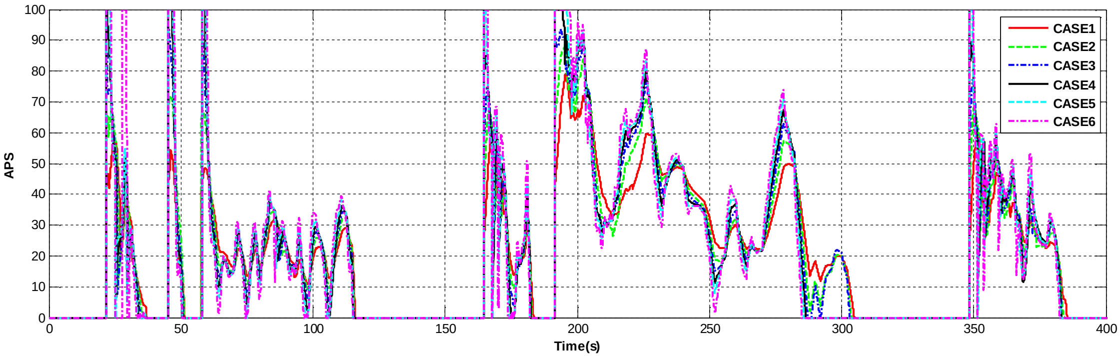
The driving pattern is defined as three. This paper assumes that various driver's characteristic are classified according to the flexibility, the sensitivity, the competence. These each patterns are orthogonal. Therefore, the each parameters of the driver model describes one pattern. Changing the pa describes Flexibility, changing the lb describes Sensitivity and changing the apdel describes competence. In this paper, parameter sets of the table 1 are used and its simulation results are figure 4.

Table 1: The parameter sets for the driver types used in the simulation

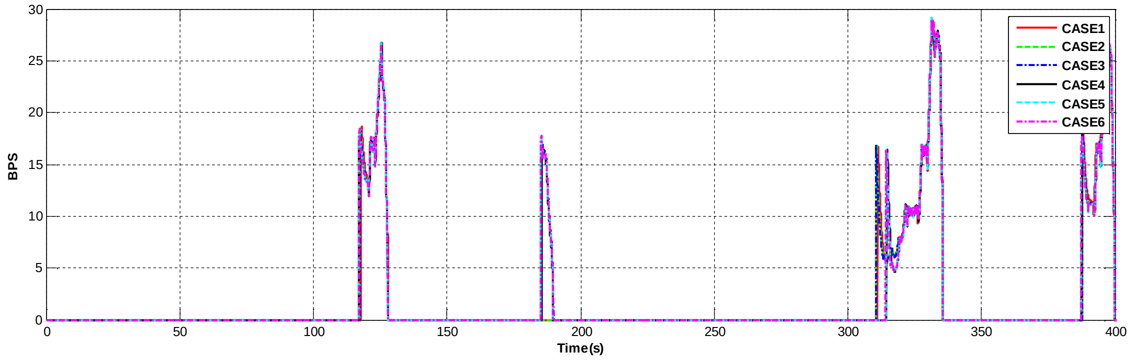
Flexibility	pa	lb	apdel
Case1	5	-1	0
Case2	8	-1	0
Case3	11	-1	0
Case4	15	-1	0
Case5	22	-1	0
Case6	30	-1	0
Sensitivity	pa	lb	apdel
Case1	20	-5	0
Case2	20	-1	0
Case3	20	1	0
Case4	20	2	0
Case5	20	3	0
Case6	20	5	0
Competence	pa	lb	apdel
Case1	20	-1	0
Case2	20	-1	0.5
Case3	20	-1	1
Case4	20	-1	1.5
Case5	20	-1	2
Case6	20	-1	2.5



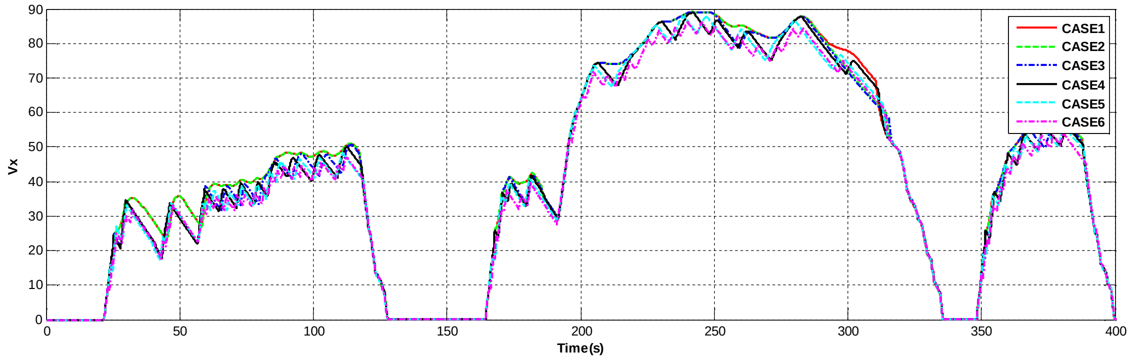
(a) Vehicle Speed (Flexibility)



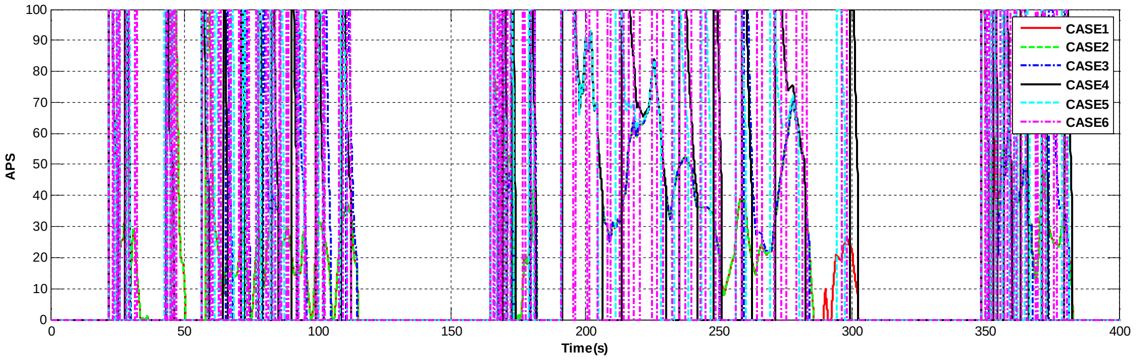
(b) APS (Flexibility)



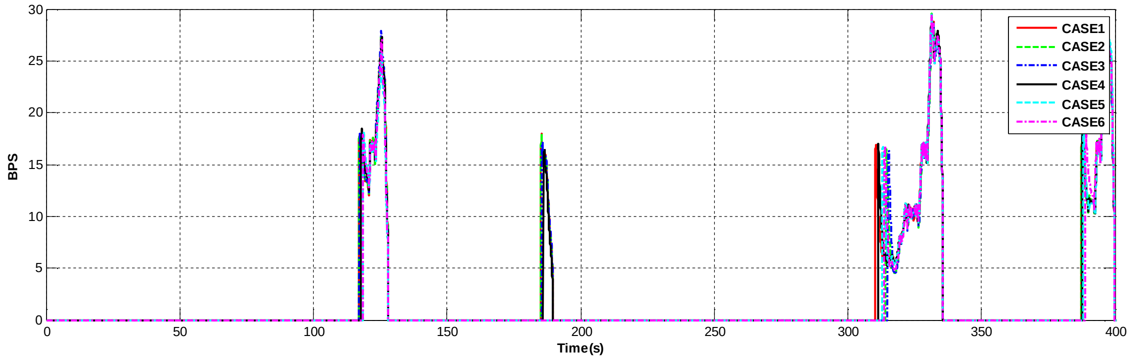
(c) BPS (Flexibility)



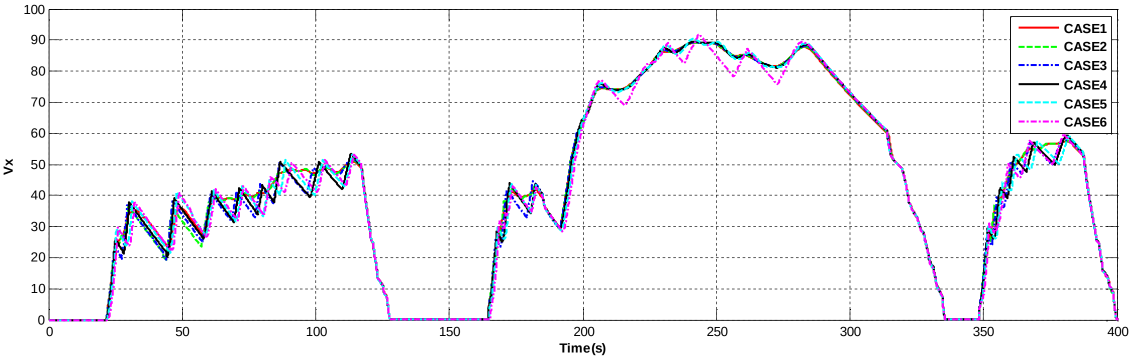
(d) Vehicle Speed (Sensitivity)



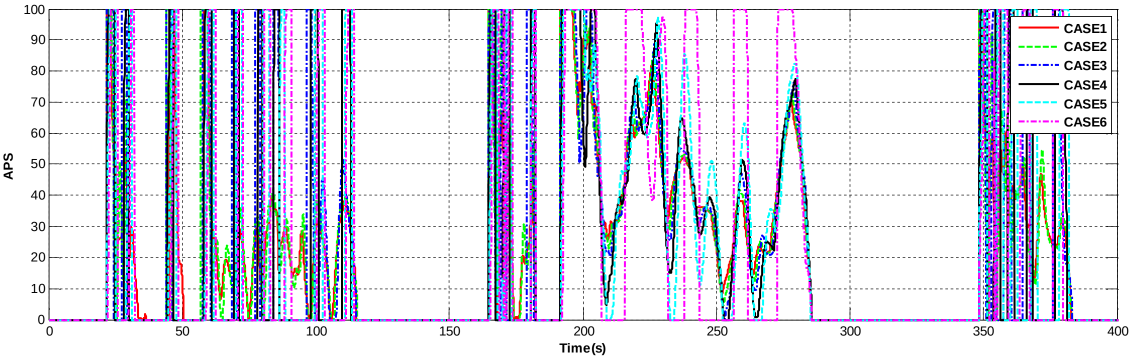
(e) APS (Sensitivity)



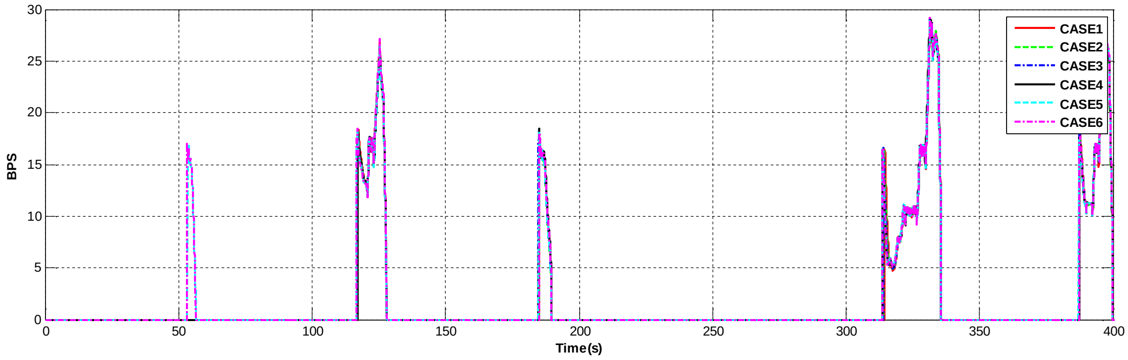
(f) BPS (Sensitivity)



(g) Vehicle Speed (Competence)



(h) APS (Competence)



(i) BPS (Competence)

Figure 4: Vx, APS, BPS of the each patterns (used parameter in the table 1)

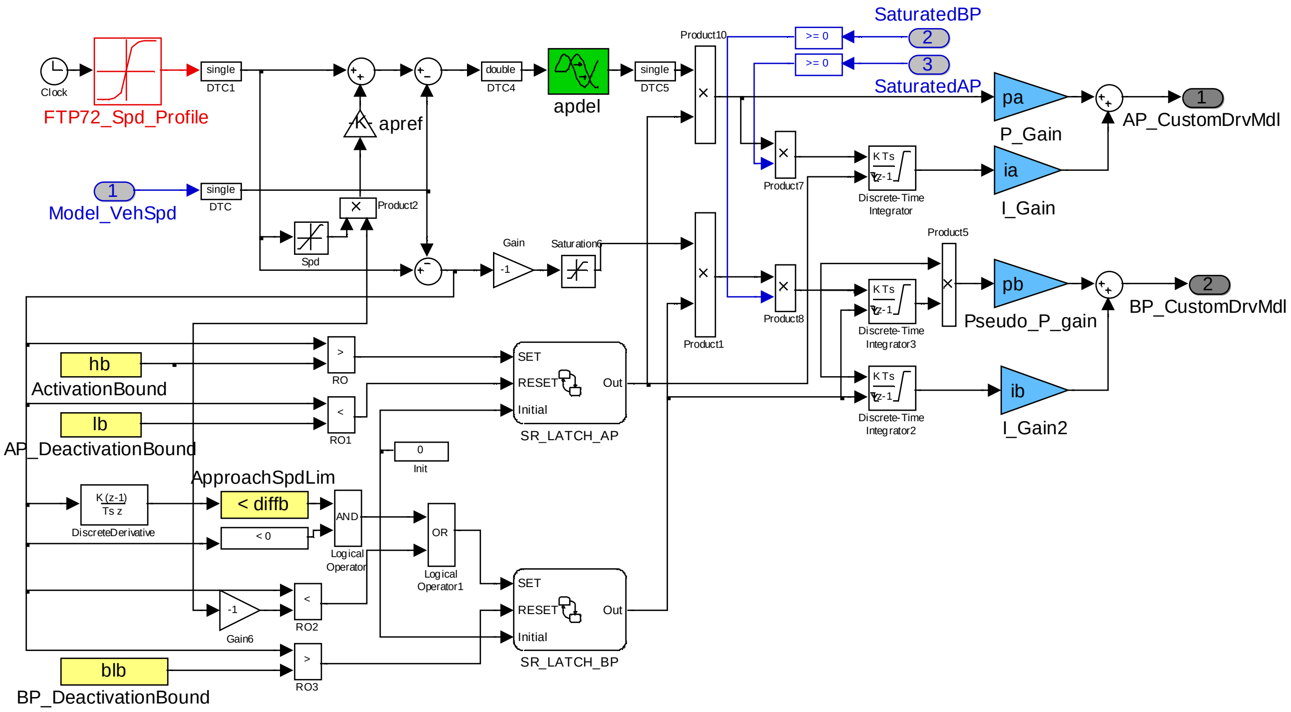


Figure 5: Longitudinal Driver Model

5. SIMULATION

The proposed Advanced HEV supervisory control algorithm adapts the equivalence factor according to driver characteristic variation. The different driving characteristic represents the driving pattern and driver type (Case1 ~ Case6) according to changing parameter pa, lb, apdel.

Chapter 5.1 depicts the simulation environment applied defined driving pattern, Chapter 5.2 shows the simulation results.

5.1. Simulation environment

The co-simulation environment configures with aforementioned (Chapter 2, 3) target vehicle and HCU control algorithm in the figure 5. The developed target vehicle by Cruise sends signals of EMS, MCU, GCU, LDC, BMS, TCU to the HCU logic, and the HCU calculates the each signals by control algorithm. As a result, the final calculated EMS, MG, ISG, LDC, TCU commands transmit the target vehicle. Above step repeat each sample time. This simulation has sampling time 0.005 sec and uses the FTP-72 profile among the driving cycles. Fuel consumption a constant speed cannot accurately represent real driving conditions. Various drive cycles have been developed to simulate real driving conditions. The drive cycles are usually represented by the speed along with the relative driving time. Legislation drive cycles – all mass produced cars are subjected to before being authorized for sale in a particular market. The total mass of emissions produced during a particular drive cycle must be below a set limit decided by the legislating authority. The most common – the cycles used by the US EPA and the European ECE. Light-Duty Vehicles (Chassis Dynamometer), FTP72 – A transient test cycle for cars and light duty trucks on a simulated urban route with frequent stops (Federal Test Procedure)

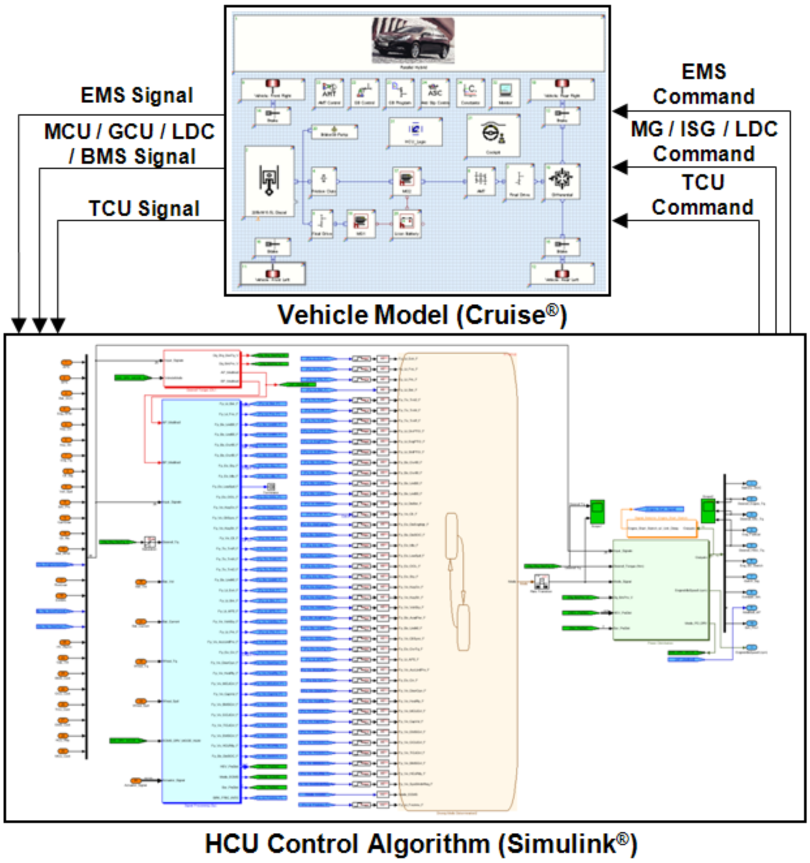


Figure 5: Simulation Environment

5.2. The simulation result for the fuel economy

The result of the driver characteristic variation works the fuel economy. This chapter shows the simulation result for the fuel economy according to changing the driving pattern. Chapter 5.2.1 is the case applied the optimal equivalence factor for each driving tendency. Chapter 5.2.2 is the case applied the non-optimal equivalence factor for each driving tendency. These two cases fuel economy made by table 2, 3.

5.2.1. The simulation for the optimal equivalence factor

The optimal equivalence factors are found for the one case selecting among Case1~6 of the each driving pattern. Case1 (Flexibility), Case6 (Sensitivity), Case4 (Competence) is simulated with it's the optimal equivalence factor and the fuel economy is calculated. In order, the fuel economy is 17.281km/l, 21.901km/l, 21.994km/l.

5.2.2. The simulation for the non-optimal equivalence factor

The equivalence factor sets selected in the chapter 5.2.1 are exchanged each other. The fuel economy appears in table 2. Comparing the table 1 and table 2, the gap of the Max and min fuel economy is the 5.831km/l in the Case1 (Flexibility), is the 0.991km/l in the Case6 (Sensitivity) and is the 1.846km/h in the Case4.

The fuel economy simulation result shows that the optimal fuel economy is made by adapting the optimal equivalence factor according to the driver characteristic.

Table 2: Optimal equivalence factor and fuel economy
*Fuel economy unit : [km/l]

Flexibility	S_{chg}	S_{dis}	Fuel economy
Case1	1.80	2.41	17.281
Sensitivity	S_{chg}	S_{dis}	Fuel economy
Case6	2.49	2.51	21.901
Competence	S_{chg}	S_{dis}	Fuel economy
Case4	1.78	2.87	21.994

Table 3: Non-optimal equivalence factor and fuel economy

Flexibility	S_{chg}	S_{dis}	Fuel economy
Case1	2.49	2.51	11.450
Case1	1.78	2.87	17.261
Sensitivity	S_{chg}	S_{dis}	Fuel economy
Case6	1.80	2.41	20.910
Case6	1.78	2.87	21.226
Competence	S_{chg}	S_{dis}	Fuel economy
Case4	1.80	2.41	20.148
Case4	2.49	2.51	20.190

*Fuel economy unit : [km/l]

6. CONCLUSION

This paper presented the concept of novel adaptive ECMS and defined the driver types in terms of the parameter set of the longitudinal driver model which is developed. The effectiveness of the equivalence factor adaptation according to the driver characteristic variation is verified through the fuel economy simulation. The simulation result shows that each driver using the optimal equivalence factor has better fuel economy than the non-optimal equivalence factor. In conclusion, the ECMS parameter adaption for the driver characteristic variation has validity.

In the future work, an algorithm identifying actual driver in real-time need to be developed, because this study analyzes the relationship between the offline-defined driver type and the equivalence factor as an initial step of novel adaptive ECMS algorithm. Taking a step forward, this study does not consider the road condition identification. Some pattern recognition algorithm for road condition identification can be combined with this study

REFERENCES

James, D.J.G., Boehringer, F., Burnham, K.J., and Copp, D.G., 2004. Adaptive driver model using a neural network. *Proceedings of artif life robotics, 7 4, 170-176.*

Majjad, R., Kiencke, U., and Körner, H., 2005. A-ECMS: An adaptive algorithm for hybrid electric vehicle energy management. *Proceedings of the 44th IEEE conference on decision and control, and the European control conference,* pp. 1816-1823. December 12-15, Seville, Spain.

Cheng, B., Fujioka, T., 1997. Hierarchical driver model. *Proceedings of the ITSC '97 IEEE conference on intelligent transportation system,* pp. 960-965. November 9-12, Boston, MA.

Gu, B., Rizzoni, G., 2006. An adaptive algorithm for hybrid electric vehicle energy management based on driving pattern recognition. *Proceedings of (ASME) international mechanical engineering congress and exposition,* pp. 249-258. November 5-10, Chicago, Illinois, USA.

Majjad, R., Kiencke, U., and Körner, H., 1998. Design of a hybrid driver model. *Proceedings of the 1998 society of automotive engineers international congress and exposition,* pp. 37-42. February 23-26, Detroit, Michigan.

McGordon, A., Poxon, J.E.W., Cheng, C., Jones, R.P., and Jennings, P.A., 2011. Development of a driver model to study the effects of real-world driver behavior on the Fuel. *Mechanical engineers, Part D: Journal of automobile engineering,* 225 11, 1518-1520.

Yang H.-H., Peng, H., 2010. Development of an errorable car-following driver model. *Vehicle sytem dynamics: International journal of vehicle mechanics and mobility,* 48 6, 751-773.

Inata, K., Raksincharoensak, P., and Nagai, M., 2008. Driver behavior modelling based on database of personal mobility driving in urban area. *Proceedings of international conference on control, automation and systems,* pp. 2902-2907. Octobor 14-17, COEX, Seoul, Korea.

Brundell-Freij, K., Ericsson, E., 2005. Influence of street characteristics, driver category and car performance on urban driving patterns. *Transportation research part D: Transport and environment,* 10 3, 213-229.

Sciarretta, A., Back, M., and Guzzella, L., 2004. Optimal control of parallel hybrid electric vehicles. *IEEE transactions on control systems technology,* 12 3, 352-363.

Hofman, T., Steinbuch, M., van Drutenm R.M., and Serrarens, A.F.A., 2008. Rule-based equivalence fuel consumption minimization strategies for hybrid vehicles. *Proceedings of the 17th world congress the international federation of automatic control,* pp. 5652-5657. July 6-11, Seoul, Korea.

Sciarretta, A., Chasse, A., 2011. Supervisory control hybrid powertrains: An experimental benchmark of offline optimization and online energy management. *Control engineering practice,* 19 11, 1253-1265.

Reñski, A., 1998. The driver model and identification of its parameters. *Proceedings of international congress and exposition,* pp. 1-8. February 23-26, Detroit, Michigan.

Macadam, C., 2003. Understanding and modeling the human driver. *Vehicle system dynamics: international journal of vehicle mechanics and mobility,* 40 1-3, 101-134.

TOWARDS A METHODOLOGY FOR MODELLING AND VALIDATION OF AN AGRICULTURAL VEHICLE'S DYNAMICS AND CONTROL

Martin Peter Christiansen(a), **Kim Bjerge**(b), **Gareth Edwards**(c), **Peter Gorm Larsen**(c)

Department of Engineering, Aarhus University
Finlandsgade 22, 8200 Aarhus, Denmark

(a)mapc@djf.au.dk, (b)kbe@iha.dk, (c)Gareth.Edwards@agrsci.dk, (d)pgl@iha.dk

ABSTRACT

A model-oriented approach aimed at cost-effective development of autonomous agricultural vehicles is presented. Here a combination of discrete-event modelling of a digital controller and continuous-time modelling of the vehicle is used for co-simulation. In order to have confidence in the simulation results it is paramount to be able to relate the simulation results to the behaviour of the real system. The cost of physical tests is high and we argue that using such collaborative models is a cost-effective way to experiment with the most significant design parameters influencing the optimal system solution. The suggested methodology is exemplified on a Lego®Mindstorms®NXT micro-tractor. Testing is performed based on measurements from a localisation system and internal sensors on the tractor. Our tests show that we are able to predict the performance with a high accuracy indicating that this is worthwhile for a full-scale model.

Keywords: Auto-steering, Bond graph, Lego Mindstorms NXT, Vienna Development Method

1. INTRODUCTION

Modern agricultural machinery is gradually moving towards a higher degree of autonomous operation (Grisson et al. 2009). Global Navigation Satellite Systems (GNSS) in combination with other sensors are used to estimate the position of the vehicle. Operational tasks like ploughing, spraying and harvesting are commenced by the autonomous vehicle. A pre-planned route for the agricultural vehicle to follow for a specific broad-acre field is supplied in advance. The onboard auto-steering system then aims to adjust the current position so it gets as close as possible to the pre-planned route. The ability to automatically correct the position helps deal with physical conditions, such as the terrain (Fang et al. 2005), which may affect the vehicle's movements in unpredictable ways.

Methods to determine the precision of the vehicle's control equipment have been proposed in a ISO test standard (DS-F/ISO/DIS 12188–2). Testing is performed over a period of more than 24 hours, repeating the testing scenarios multiple times. Full-scale testing of performance and operation is both time-consuming and very costly. Utilising a simulated model of an agricultural machine and auto-steering system, could lower some of these costs. Relevant testing scenarios can be determined based on flaws found through evaluation of the simulations. These scenarios could then be tested to determine if they would produce similar results as in the real physical system.

The aim of this work is to develop collaborative models of agricultural vehicles and their auto-steering systems, combining discrete-event models of control elements with continuous-time modelling of the physical elements and the surrounding environment. The Vienna Development Method (VDM) is utilised for discrete-event modelling of the vehicles control equipment and 20-sim is used as the continuous-time framework for modelling the tractor. In this paper collaborative modelling (co-modelling) is used to model a concrete physical system and its controller.

The aim of this work is to develop collaborative models of agricultural vehicles and their auto-steering systems, combining discrete-event models of control elements with continuous-time modelling of the physical elements and the surrounding environment. The Vienna Development Method (VDM) is utilised for discrete-event modelling of the vehicles control equipment and 20-sim is used as the continuous-time framework for modelling the tractor. In this paper collaborative modelling (co-modelling) is used to model a concrete physical system and its controller.

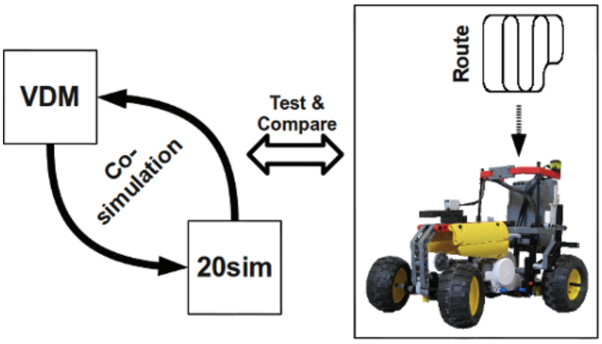


Figure 1: Overview of the micro-tractor and the co-modelling method. 20-sim models the vehicle and VDM the control part

An agricultural tractor system is a complex system to model, simulate and test. Many of the parameters in such a complex system are unknown, making it difficult to verify the complete model based on testing and analysis. To simplify the process, a model and co-simulation based on a Lego®Mindstorms®NXT tractor (micro-tractor) has been developed. The micro-tractor is a representative scaled model of agricultural machinery used to test and demonstrate autonomous operations.

This prototype model of the autonomous vehicle is intended to provide an abstraction with key components in autonomous vehicle steering and explore alternative requirements and design decisions.

The model of the system will provide the functionality to control the motors for the front and back wheels, using the inputs from motor encoders and Inertial Measurement Unit (IMU) in VDM. Subparts of the 20-sim model were modelled separately and combined after verification of each subpart (component). The output from the co-simulation model is the dynamic movements of the vehicle while commencing a pre-planned route.

The article is structured as follows: a short presentation of the underlying method and technologies used for co-modelling and co-simulation is given in section 2. The case study with the scaled-down tractor and a short description of the proposed development process is provided in section 3. The verification and validation of the model in relation to the real physical system is found in section 4. Finally section 5 provides concluding remarks together with directions for future work.

2. TECHNOLOGIES APPLIED

Models of control systems can be complex when they account for many different scenarios. Testing the final model to determine the sources of a specific problem is complicated and time consuming. The work presented in this paper uses a methodology that integrates tests as an essential part of the development process.

The idea is to discover errors and faulty assumptions at an early stage in the development process. It is expected that combined analysis and testing throughout the development will provide a good methodology for developing the multi-domain models.

In the initial part of the development phase a subcomponent of the system is selected. This subcomponent is analysed and an initial sub-model is created. A test scenario will then be created to determine the sub-model's accuracy compared to the actual setup. If flaws are found in the sub-model, extensions and improvements are made until the model represents the actual sub-component.

After the subcomponents have been modelled independently they are put together as a first version of the system intended to be modelled. This model will then be tested using the same process as the sub-models. The process is an iterative incremental development process that improves and extends the model.

2.1. DESTECS and Co-simulation

This paper is based on the DESTECS (”Design Support and Tooling for Embedded Control Software” (see www.destecs.org).) co-simulation technology (Broenink et al. 2010) that supports a model-based approach to the engineering of embedded control systems. Models are built in order to support various forms of analysis including static analysis and simulation — the latter is our focus here.

The technology supports models where the controller and plant or environment is modelled using different specialized environments and tools. In particular, it supports co-simulation by allowing the collaboration of two simulation engines in order to produce a coherent combined simulation of a co-model of a digital controller expressed in a Discrete-Event (DE) formalism and a model of the plant/environment expressed in a Continuous Time (CT) notation.

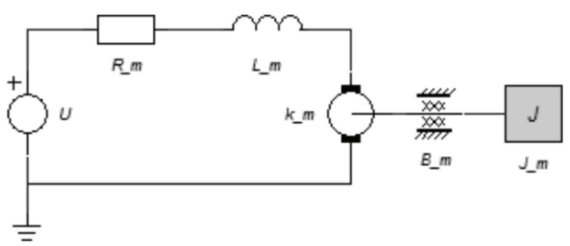
In order to link the DE and CT models together, a contract is established between them. The contract includes information about the shared design parameters as well as monitored and controlled variables exchanged between the two simulators. Once co-models have been constructed, they can be evaluated by co-simulation. Evaluation is done using criteria's chosen by the developer, intended to select the best candidate co-model termed Design Space Exploration (DSE).

VDM is used for modelling DE controllers, and 20-sim as the CT framework for modelling the environment. VDM Real Time (VDM-RT) is the dialect used in DESTECS (Verhoef et al. 2006; Verhoef 2009). Both VDM and 20-sim are well-established formalisms with stable tool support and a record of industry use.

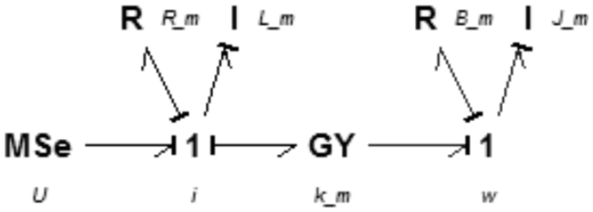
2.2. 20-sim and Bond graph modelling

20-sim is a modelling and simulation tool, developed by Controllab Products in the Netherlands. The tool is able to model complex multi-domain dynamic system, such as combined mechanical, electrical and hydraulic systems. 20-sim models (Kleijn 2006) may use iconic diagrams, Bond graphs and equation models. Iconic diagrams generally contain a sub-model based on equations or Bond graph models. In this context a sub-model means a part on the overall model describing a dynamic system.

Bond graphs are a type of directed graph representing the idealized power flows in a dynamic system (van Amerongen 2010). Every element in a Bond graph is represented by a multiport, describing a subpart (sub-model) of the system. The connections between sub-models, called bonds, represent the exchange of energy. Each port element describes the energy flow, using the product of the variables effort (*e*) and flow (*f*). The meaning of different ports elements changes based on the current system domain. In an electrical domain (*e*) and (*f*) could represent voltage and current and in a mechanical domain torque and angular velocity. This abstraction of ports provides a huge advantage, in terms of reuse and movement between different physical domains.



(a) Iconic diagram representation



(b) Bond graph representation

Figure 2: Motor representations using iconic diagrams.

An iconic diagram representation of a DC-motor can be seen in Figure 2(a). In a Bond-graph terminology voltage source would mean an effort source for the motor system. In Figure 2(b) the bond-graph represents the Direct Current (DC)-motors mechanical and electrical domains. In the electrical domain the I-element represents the inductance L_m and R the resistance R_m connected to the electrical 1-junction. The gyrator GY relates the effort and flow between the electrical and mechanical domain. In the mechanical domain the R-element represents the internal friction B_m and the I-element the moment of inertia J_m. In the bond-graph model *i* represent the current and *w* the rotational speed of the DC-motor.

2.3. The Vienna Development Method

The 20-sim tool environment also provides techniques to allow for mixed modelling and simulation of digital controlled physical systems. Using the equation models different discrete-event scenarios can be simulated and tested. This provides the ability to simulate both continuous and discrete time events of a dynamic system. Most modern digital control systems are complex and hard to model in a single model block. The 20-sim tool provides the ability for external software to connect and communicate with a specific model. An external environment could therefore be used to model the discrete time event parts of a dynamic model (Fitzgerald et al. 2011).

VDM is a formal method for specification, analysis, modelling and identification of significant features in a computer system. VDM originate from work done at IBM's Vienna Laboratory in the 1970's on semantics of programming languages (Bjørner and Jones 1978). VDM provides the ability to model at a level higher of abstraction, than is realizable in a normal programming language. Validated models can then be turned into a concrete implementation in a programming language. The current tools focus are to provide modelling and analysis techniques used for simulation rather than proof checking. VDM tool support is provided by the open-source Overture tool (Larsen et al. 2010).

The demands and assumptions about the system intended for modelling is a significant part, when describing the functionality of the system. Functionality is performed on different data types, ranging from basic type like Booleans, tokens, integers and real numbers and collections such as sets, mappings and sequences. Functions can be either implicit or explicit specified in VDM, for a modelled system in terms of describing the relations. The VDM functionality has been extended to include the Object-oriented structuring (Fitzgerald et al. 2005) using the VDM++ extension. For VDM to be used in a real-time embedded system context, it requires explicit modelling of computation time. The capabilities to describe real-time, asynchronous, object-oriented features are provided in the VDM-RT extension. Using the VDM formalism for both control and modelling of the environment is not an ideal solution, since the CT environment would be expressed in DE formalism based on simplifying assumptions. A co-simulation with a CT event tool would be a significant improvement, in terms of simulating the VDM controller.

3. THE MICRO-TRACTOR CASE STUDY

The micro-tractor was developed to represent an average 150 bhp tractor. A scale ratio of 1:14 was used. The micro-tractor is described in (Edwards et al. 2012).

The micro-tractor's steering range was between +/- 30 (degrees) and controlled by an NXT servo motor and gearbox.

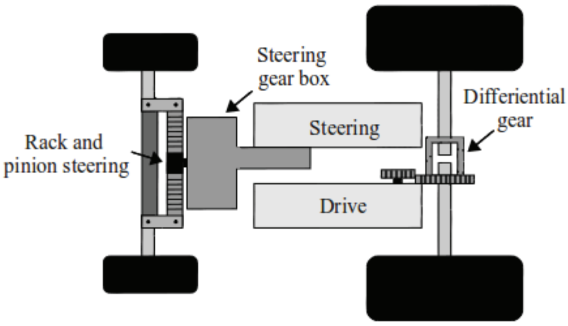


Figure 3: Sketch of the micro-tractors steering and drive components.

The rear wheels were powered by another NXT servo motor. A differential gear was used on the drive axle to allow the wheels to turn at different speeds and reduce slip (see Figure 3).

The navigation sensor is the CruizCore R XG1300L IMU. The IMU measures heading of the micro-tractor based on relative initial heading. The IMU contains a single axis MEMS gyroscope and a three axis accelerometer. The signals from these sensors are processed onboard the IMU.

3.1. Co-simulation

A co-simulation engine is responsible for exchange of shared parameters and variables between the CT and DE models. The co-simulation engine coordinates the 20-sim and VDM simulation by implementing a protocol for time-step synchronisation between the two simulation tools. A contract defines the parameters and variables to be exchanged during simulation as illustrated in Figure 4. Here the start and stop times of the co-simulation are shared, to ensure common reference. The micro-tractor (CT Model) updates the shared angle parameters of the motor encoders and IMU and the controller (DE model) drives the shared input parameters the motors.

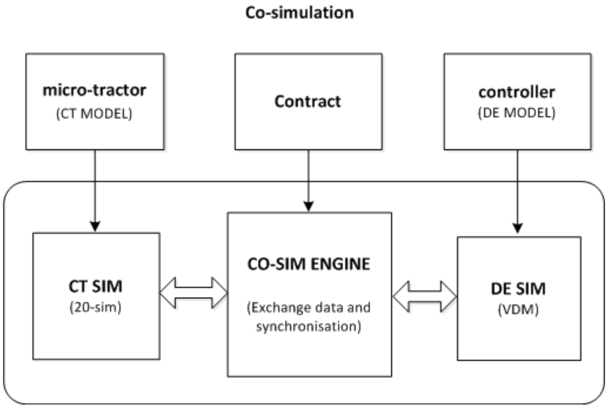


Figure 4: Co-simulation engine and synchronisation of the CT and DE simulation.

3.2. Bond graph model

The motor is modelled as the first component of the vehicle model, since it is used for control of both drive and steering.

To exemplify the development method, the motor subpart will is described in detail. The NXT controls the average voltage output to the motors using Pulse Width Modulation (PWM). This makes it possibly to compare input/output between model and actual Lego DC-motor: A representation of a bond graph DC-motor model can be seen in Figure 2(b).

To test if the model works as intended, the bond graph is supplied with motor parameters for the Lego DC-motor. An impulse-function (0V-7.4V-0V) is used to apply a voltage to both Bond graph and real DC-motor.

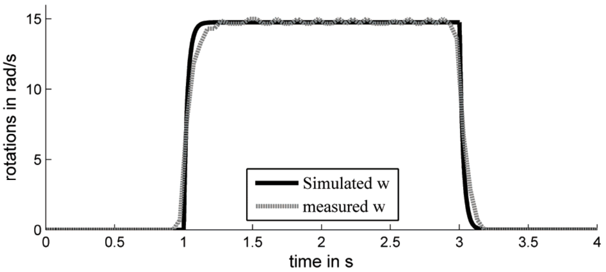


Figure 5: Impulse test of DC-motor model and real system.

From the plot in Figure 5 it can be seen, the plots correlate to a high degree. Based on these findings, the DC-motor bond-graph is accepted and development on other subparts is initiated.

A similar methodology is applied when modelling the remaining components in 20-sim.

The controller is connected to the DC-motor implementation, providing the interface for communication with VDM. Gearing for steering and drive components are modelled using Transformer TF elements. Each TF element corresponds to a gearing ratio effort with effort out causality. A Bond-graph equivalent of a spring damper (C_g , R_g) system is used to model the effects of rotating gears using a 0-Junction.

Change in angle of the front wheels are represented using a friction and moment of inertia (B_{fwheel}, J_{f_wheel}). Interactions between wheels and ground plane are only considered for a smooth surface to keep the complexity of the model down.

Only the longitudinal effects on the wheel are considered, since tire-road normal effort (Merzouki et al. 2007) is expected to be minimal. Effects of the wheels moment of inertia is represented with a 1-junction and an I-element J_{b_wheel}. A TF-element converts between rotational and linear speed. A spring-damper system (C_{b_wheel} , R_{b_wheel}) is used to represent the longitudinal surface interaction with the wheel contact-point.

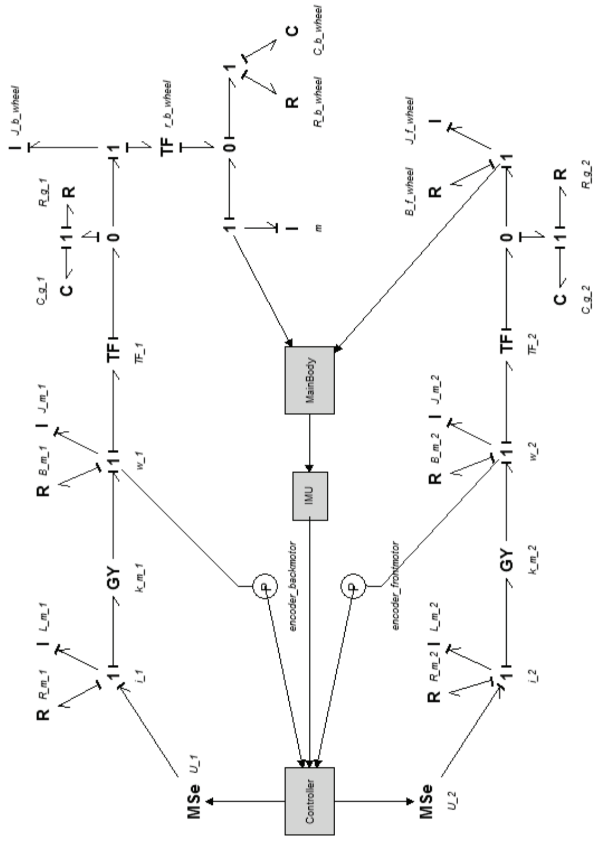


Figure 6: The micro-tractor model in 20-sim.
A combination of bond graphs and iconic diagram blocks are used for modelling the micro-tractor.

To combine the dynamic effects of the back and front wheels a first order bicycle model (Figure 7) has been chosen. The first order bicycle model is a pure kinematic model (Rovira M Rovira Más et al. 2011) of the chassis movements, without regards for the forces acting on the body.

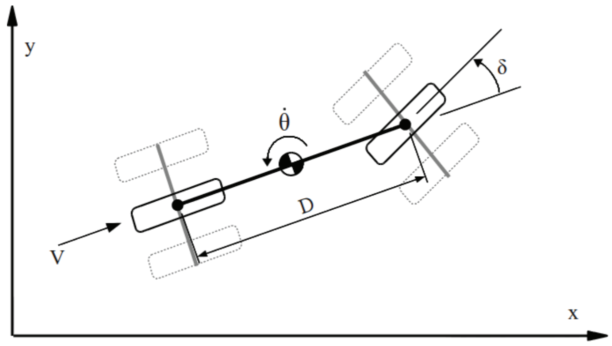


Figure 7: Bicycle vehicle model used to model the body of the micro-tractor.

Translational speed (V) from the back wheel in combination with the rotational speed (δ) of the front steering system is used as input for the model.

Equation (1), (2) and (3) are used to calculate the vehicle rotation and speed in x,y direction in a global reference frame.

$$\dot{x} = cos(\theta) \cdot V \tag{1}$$

$$\dot{y} = sin(\theta) \cdot V \tag{2}$$

$$\dot{\theta} = (tan(\delta)/D) \cdot V \tag{3}$$

D represents the distance between front and back wheels, δ the orientation of the front wheels. To represent backlash in the steering system Gaussian noise is added to the rotational change.

A Bond graph model of backlash in the front wheel orientation is not incorporated, since modelling of the external forces on both front wheels would be needed. Position and orientation of the vehicle over time is calculated using numerical integration of equation (1), (2) and (3). The positioning and orientation of the vehicle is the intended output from the model.

3.3. VDM model

The discrete event control system modelling the NXT's steering of the micro-tractor is modelled in VDM-RT. A pre-planned route is given to the autonomous system. This route is used when the micro-tractor commences its task in an area.

The route is based on a collection of continuous curve elements. Each continuous path element is either a line segment or circular arc with constant radius, containing a start and stop waypoint (Bevly 2009).

The micro-tractor is aware of its current position and is able to use this information when following the route. A route-manager ensures that each route segment is performed in the order described in the route. The VDM model uses invariants and pre and post-conditions to ensure only a viable route and route segments are commenced. The description is given to provide the reader with a perspective of what the VDM capabilities could be used for. Details of the route-manager will not be given in this paper but similar systems can be found in (Fitzgerald et al. 2005).

When a route element is commenced a control loop is needed to keep the micro-tractor on track. In this model the inputs from the back-motor encoder and the IMU is used to determine the current position and orientation. The model for executing a line function segment can be seen in Listing 1.

```
class controlStraight

instance variables
public rotations: real; -- 20-sim variable
public ImuOrient: real; -- 20-sim variable
P: real; --proportional control factor
MotorOutput: real; -- output value in %
distance: real; -- distance to travel
wOrient: real; -- Wanted orientation

operations
public ControlStep: () ==> ()
ControlStep() ==
( if abs rotations >=
distance/(2*MATH‘pi*R_BACK_WHEEL)
then drivingMotor := 0.0;
else driveMotor := MAX_OUTPUT*MotorOutput;
steerMotor := MAX_OUTPUT*P*(ImuOrient-
wOrient)
);

thread
periodic(10E5,0,0,0)(controlStraight); -- 100Hz
end controlStraight
```

Listing 1: VDM++ model of control loop for driving a straight.path.

A line segment is followed until the distance between start and stop point is reached. The control loop allows the micro-tractor to steer off track, since any positioning error is accumulated. A more advanced control system could compensate for this and is intended for the future.

4. MODEL VERIFICATION AND VALIDATION

Measurement data from the testing of the micro-tractor is compared against the vehicle co-model to determine the accuracy of the co-model. To accomplish this task measurement data from an external source is compared against data from the co-simulation. The testing, measurement method and results thereof are presented in this section.

4.1. Test scenario

The testing will determine the difference between the actual system and the co-simulation, in terms of position and orientation. Running the same route-scenario in will make them comparable and provide a means of comparing different parts of the route. Testing is performed on a route with 3 straight segments of 2-3 meters and 2 circle arcs in opposite direction. A more complex route with more route-segments could introduce larger accumulative errors in terms of position.

Since this is not taken into account in the current co-model, comparison would clearly fail. The selected route ensures the testing is done for movements with different rotational speeds θ of the micro-tractor body, which is a major part of the micro-tractor dynamics. The model parameters used in the co-simulation of the bicycle vehicle model are given in Table 1.

Table 1: Testing and co-simulation parameters

Sub-system	**Parameter Values**	
Motor:	$R_m = 5.2637(\Omega)$ $k_m = 0.4952$ $J_m = 0.0013(kgm^2)$	$L_m = 0.0047(H)$ B_m=6^{-4} Nm/(rad/s)
Gearing:	$TF_{g1} = 20/28$ $Cg = 10^{-5}$	$TF_{g2} = 3/70$ $Rg = 10^{-5}$
Back-Wheels :	$J_{wheel} = 3.67^{-6}$ $C_{b\_wheel} = 1.1\mu$	$r_{b\_wheel} = 0.0408(m)$ $R_{b\_wheel} = 0.3$
Body:	$m = 2.2374(kg)$	$D = 0.175(m)$

Parameters with the equal value in Figure 8 like k_m_1 / k_m_2 is represented with same symbol (ex k_m).

4.2. Testing equipment

To determine the position of the micro-tractor over time the iGPS system from Nikon is used. iGPS measurement technology is a laser-based indoor system with optical sensors and transmitters to determine the 3D position of static or moving objects. The iGPS technology is based on internal time measurements related to spatial rays that intersect at sensor positions in the measuring area. The iGPS measurement system has been evaluated in experimental studies (Depenthal, 2010) of the capabilities for tracking applications.

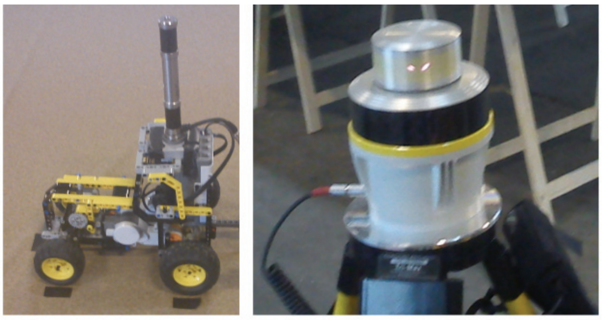

(a) (b)

Figure 8: Nikon iGPS receiver mounted at the micro-tractors CoR (a) and transmitter used in the testing (b).

The intention is to use the iGPS technology to evaluate full-scale autonomous agricultural vehicles based on the ISO. The system is able to measure the micro-tractor position over time to provide capabilities for direct analysis of auto-steering system.

The iGPS sensor is mounted on top of the micro-tractor close to the Centre of Rotation (CoR) (see in Figure 8(a)). Using the CoR as measurement point ensures measurement data is comparable directly with co-simulations. The vehicle was driven at 20% of full motor power output when running the pre-planned route. The low motor-output was chosen to ensure safe driving, when using the iGPS sensor system.

The testing was repeated 10 times to account for any variation in performance.

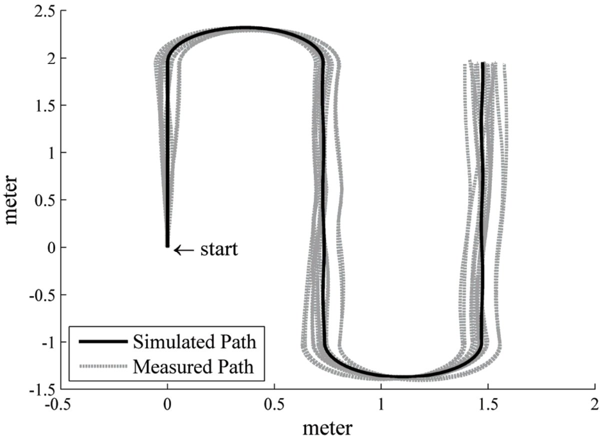


Figure 9: Measured and simulated path of the micro-tractor. Measured path shows the variation in 10 runs.

Figure shows the micro-tractor drifts in position but keep it's heading throughout the path. The small variations in angle can mostly be described to variation in initial placement, since the IMU is a relative sensor any misalignment is kept throughout the path. These initial misalignments are not part of the simulation, resulting in a path moving in the middle of the actual measure paths.

Position measurements are determined to have a precision of 0.4-0.5 mm, based on estimates provided by the iGPS system.

5. CONCLUDING REMARKS AND FUTURE WORK

Based on testing of the current co-model it can be concluded, that the model can emulate the basic performance intended for the micro-tractor. Visual presentation of results and comparisons show a high consistency between actual system and co-model. The current model can be seen as a first step towards a full scale co-model of an autonomous agricultural vehicle. Different routes can be tested in the co-model, to determine their efficiency on the actual system. The co-model will allow for detailed analysis, without the need to start the process of the testing each time.

The need for testing can be diminished significantly and thereby saving development costs. The development process has shown exemplary initial results, to produce dynamic models. Splitting design of the model into smaller steps helps ensures sub-part errors could be determined easily throughout testing. Experience from this project has indicated an iterative modelling method with testing to be a beneficial development and modelling approach.

Continuing to use the current approach to model development is therefore seen as a promising way of continuing the development.

The current version of the co-model is used to test and develop more advanced control algorithms for the micro-tractor driving and steering system. Only selected versions of the control system need to be tested on the actual system, to confirm any improvements. Obtained measurements could be used to improve the model, should a different result be produced in the testing from simulation using the co-model. Many errors and shortcomings in the control loop can also be tested using the co-model to determine their source and provide the mean to test solutions to the problems.

The co-model currently has a number of shortcomings in terms of describing the dynamics between the front and back wheels. At the current state, a kinematic description is used to describe the overall changes to the vehicles placement and orientation. Forces acting on front and back wheels need to be described in more detail in the co-model to account for their interaction. The forces introduced by the back wheels when rotating and thereby driving the vehicle forward influences the front wheels and the backlash introduced in the orientation. Occurrences of backlash are seen when the micro-tractor is moving in a straight line in figure 9 as small changes in orientation over time. Body rotation of the vehicle introduced by the front wheels will introduce forces on the back wheel and the differential gearing drive. Accounting for these factors is expected to provide a model able to run more complex route scenarios and provide a reliable estimate of the real system.

These improvements to the model are planned to be part of the next stage of the co-model development. Later versions should also account for the external factors like uneven terrain and 3-dimensional movement.

ACKNOWLEDGMENTS

The authors want to thank Ole Green for supervision and support of the work. We would like also to acknowledge Joey W. Coleman and Sune Wolff for provided invaluable input regarding the content and structure of this work. Thanks are due to Jens Kristian Kristensen at Research Centre Foulum for providing assistance in connection with the experiments. We acknowledge partial support from the EU FP7 DESTECS project on co-simulation. Financial support given by the Danish Ministry of Food, Agriculture and Fisheries is gratefully acknowledged.

REFERENCES

Bevly, D.M., Cobb, S., 2009. *GNSS for Vehicle Control.* 1st ed. Norwoord: Artech House.

Bjørner, D., Jones, C.B., 1978, The Vienna Development Method: The Meta-Language, *Lecture Notes in Computer Science*, 61, 24-217.

Broenink, J. F., Larsen, P.G., Verhoef, M., Kleijn, C., Jovanovic, D., Pierce, K., Wouters, F., 2010, Design Support and Tooling for Dependable Embedded Control Software, *Proceedings of Serene 2010 International Workshop on Software Engineering for Resilient Systems*, 77-82, April, New York (USA).

Claudia, D., 2010, Path tracking with iGPS, *Proceedings of International Conference on Indoor Positioning and Indoor Navigation*, 1-6, September. Zurich (Switzerland).

DSF/ISO/DIS 12188–2, 2011, Tractors and machinery for agriculture and forestry - test procedures for positioning and guidance systems in agriculture - part 2: Testing of satellite-based auto-guidance systems during straight and level travel, *International Organization for Standardization*, Denmark

Edwards, G., Bochtis, D., Sørensen, C.G., 2012, Developing a field robotic test platform using Lego®Mindstorm®NXT. *Proceedings of CIGR International Conference on Agricultural Engineering*, July 8-12, Valencia (Spain).

Fang, H., Lenain, R., Thuilot, B., Martinet P., 2005, Robust Adaptive Control of Automatic Guidance of Farm Vehicles in the Presence of Sliding, *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 3102–3107, April 18-22, Barcelona (Spain).

Fitzgerald, J., Larsen, P.G., Mukherjee, P., Plat, N., Verhoef, M., 2005, *Validated Designs for Object–oriented Systems*, 2nd ed. New York: Springer.

Fitzgerald, J., Larsen, P.G., Pierce, K., Verhoef, M., 2011, *A Formal Approach to Collaborative Modelling and Co-simulation for Embedded Systems*, Technical Report, Newcastle University.

Grisson, R., Alley, M., HeatWole, C., 2009, Precision Farming Tools: Global Positioning System (GPS). *Virginia Cooperative Extension*, 442-503.

Kleijn, C., 2006, Modelling and Simulation of Fluid Power Systems with 20-sim. *Intl. Journal of Fluid Power*, 7(3), November, West Lafayette (USA).

Larsen, P.G., Battle, N., Ferreira, M., Fitzgerald, J., Lausdahl, K., Verhoef, M., 2010, The Overture Initiative – Integrating Tools for VDM. *ACM Software Engineering Notes*, 35(1), January.

Merzouki, R., Ould-Bouamama, B., Djeziri, M.A., Bouteldja M., 2007, Modelling and estimation of tire-road longitudinal impact efforts using bond graph approach, *Mechatronics*, 17(2-3), 93–108, March-April.

Rovira Más, F., Zhang, Q., Hansen, A.C., 2011, *Mechatronics and Intelligent Systems for Off-road Vehicles*, 1st ed. London: Springer-Verlag

van Amerongen, J., 2010, *Dynamical Systems for Creative Technology*, 1st ed. Ensched (Netherlands): Controllab Products.

Verhoef, M., 2009, *Modeling and Validating Distributed Embedded Real-Time Control Systems*, PhD thesis, Radboud University Nijmegen.

Verhoef, M., Larsen, P.G., Hooman, J., 2006, Modeling and Validating Distributed Embedded Real-Time Systems with VDM++, *FM 2006: Formal Methods, Lecture Notes in Computer Science 4085*, 147–162, Springer-Verlag.

MATHEMATICAL MODELING OF A 500 W_e POWER MODULE COMPOSED BY A PEM FUEL CELL COMBINED WITH A DC-DC ENHANCED POTENTIAL OUTPUT CONVERTER

Roque Machado de Senna (a), Marcelo Linardi (a), Douglas Alves Cassiano(b),
Ivan Santos (a), Edgar Ferrari da Cunha (a), Henrique de Senna Mota (a), Rosimeire Aparecida Jerônimo(c).

(a) Instituto de Pesquisas Energéticas e Nucleares - Universidade de São Paulo - IPEN/USP. Av. Lineu Prestes, 2242, Cidade Universitária, São Paulo, SP, Brasil, CEP05508-000
(b) Universidade Federal do ABC. Rua Santa Adélia, 166, Bangu, Santo André, SP, Brasil, CEP09210-170.
(c) Universidade Federal de São Paulo, Rua Prof. Artur Riedel, 275, Eldorado, Diadema, SP, Brasil, CEP09972-270.

(a)rmdesenna@usp.br, (b)douglas.cassiano@ufabc.edu.br, (c)rosijeronimo@yahoo.com.br

ABSTRACT

This work presents the development of a mathematical modeling of a 500 W_e PEMFC fuel cell stack (MCC500) system combined with a dc-dc enhanced potential output converter. The MCC500 was developed at IPEN (Nuclear and Energy Research Institute) and the company Electrocell, using Brazilian technology exclusively. Mathematical developments and modeling have been performed, relying on experimental data collected at IPEN laboratory. The first step was to prepare an electrical system (pre-design) for the proposed model, which included the MCC500 parameters, like: membrane ohmic resistance, activation resistance, electric double layer capacitance, open circuit potential, as well as DC-DC converter parameters, like the inductor and the transistor switching frequencies. Using the obtained parameters and a linear differential equation system with some mathematical manipulations, an electrical system model was determined. Simulations experiments demonstrated that the system was very stable. This contribution showed to be very important tool to generate useful potential for practical purposes, increasing the overall system electrical efficiency.

Keywords: Fuel Cell and Hydrogen. Fuel Cells System Model. Energy Efficiency. Mathematical Modeling.

1. INTRODUCTION

Electricity, as well as all other energy forms, has played a strategic role at the population life quality indicators, increasing every day in importance (Adam 1991).

Despite its huge social and economic relevance all activities related to energy exploration, production, distribution and use cause environmental impacts, concerning the substances that can release into the atmosphere, the water sources and the soil, endangering the health and survival any terrestrial ecosystem. The various energy systems steps are closely related to each other and to all mankind development.

There is a growing need to find solutions to the related negative effects of the energy production processes chain, with goals to minimize the social, environmental and economic injuries (Szwarc 2007).

The desirable characteristics for power sources are: clean, renewable, low noise emission, low operating costs, reliability, among the others desirable characteristics. Among the options for power sources from renewable sources and environmentally sustainable is the hydrogen gas, which has been used as an energy vector therefore, it is electricity energy storage for many energy sources types, it can be obtained from several energy processes (Linardi, 2010). Three main options may be addressed:

1. The first is the hydrogen production by water electrolysis, from use the electricity production excess from photovoltaic panels, wind generators, hydro, geothermal and nuclear power plants.

2. The second, through the energy use from biomass gasification and further purification, such as garbage, sewage, as well as the forest production remains, agribusiness and agriculture.

3. The third energy process, which has found a higher economic viability in the present day, is supported on catalytic reforming and gasification processes that lead to improved efficacy and reduce the fuel pollution potential already in use, such as ethanol, biodiesel, coal, oil and natural gas. The Brazil Hydrogen production tends to grow building on bioethanol (Brown, 2001), (Liguras et al., 2004; Vaidya and Rodrigues, 2006; Linardi, 2010).

An option for power generation, supported on hydrogen is the power module of PMFC Fuel Cells technology-based. According to the DOE's Fuel Cell Technologies Market Report (DOE, 2011), PEM fuel cells ranging from 0.5 kW to 2 kW were suitable and indicated for residential applications, such as uninterruptible power supply and for combined heat and power supplying. Since the 2010 decade, 0.5 kW to 2 kW PEM fuel cell modules were available in the market, being commercialized for these stationary applications by Horizon and ClearEdge Power, among other companies. The choice of Mathematical Modeling developed for the 500 W PEMFC module in this present work was determined by these facts, although there were also recent mobile applications for this fuel cells class reported on the literature (Shang and Pollet, 2010).

These facts attest that it is extremely necessary to detail further systems studies, like developments based on fuel cells power modules and other renewable and environmentally sustainable technologies (Schoots *et. al.* 2010).

2. OBJECTIVES

The two main objectives of this work are:

1. Development of a mathematical modeling for a system including a PEM fuel cell stack (MCC500) combined to a dc-dc enhanced potential output.
2. Perform simulations using the mathematical model in order to enable the dc-dc converter to raise the bus potential and keep it constant, through the computer simulation on Matlab7® (Mathworks 2007).

3. STATE OF THE ART

3.1 The PEM Fuel Cell Stack - MCC500

The fuel cell is a direct electrochemical power converter (Wendt *et. al.* 2002). At a PEM fuel cell type, two half-cell reactions occur simultaneously, with an oxidation reaction (losing electrons) at the anode and a reduction reaction (gaining electrons) at the cathode.

These two reactions account for the oxidation-reduction reactions at the device, resulting in water formation, due to external fed gaseous hydrogen and oxygen combination, releasing thermal and electrical energy, this one flowing to an external circuit (Kinoshita 2001).

At the anode, the hydrogen molecules come into contact to the platinum catalyst sites (adsorption) on the gas diffusion electrode surface. The hydrogen molecules break their bonds at the platinum surface to form weak links H-Pt. Each hydrogen atom loses its electron to an external circuit, connected to a load, to meet the oxygen ions at the cathode. In turn, the hydrogen ions associated with water (H_3O^+) passes through the proton exchange membrane, reaching the cathode (Linardi 2010).

At the cathode, the oxygen molecules come into contact with the platinum catalyst on the gas diffusion electrode surface. The Oxygen molecules adsorb at the platinum electrode surface, where the oxygen-oxygen bond (O-O) is weakened, and so conditions for the reduction reaction are created. Each oxygen atom then combines with two electrons and two hydrogen ions to form a water molecule. The platinum catalyst at the cathode gas diffusion electrode is now free to weaken new oxygen molecules bonds (Spinacé 2003).

Practical system design requires higher power output than available in a single cell, accordingly, several cells in series association is necessary, named stack, as shown on figure 1, to reach this requirement.

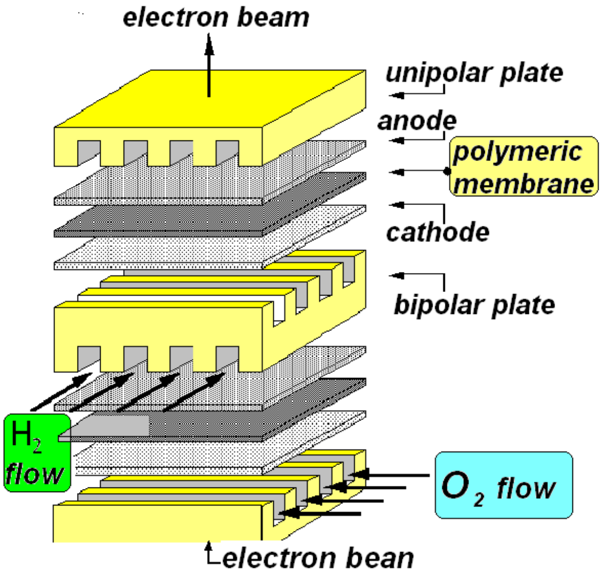


Figure 1: Schematic PEM Fuel Cell Stack (Adapted From Kinoshita 2001).

3.2 The MCC500 Developed at IPEN

The called MCC500 system, as shown at figure 2, is a stack composed by 10 single fuel cells in a series association. The produced electrodes could have, each, until 144 cm^2 of geometrical area, depending on the employed technology. In this work 10 electrodes of 144 cm^2 area were fabricated by the IPEN-Screen Printing Method (Boniface 2011), to be applied to the MCC500 system.

The IPEN-Alcohol Reducing Process was used to produce the Pt/C nano structured electrocatalysts in the gas diffusion electrodes (anode and cathode) (Spinacé 2003).

For the bipolar plates design, a computational simulation of fluid dynamics on gas flow channels was used (Cunha 2009).

In this way, a pre-commercial 500 W_e PEMFC power module was developed, as support for the distributed electricity generation industry, using only Brazilian technology. The Brazilian company ELECTROCELL also contributed to the stack design;

bipolar plate production; sealing, cooling systems and cells assembly.

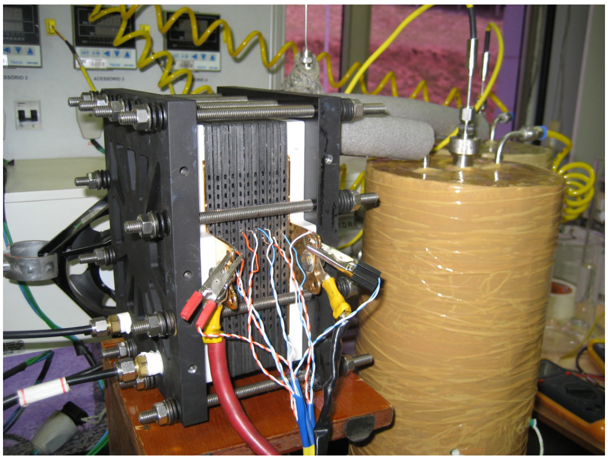


Figure 2: MCC500 Photography Developed at the IPEN and Electrocell® Laboratories.

Previous MCC500 tests showed stable operation, achieving the output power of 500 W_e (77.7 A, current at 6.43 V). The device also showed to be able to produce a maximum power of 574 W_e. For heat recovery studies, it was estimated that the thermal power output developed by MCC500 was 652 W_t, at 500 W_e nominal power. The total stack materials cost was estimated to be around US$ 2,531.70 (Cunha 2009).

3.3 DC-DC Step-up Converter with Enhanced Potential Output Converter

The DC-DC converter is able to receive the electrical potential produced (generated) at MCC500 and make it available on relatively stable potential for the load use, and thereby improve the conversion efficiency, especially when demand is high and the potential generated are small, as shown on Figure 3.

4 METHODOLOGY

4.1 Stability Analysis

In the development of stability analysis, two forms of Nyquist Methods were used. The simulation using Matlab7® computational program proved to be very effective in determining system behavior and in allowing the evaluation of either the output potential elevation or the computational disturbance stability. The system showed to be very stable both by the characteristic equation Root Locus Analysis and by the Nyquist Mapping Theorem (Jonckheere *et. al.* 2002).

4.2 The Mathematical Model Development of the DC-DC Enhanced Potential Output Converter (DC- DC Step-up Converter)

The dc-dc converter mathematical modelling began with a preliminary design of an electrical system, requiring additional steps, as follows:

1. Elaboration of the required polarization curve from current and potential measurements on the MCC500 stack, as shown by Figure 3.
2. Determination of the stood up parameters: inductance of the DC-DC converter inductor (L); the MCC500 electric double layer capacitance (C); the MCC500 membrane ohmic resistance (R_1); the MCC500 resistance activation (R_2), open circuit potential (E), as shown by Figure 4.
3. Combination of the parameters using some mathematical manipulations and the linear differential equations system, combined with mathematical calculations and module parameters (SR-12, Avista Labs). These data permit the electrical system lifting and support the parameterized circuit diagram form for the model presented, as shown by Figure 4 and table 1.

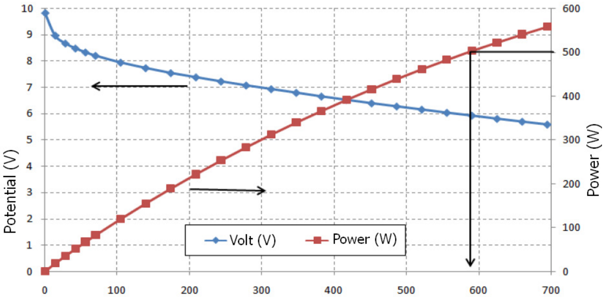


Figure 3: The PEM Fuel Cell Stack - MCC500 Polarization and Power Curves, (Adapted from Cunha 2009).

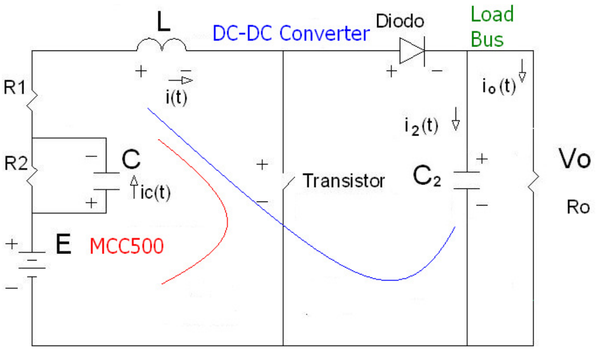


Figure 4: Schema of the Electrical Model for the MCC500 with Dc-Dc Converter, Parameterized to Control the Load Bus Potential, Author.

4.3 Simulation of the Converter Potential to Increase Capacity and Keep it Constant

This step was started by defining to apply the steps and pulses number. Then the program routine was developed and the assisted simulations on Matlab7® were carried out, culminating with the graphic form and table to results presentation.

5 RESULTS AND DISCUSSION

5.1 Determining the inductor to the system

The inductor (L) was obtained based on the equations (1) to (6) as described by Ahmed (2000) and Linardi (2010), and the data in the table (1) and (2), where (T_{on}) corresponds to half the period of the transistor switching frequency (fsh), (T) was period of the transistor switching frequency, (E_{op}) was the value of the standard potential operation, (I_{pp}) was the current ripple that was allowed the peak to peak for the inductor, (g_e) was the static gain, (D) was the commutation duty cycle, (Vi) was the linear initial potential, (L_{min}) was the minimum inductor value for warranting continuous current conduction into the diode, (R_o) was the load nominal resistance, and (V_o).was the stable potential in load bus.

$$g_e = \frac{V_o}{V_i} \tag{1}$$

$$D = \frac{g_e - 1}{g_e} \tag{2}$$

$$T_{on}(s) = D.T \tag{3}$$

$$L_{min}(H) = \frac{R_o . T_{on} . (1-D)^2}{2} \tag{4}$$

$$L(H) = \frac{E_o . T_{on}}{I_{pp}} \tag{5}$$

Table 1: Date of the SR12, Avista Labs

Technology	PEMFC
MEAs	48
Open Circuit Potential (Module)	48 V
Electric Double Layer Capacitance	0.015F
Electric Double Layer Mean Charge	0.72 C
Nominal Power	500 We

Table 2: Date of the MCC500 and Electrical Model

Technology	PEMFC
MEAs	10
Area	144 cm^2
Nominal Power	500 We
Open Circuit Potential (Module) E	9.8 V
Nominal Potential (Module)	6.43 V
Potential (Regulated) in load bus V_o	14.4 V
Linear Low Potential V_L	5.3 V
Nominal Current (Module)	77.7 A
Load Mean Current I_o	34.6 A
Potential static gain g_e	2.7
Efficiency	43.4%
Operating Temperature	65 ° C
Electrolyte Resistance R_1	29.1 mΩ
Activation Resistance R_2	14.9 mΩ
Load Resistance R_o	417.7 mΩ
Load Capacitor C_2	0.005F
Electric Double Layer Capacitance C	0.0367F
Inductor L_{min}	360 pH
Inductor L	228 μH
Transistor Switching Frequency Period T	20 μs
Transistor Switching Frequency Period T_{on}	12.6 μs
Duty Cycle D	0.63

5.2 The Electrical Model

The electrical model is based on the inductor current coming from the MCC500 module, as shown by figure 4.

The six system equations on the time domain (t) and one equation on the complex domain frequency (s) were provided. Results were then obtained as described on the following steps:

Equation 7 inserted the increase due the modification in duty cycle (complementary, D' and d'(t)) on the transistor step or impulse by the controller;

$$D' = \frac{T_{off}}{T} \tag{6}$$

$$d'(t) = D' - \Delta d(t) \tag{7}$$

Equations 8 to 10 showed algebraically the change that occurred in the inductor current i (t) and in the regulated potential of load bus v_o (t) and in the electric double layer potential v_c (t) by the controller step or by impulse;

$$i(t) = I + \Delta i(t) \tag{8}$$

$$v_o(t) = V_o - \Delta v_o(t) \tag{9}$$

$$v_c(t) = V_c - \Delta v_c(t) \tag{10}$$

Equation 11 showed the MCC500 current in the electric double layer, due the potential variation by the variation during the time.

$$\frac{C\,.\,dv_c(t)}{dt} = i(t) - \frac{v_c(t)}{R_2} \tag{11}$$

Equation 12 showed the capacitor (C_2) current, due to the potential variation by to the variation during the time.

$$i_2(t) = \frac{C_2\,.\,dv_o(t)}{dt} = d'(t)\,.\,i(t) - \frac{v_o(t)}{R_o} \tag{12}$$

Equation 13 showed the potential differences algebraic sum along the closed path, including C;

$$E - d'(t)\,.\,v_o(t) = \frac{Ldi(t)}{dt} + R_1\,i(t) + v_c(t) \tag{13}$$

Equation 14 showed the MCC500 current in the electric double layer i_c (t) current, due to the potential variation by to the variation during the time.

$$i_c(t) = \frac{C\,.\,dv_c(t)}{dt} = i(t) - i_{R2}(t) \tag{14}$$

Equation 15 showed the MCC500 current in the electric double layer resistor (R_2), due the potential variation by the variation during the time.

$$i_{R2}(t) = \frac{v_c(t)}{R_2} \tag{15}$$

Equation 16 showed the switching frequency period (T), during the static duty cyrcle.

$$T\ (s) = T_{on} + T_{off} \tag{16}$$

The system composed by equations 1 to 16 is designed to support the Kirchhoff potentials law (the potential on closed path sum is always equal to zero), as described as by Nahvi and Edminister (2011). Equations 6 to 10 were obtained from Granville (1998). Equation 17 originated by solving the equations 7 to 16, wherein the DC terms, in the time (t) were excluded. Immediately after this operation was used Laplace transformation, and the result was showed in the frequency domain complex (s). Then, the terms have been grouped together as a function of (s), as showed in Literal Equations 17, and 17a to 17g, Close (1980) and Nahvi and Edminister (2011).

$$G(s) = \frac{\Delta v_o(s)}{\Delta d(s)} = \frac{V_0.(n_1+n_2+n_3)}{D'.(d_1+d_2+d_3+d_4)} \tag{17}$$

$$n_1 = s^2.L.C.R_2 \tag{17a}$$

$$n_2 = s.\{(D'^2.R_o.C.R_2 - L - C\,.\,R_1\,.\,R_2)\} \tag{17b}$$

$$n_3 = D'^2.R_o - R_1 - R_2 \tag{17c}$$

$$d_1 = s^3 . L . C . R_o . R_2 . C_2 \tag{17d}$$

$$d_2 = s^2 . (L . R_o . C_2 + L.C.R_2 + R_o . R_1 . R_2 . C . C_2) \tag{17e}$$

$$d_3 = s . (D'^2 . R_o . C . R_2 + L + R_1 . R_2 . C + R_o . R_1 . C_2 + R_o . R_2 . C_2) \tag{17f}$$

$$d_4 = D'^2 . R_o + R_1 + R_2 \tag{17g}$$

Using this methodology, a transfer function was obtained coupling experimental data obtained from the power module and some mathematical transformation as shown by Figure 4 and tables 1and 2.

The final result is shown in equation 8.

$$G(s)_p = \frac{\Delta v_O(s)}{\Delta d(s)} = \frac{V_O . (n_1 . n_2)}{D' . (d_1 + d_2)} \tag{18}$$

$$n_1 = 25.3815.10^3 \tag{18a}$$

$$n_2 = s^2 + s.1.8225.10^3 + 108.2598.10^3 \qquad \text{(18b}$$

$$d_1 = s^3 + s^2 . 2.4352.10^3 \tag{18c}$$

$$d_2 = s.1.3536.10^6 + 286.1094.10^6 \tag{18d}$$

5.3 Simulations Using the Mathematical Model

The simulation performance result of mathematical model could be seen in Figure 5 and Table 3. Then to this it was used the mathematical model described in equation 18. It was shown that the end of 10 cycles of operation of the control system is increased the potential for the desired set value.

The horizontal axis indicates the time in seconds and the vertical axis the potential correction in volts, due the action of system control. The blue curve represents the response to a control step 0.0952 V applied to the first cycle of 50 kHz operation.

The simulation began with the definition of gains to be applied to the mathematical model, the potential sensor, the pulse width modulation controller (PWM). Then, it was adopted for the mathematical model the value of 0.095 V by the control step, in this case the gain of the plant model was 0.01G (s), and by way of simplification, the sensor and the gains of the controller PWM were modeled with unit values.

As shown in Table 3 it could be seen that in the course of 10 cycles, the system reaches the set point, showing that the setting adopted was successful.

As showed as on Figure 5 for the applied step, it can be seen that, from start, 90% of the additional set occurs on 2 ms (less than 60 Hz wave period), and the system is stable on 14 ms.

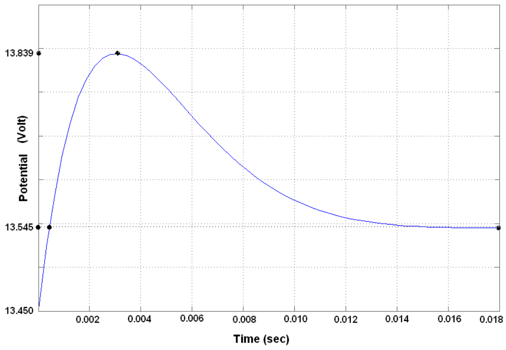


Figure 5: Performance of the controller to implement the first step to correct the potential to the controller. It was using the electrical model shown in equation 17, Author.

Table: 3 Mathematical Simulation of closed loop implementation and controller design

Cycle	Potential (V)	Reference (V)	Action
0	13.450	14.400	on
1	13.545	14.400	on
2	13.640	14.400	on
3	13.736	14.400	on
4	13.831	14.400	on
5	13.926	14.400	on
6	14.021	14.400	on
7	14.116	14.400	on
8	14.212	14.400	on
9	14.307	14.400	on
10	14.402	14.40	off

The basic feature of the transient response of a closed loop system depends on the closed loop poles location. If the system loop gain is variable, then the location of the closed loop poles will depend on the loop gain value selected. Therefore it is important that the poles are known as closed-loop moving on the complex plane (s), as the loop gain varies, as shown in Figure 7.

The closed-loop poles are the roots of the characteristic equation. A method for determining the roots of a characteristic equation is called the root locus, and allows that the roots of the characteristic equation are represented graphically for all values of the system parameter. Usually the gain varies between zero and infinite.

Figure 7 was used to evaluate the stability of the DC-DC converter with the support of equation 18 and the graphical representation of the characteristic equation roots for the closed loop. Then, it can be concluded that this closed loop system is rather stable, because the roots of its characteristic equation stood in the left half of the s-plane (on the complex frequency domain), (Ogata 2011)

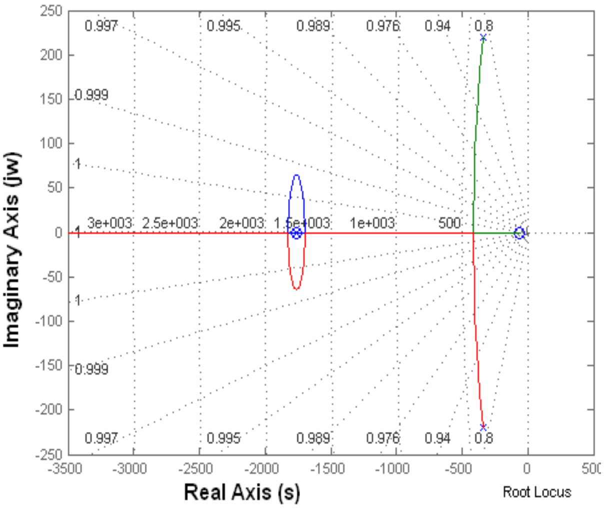


Figure 7: The Characteristic Equation Stability Analysis by the Root Locus Method, Author.

It is also useful to analyze the system using the Nyquist stability criteria for Mapping Theorem, which states that a closed loop system is stable if the outline of the entire right half of "s" (clockwise) with the frequency (w), varying from - ∞ to + ∞, on the imaginary axis (jw), there is no involvement on the origin path of the characteristic equation 18. At this situation there will be no poles and then the closed-loop system is stable. Also, Figure 8 is able to show that the system model is stable for graphical representation based on Nyquist path, because there is no engagement of the source (Ogata 2011).

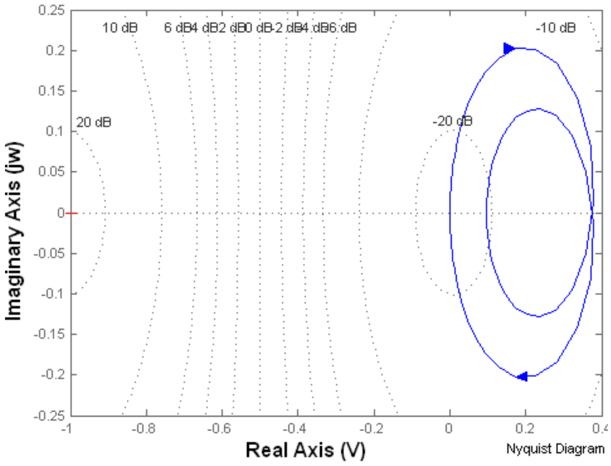


Figure 8: The Characteristic Equation Stability Analysis by the Nyquist Mapping Theorem, Author.

6 CONCLUSIONS

The developed system proved to be stable according to the Root Locus Method and Nyquist mapping theorem analyses performed.

The contribution given by the present developed model showed to be very important to generated useful potential for practical purposes, increasing the overall electrical efficiency, using the MCC500 system, built using Brazilian technology exclusively for stationary purposes power systems.

REFERENCES

Alam, M.S., Bala, B.K., Huq, A.M.Z., Matin, M.A., 1991. A Model for the Quality of Life as a Function of Electrical Energy Consumption. *Energy*, 16 (4), 739–745.

Bonifácio, R. N., Paschoal, J. O. A., Linardi, M. 2011. Catalyst Layer Optimization by Surface Tension Control During Ink Formulation of Membrane Electrode Assemblies in Proton Exchange Membrane Fuel Cell. *Journal of Power Sources, published by Elsevier,* 196, 4680-4685.

Brown, L. F., 2001. A Comparative Study of Fuels for On-Board Hydrogen Production for Fuel-Cell-Powered Automobiles, in*: International Journal of Hydrogen Energy*, 26(4), 381–397.

Close, M. C, 1980. *Circuitos Lineares.* 1980. 2nd edition, 550p, Rio de Janeiro, Brazil: LTC Editora.

Cunha, E. F., 2009. *Avaliação e Aplicação de Tecnologias de Células a Combustível Tipo PEMFC Desenvolvidas no IPEN em um Módulo de 500 We de Potência Nominal.* Thesis (PhD). Instituto de Pesquisas Energéticas e Nucleares - Universidade de São Paulo.

DOE - U.S. Department of Energy, 2011, Fuel Cell Technologies Market Report. U.S. Department of Energy publication, 68 p.

Friedrich, A., Carrette, L., Stimming, U., 2000. Fuel Cells: Principles, Types, Fuels, and Applications. *European Journal of Chemical Physics and Physical Chemistry*, 1(4), 162-193.

Granville, Smith, Longley, 1998. *Elementos de Cálculo Diferencial e Integral,* Edição Brasileira. Rio de Janeiro: Editora Científica.

Jonckheere, E. A., Coutinho, M. G. Chih-Yung Cheng, A., 2002. Computational Geometry Approach to Simplicial Nyquist Maps in Robust Stability, American Control Conference, Los Angeles, USA. *Proceedings of the American Control Conference Baltimora,* Maryland, June 1994. (213)740-4457

Kinoshita, K., 2001. *Electrochemical uses of carbon*. Available from : http://electrochem.cwru.edu/ encycl/art-c01-carbon.htm [accessed 23 March 2012].

Liguras D. K.; Goundani, K.; Verykios, X. E. 2004. Production of hydrogen for fuel cells by catalytic partial oxidation of ethanol over structured Ni catalysts, in: Journal of Power Sources, published by Elsevier, 130(1-2),30–37.

Linardi, M., 2010. *Introdução à Ciência e Tecnologia de Células a Combustível*. São Paulo: Artliber Editora.

Mathworks, 2007. *Matlab7/Simulink: Real-Time Workshop, TargetBox*. Available from: http://www.mathworks.com/ [accessed 18 October 2011]

Nahvi and Edminister, J., 2011. *Schaum's Outline of Electric Circuits,* Fifth Edition. São Paulo: Publisher: MCGRAW-HILL.

Ogata, K. *Modern Control Engineering*. USA, Prentice Hall Inc, 5th edition, 2011. 824p

Schoots, K., Kramer, G. J., Van Der Zwaan, B. C. C., 2010. Technology Learning for Fuel Cells: an Assessment of Past and Potential Cost Reductions. *Energy Policy*, 38(6), 2887-2897.

Shang, J. L., Pollet , B. G., 2010, Hydrogen fuel cell hybrid scooter (HFCHS) with plug-in features on Birmingham campus. International Journal of Hydrogen Energy, 35, 12709-12715.

Spinacé, E. Oliveira Neto, A., Linardi, M., 2003. Electro-Oxidation of Ethanol on PtRu/C Electrocatalysts Prepared from (n-C2H4)(Cl)Pt (uCl)2Ru(Cl)(n3, n3-C10H16). *Journal of Power Sources*, 124, 426-431.

Szwarc, A., 2007. *Bioenergia e Meio Ambiente.* Proceedings of Conferência Nacional de Bioenergia - BIOCONFE, USP. Available from: http://www.usp.br/bioconfe/downloads.htm [accessed 18 October 2011].

VAIDYA, P. D.; RODRIGUES, A. E. 2006. Insight into Steam Reforming of Ethanol to Produce Hydrogen for Fuel Cells, in: *Chemical Engineering Journal*, 117(1), 39-49.

Wendt, H.; Linardi, M.; Aricó, E. 2002. Células a Combustível de Baixa Potência Para Aplicações Estacionárias. *Química Nova*, 25, 470-476.

A METHOD PREDICTING TEMPERATURE RISE OF OIL-HYDRAULIC SYSTEM CONSIDERING HEAT BALANCE BETWEEN OIL-PASSAGE AND HOUSING

Kouki TOMIOKA[a], Kazuhiro TANAKA[b], Hiroshi HIGO[b], Fumio SHIMIZU[b]

[a] NITTO DENKO Co., Ltd. Engineering Center, 18 Hirayama Nakahara-cho Toyohashi Aichi, 441-3194, Japan
[b] Dept. of Mechanical Information Science and Technology, Kyushu Institute Technology,
680-4 Kawazu Iizuka, Fukuoka, 820-8502, Japan

[a]kouki_tomioka@gg.nitto.co.jp, [b]kazuhiro@mse.kyutech.ac.jp,
[c]higo@tech-i.kyutech.ac.jp, [d]shimizu@mse.kyutech.ac.jp

ABSTRACT

Aircrafts have actuators as key elements such as electro-hydrostatic actuators (EHA). They are so compact and highly-pressurized. The small heat capacity or the small surface of heat dissipation for high heat load causes high temperature rise of the system. This paper summarizes the thermal modeling effort and its utility in designing a thermal predicting solution for an oil-hydraulic system. Lumped-parameter thermal models coupled with three-dimensional analysis method by CFD (Computational Fluid Dynamics) have been developed to characterize component heat transfer within their operating environments. The results on time-variant temperature rise in each component, simulated by the method, agree well with the experimental data.

Keywords: oil-hydraulic system, temperature prediction, bondgraphs, CFD

1. INTRODUCTION

An oil-hydraulic system has become smaller and highly-pressurized especially in aircrafts, where electric actuators, such as electro-hydrostatic actuators (EHA), are used for moving flight control surfaces. In the industrial field, the use of EHA is an integral part of the more-electric aircraft concept to replace inefficient centralized hydraulic systems with power-on-demand electrical systems (Navarro, 1997; Johansson, Anderson, and Krus 2001; Takebayashi and Hara 2004; Rito, Denti, and Galatolo 2010). Removing the centralized hydraulic system will, however, eliminate an effective heat transfer network, thus resulting in an aircraft with less overall heat to reject but with localized “hot spots”. Because of small heat capacity and small surface area for heat dissipation in EHA, the system is raised to high temperature while it works for long time under highly loaded condition. The high temperature rise causes deterioration of working oil and the system may become hard to control finally. From this viewpoint, it is important to study a practical way to predict precisely system temperature rise.

Energy losses cause temperature rise in an oil-hydraulic system. In this case, modeling and simulation by the bondgraphs method is effective to predict the temperature rise because the bondgraphs method is based on energy balance (Karnopp, Margoils, and Rosenberg 1975; Dransfield 1981; Thoma 1990). Some studies have been performed to predict the temperature rise of an oil-hydraulic system considering heat generation and heat transfer in components. Yamamoto, Tanaka, Nakanishi, and Tarumi (1997) analyzed heat generation and heat transfer in an oil circuit of a mobile crane by the bondgraphs method and obtained good agreement between the calculated results and the experimental results. Johansson, Anderson, and Krus (2001) analyzed heat generation from a hydraulic motor embedded in an aircraft EHA system and heat transfer from oil passages. Takebayashi and Hara (2004) and Tomioka, Tanaka, and Nagayama (2005) proposed a new idea to predict the temperature rise of a hydraulic pipe by considering heat exchange between the working oil and the pipe-housing. In their studies, lumped-parameter thermal models were developed to characterize component heat transfer within their operating environments and the bondgraphs method based on the models was applied to predict temperature rises of the working oil and the pipe-housing by coupling with three-dimensional (3D) heat conduction analysis. The time-variant temperature distributions were well predicted both in the pipe housing and in the working oil. Thermal management designed for time averaged heat load for an oil-hydraulic system is generally considered to be not adequate since those heat loads are highly transient and localized in nature. However, the studies prove reliability and effectiveness on temperature prediction using the bondgraphs simulation where the calculating time step is very small.

Then, Tanaka, Tomioka, Fuchiwaki, and Suzuki (2011) developed the above idea to predict time-variant temperature distributions in an oil-hydraulic cylinder by comparing with 3D Computational Fluid Dynamics (CFD) analysis, because the temperature distribution changed while the cylinder works for long time. The study showed that in order to predict more precisely the temperature change inside the cylinder while the piston was moving, it was unsuitable to increase cell number merely and the optimal number of cells existed in modeling using lumped parameter modeling system. Because there were two characteristic regions inside the cylinder, one-dimensional flow pattern region like laminar flow and three-dimensional flow pattern region like turbulent flow. The latter flow pattern region should be modeled as one lumped cell.

In this study, the above-mentioned predicting methods were applied to an actual hydraulic system and the results calculated by the method were compared with the experimental data in order to validate the temperature-predicting method.

2. NOMENCLATURE

A [m^2]	heat dissipation surface area of cylinder
a [m]	representative length of inlet head port
b [m]	length from cylinder end to inlet port center
c_p [J/kgK]	specific heat at constant pressure
D [m]	cylinder diameter
E [J]	internal energy
H [J]	enthalpy
h [W/m^{2}K]	heat transfer coefficient
LoD [m]	characteristic distance of boundary between 3D and 1D flow pattern region
N [s]	reciprocating motion cycle of piston
P [Pa]	pressure
Q [m^3/s]	volumetric flow rate
R_c [m]	standard diameter of inlet head port
$\dot{q}$ [W]	amount of heat inflow
T [K]	temperature
t [s]	time
$\mathbf{U}$ [m/s]	velocity vector
V [m^3]	volume of cylinder head chamber
v [m/s]	velocity
W [W]	power consumed for outside
x [m]	axial location of piston head
δ [-]	Kronecker delta
λ [W/mK]	thermal conductivity
μ [Pas]	viscous coefficient
ν [m^2/s]	kinematic viscous coefficient
ρ [kg/m^3]	density

Suffix

c	cylinder
in	inflow
p	piston
$1D$	one-dimensional
$3D$	three-dimensional

3. TEMPERATURE PREDICTING METHOD IN EACH COMPONENT OF HYDRAULIC CIRCUIT

3.1. Test Experimental Circuit

Figure 1 shows a diagram of the test hydraulic circuit, which mainly consists of a pump, pipes, a control valve, a relief valve, a hydraulic cylinder, and a tank. The control valve is a closed center proportional solenoid valve. TA, TB, and TT indicate the temperature measurement of working oil at Port A, Port B, and the tank, respectively.

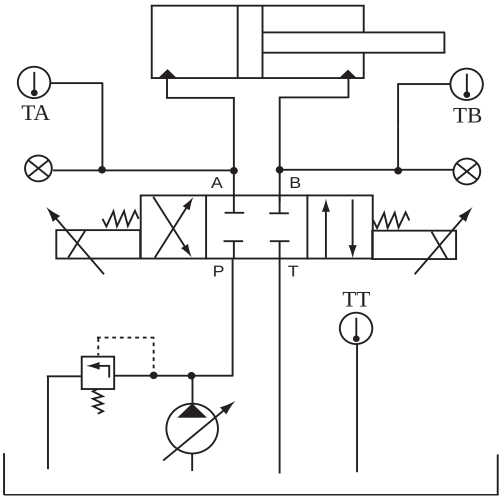


Figure 1: Oil-Hydraulic Circuit

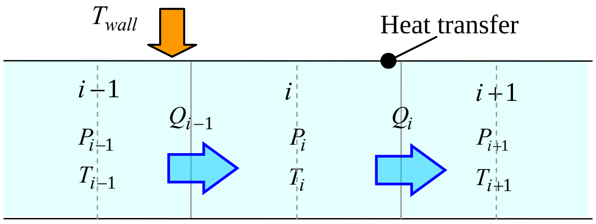


Figure 2: Pipe Model for Heat Balance

3.2. Temperature Rise Prediction in Pipe and Pipe-Housing (Tomioka, Tanaka, and Nagayama 2005)

The diagram of this heat balance type is shown in Fig. 2, where suffix i means the index of the pipe component in hand, i-1 the index of a neighboring upstream side component and i+1 the neighboring downstream side component. Temperature of each pipe component is simulated by heat balance among the heat generation in the component, the heat transfer from the pipe wall to the internal oil and heat transportation between the neighboring components of the pipe. Assuming that working oil does not work to outside of the pipe component i, rate of internal energy increase is expressed as follow.

$$\frac{dT_i^n}{dt} = \frac{1}{c_v \rho V_i}\left\{(P_{i-1} - P_i)^n Q^n + c_v \rho Q^n (T_{i-1} - T_i)^n + \dot{q}_i^n\right\} \quad (1)$$

The first term in the right-hand side indicates energy loss by pressure difference and the second term energy transferred from the upstream side component. Third term $\dot{q}$ is expressed by Eq. (2).

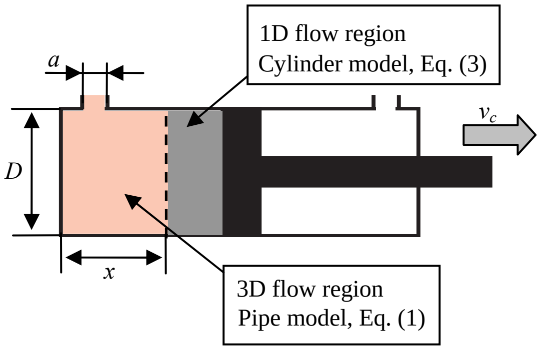


Figure 3: Heat Balance in Cylinder Head Chamber

$$\dot{q}_i^n = hA_h(T_{wall} - T_i^n) + \lambda A \frac{T_{i-1}^n - 2T_i^n + T_{i+1}^n}{2l} \tag{2}$$

The first term indicates heat amount transferred from the pipe wall to the inner working oil and the second term heat conduction between neighboring components of pipe. Heat generation and heat transfer in the pipe is calculated by Eq. (1) in each component.

3.3. Temperature Rise in Piston Head Chamber

(Tanaka, Tomioka, Fuchiwaki, and Suzuki 2011)

Figure 3 shows a schematic figure of heat balance in a cylinder head chamber. Temperature in the chamber is calculated as the sum of amount of heat inflow and outflow of working oil, volumetric change of working oil, and heat dissipation to cylinder housing, which is represented by Eq. (3).

$$\frac{dT_c}{dt} = \frac{1}{c_p \rho V_c}\left\{ c_p \rho Q T_i + hA_c(T_{wall} - T_c) - c_p \rho T_c \frac{dV_c}{dt} \right\} \tag{3}$$

$$\begin{aligned} &if \quad Q \geq 0(in) \; T_i = T_{in} \\ &\qquad Q \leq 0(out) \;\; T_i = T_c \end{aligned}$$

In the above equation, the first term of right hand side indicates the heat flux driven by inflow or outflow with piston motion, the second the heat flux transferred between the cylinder chamber and cylinder housing, and the third the effect of cylinder volume change, respectively.

From the results by 3D CFD analysis on the flow inside the cylinder chamber during the piston movement, it was verified that the inside flow could be modeled into two patterns characteristically, one of which was 3D vortex flow pattern near the inlet port and another of which was 1D parallel flow pattern near the piston head. So, it became possible to predict precisely the temperature distribution in the cylinder head chamber preferably by incorporating the internal flow patterns with the lumped parameter models.

The dimensionless distance from the end-wall of the head chamber to the location of boundary between 3D and 1D flow pattern region was defined as *LoD*, where the representative length was *a*. Dimensional analysis was performed and derived that *LoD* was a function of the following three terms, such as the product of the dimensionless length *D/a* and one Reynolds number expressed by $\rho D v_p/\mu$, another Reynolds number expressed by $\rho D v_{in}/\mu$, and another dimensionless length *b/a*.

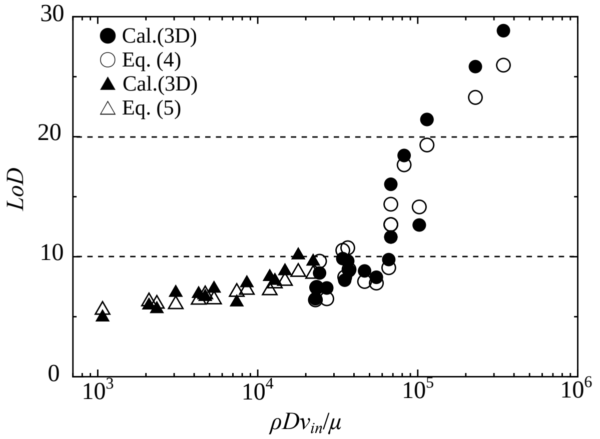


Figure 4: *LoD* with a Reynolds Number

After many simulations based on 3D CFD method, the results were obtained in Fig. 4, where the data expressed by “Cal.(3D)” were the calculated results. The abscissa and vertical axis indicate the Reynolds number expressed by $\rho D v_{in}/\mu$ and the dimensionless length *LoD*, respectively. Figure 4 gives the line of splitting at $\rho D v_{in}/\mu = 2.3\times10^4$ between one region with lower gradient and another region with higher gradient. By coupling this result with the above described result of dimensional analysis, the following characteristic equations were derived crossing the split line.

LoD:

(i) $\rho D v_{in}/\mu \geq 2.3\times10^4$

$$\frac{x}{a} = 1.44\left(\frac{D}{a}\cdot\frac{\rho D v_p}{\mu}\right)^{-0.014}\left(\frac{\rho D v_{in}}{\mu}\right)^{0.19}\left(\frac{b}{a}\right)^{0.27} \tag{4}$$

(ii) $\rho D v_{in}/\mu < 2.3\times10^4$

$$\frac{x}{a} = 0.015\left(\frac{D}{a}\cdot\frac{\rho D v_p}{\mu}\right)^{-0.69}\left(\frac{\rho D v_{in}}{\mu}\right)^{0.96}\left(\frac{b}{a}\right)^{0.098} \tag{5}$$

Figure 4 shows the comparison of *LoD* obtained through the above characteristic equations, expressed by “Eq. (4)” or “Eq. (5)”, with *LoD* obtained through 3D CFD calculations. Though there is small difference less than 14%, the both results agree well. It is verified that *LoD* can be predicted suitably through the above characteristic equations.

In predicting the temperature rise in the chamber one-dimensionally, Eq. (1) is used in the 3D flow pattern region because its volume does not change and Eq. (3) is used in the 1D flow pattern region because its volume changes with time when the piston moves.

3.4. Temperature Rise in Valve and Tank

Because a hydraulic valve generates the highest pressure loss among any component in a hydraulic system, it is very important to predict heat generation there. A schematic diagram to predict temperature rise

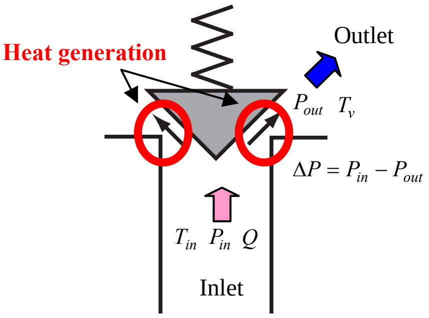


Figure 5: Valve Model

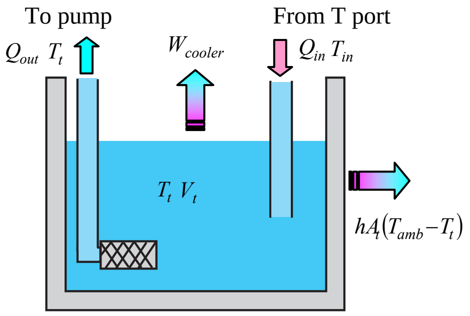


Figure 6: Tank Model

in a valve is shown in Fig. 5, which shows an example of poppet type valve. When working oil flows through throat part between the valve body and valve housing, high pressure loss appears and it changes to be heat generation.

$$\frac{dT_v}{dt} = \frac{1}{c_p \rho V_v}\left\{\Delta PQ + c_p \rho Q\left(T_{in} - T_v\right) + hA_h\left(T_{wall} - T_v\right)\right\} \quad (6)$$

A heat conduction term is neglected in Eq. (6), because the working oil velocity is fast in the valve. This equation covers the case of spool type valve.

Figure 6 shows a schematic diagram for heat balance in a tank. A volume of working oil in the tank changes time-dependently due to the sum between inflow rate from T-port of a valve and outflow rate to a hydraulic pump. Consequently, Eq. (3) with a term for time-variant volume is applicable in this case. The next equation consists of terms included in Eq. (3), heat flux flowing out, and heat release by an oil cooler.

$$\begin{aligned}\frac{dT_t}{dt} = \frac{1}{c_p \rho V_t}\Big(& c_p \rho Q_{in} T_{in} - c_p \rho Q_{out} T_t \\ & + hA_t\left(T_{amb} - T_t\right) + W_{cooler} - c_p \rho T_t \frac{dV_t}{dt}\Big)\end{aligned} \quad (7)$$

In this equation, the heat release from the tank wall and the oil cooler is determined by the experimental measurement.

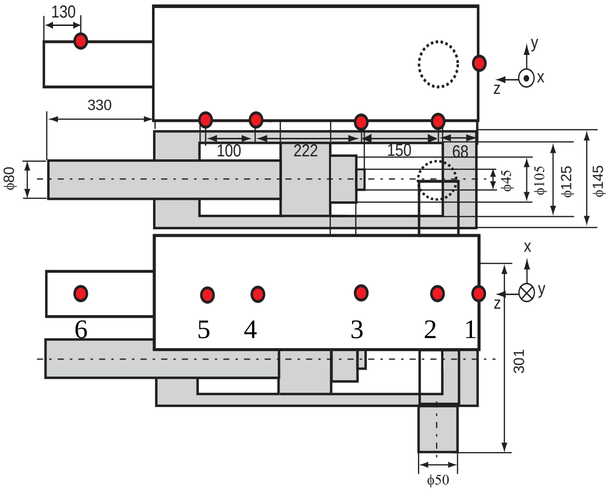


Figure 7: Schematic Figure of Test Cylinder and Measurement Points

3.5. Temperature Rise in Housings

Temperature distribution in housings can be calculated by use of a heat conduction equation, shown in Eq. (8).

$$\rho c \frac{\partial T}{\partial t} = \nabla \bullet \lambda\left(\nabla T\right) \quad (8)$$

In this study, a calculating program was made according to a finite volume method is used with Crank-Nicholson method for time-iteration. Reliability of the program was verified by comparing with the result by a commercial code, ANSYS CFX-10 (ANSYS 2005).

4. EXPERIMENTS AND CONDITIONS

The experimental measurements of temperature were conducted using the equipment shown in Fig. 1. Figure 7 shows a schematic figure of the test cylinder and measurement points in the experiment. The cylinder is a widely-used type, of which outer diameter is 145 mm and inner diameter 125 mm. In this temperature measurement test, a load for the cylinder is zero.

Temperature measurement was performed at three points in the circuit at TA, TB, and TT, which were shown in Fig. 1, and at six points on the cylinder outer surface, which were shown in Fig.7 as the number from 1 to 6.

The test conditions are shown in Table 1. In the test experiment, the temperature changes were measured when the piston was moving with constant oscillation amplitude and frequency.

Table 1: Experimental Conditions

Motion of Piston	Frequency : 0.2 Hz Amplitude : 50 mm
Supply Pressure	10 MPa
Experimental Time	1200 s
Working Fluid	ISO-VG32
Ambient Temperature	23 deg.

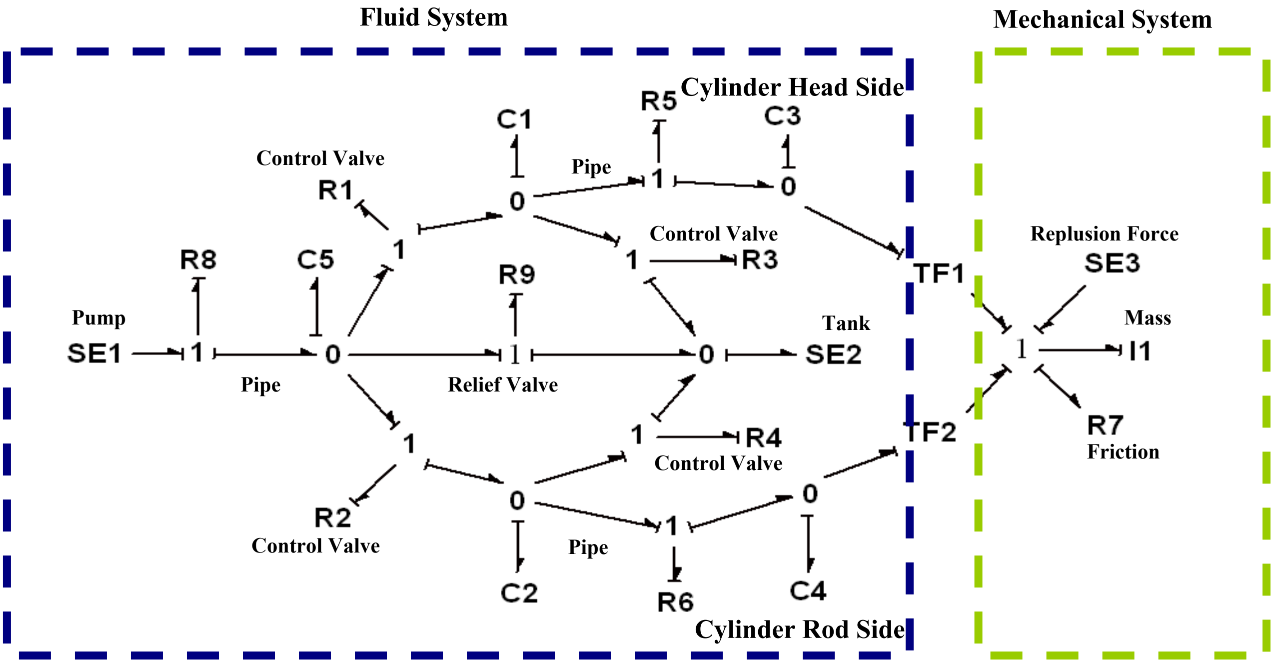


Figure 8: Bondgraph Model for Oil-Hydraulic System

5. MATHEMATICAL PREDICTION MODEL

The dynamic characteristics and temperature rise of the working oil in the system were calculated by 1D model and the temperature rise in the cylinder housing was calculated by 3D numerical method. By coupling these calculation methods, the temperature rise prediction method was established with considering dynamic characteristics of the system. This chapter shows models and parameters used in 1D modeling and simulation of dynamic characteristics and 3D numerical analysis method.

5.1. Bondgraphs model for Dynamic Characteristics

Dynamic characteristics of the working oil were analyzed by 1D modeling and simulation method using the bondgraphs. The system bondgraph correspondent to the test hydraulic system, shown in Fig.1, is shown in Fig. 8, which was introduced as a representative bondgraphs in (Dransfield 1981).

The respective element such as R1, R2, R3, and R4-element indicates resistance effect of the control valve. When the piston moves from the piston-head side to the piston-rod side, the valve-ports corresponding to R2 and R3-elements are closed and those to R1 and R4 are opened. When the piston moves in the opposite direction, the valve-ports corresponding to R2 and R3-element are opened and those to R1 and R4 are closed. The signals of valve control are given by an external program. TF-element indicates energy transfer between the fluid system and the mechanical system. C1, C2, and C5-element indicate the capacitance effect of the pipe corresponding to the respective pipe in the actual circuit. R5, R6, and R8-element indicate the pressure loss in the pipe from the control valve to the piston head

Table 2: Characteristic Equations for System Bondgraphs

Elements	Characteristic equation
SE1 (Pump)	P = Supply pressure (10 MPa)
SE2 (Tank)	$P = 0.0$ Pa
SE3 (Stop piston motion at the dead-end location)	$F = \mathit{Repulsion\ force}$
R1, R2, R3, R4, R9 (Pressure loss in Valve)	$Q = cA\sqrt{\frac{2\Delta p}{\rho}}$
R5, R6, R8 (Pressure loss in pipe)	$Q = \frac{\pi d^4}{128\mu l}\Delta P$
R7 (Friction loss between piston and cylinder)	$F = cv$
C1, C2, C5 (Capacitance effect of pipe volume)	$P = \frac{K}{V}\int Qdt$
C3 (Capacitance effect of piston head chamber; A_h=cross-section of piston head chamber, *disp.*=piston displacement)	$P = \frac{K}{V + A_h \cdot disp}\int Qdt$
C4 (Capacitance effect of piston rod chamber; A_r=cross-section of piston rod chamber)	$P = \frac{K}{V - A_r \cdot disp}\int Qdt$
TF1, TF2 (Energy transfer between fluid and mechanical system)	$F = PA\ ,\quad v = \frac{Q}{A}$
I1 (Inertial effect of piston mass)	$v = \frac{1}{m}\int Fdt$

Table 3: Parameters in each element

Elements		Variables
Pipe length and diameter	R8, C5	Length = 1.0 m, Diameter = 10 mm
	C1, R5	Length = 6.0 m, Diameter = 9.5 mm
	C2, R6	Length = 5.5 m Diameter = 14 mm
Discharge coefficient	R1, R2, R3, R4	Discharge coefficient = 0.65 Valve opening area = Variable area
Valve opening area	R9	Discharge coefficient = 0.65 Valve opening area = 1.0×10^{-5} m^2
Cylinder head	C3, TF1	Length = 0.177+ *dips.* m Diameter = 0.125 m
Cylinder rod	C4, TF2	Length = 0.177- *dips.* m Head diameter = 0.125 m Rod diameter = 0.08 m
Mass of piston	I1	39 kg
Piston friction	R7	5,000 N/m
Tank surface	SE2	0.5 m^2

chamber, from the valve to the piston rod chamber, and from the pump to the control valve, respectively. C3 and C4-element indicates the capacitance effect of the cylinder head chamber and the cylinder rod chamber, respectively. SE1 and SE2 indicate the pump supplied pressure and the tank pressure. I1-element indicates the inertial effect of the piston mass and R7-element the frictional force between the piston and the cylinder. SE3-element is added as the repulsive force to stop the piston motion when the piston reaches the dead-end location of the cylinder.

Table 2 shows a list of characteristic equations used in each element. In this table, the volume of the cylinder chamber is variable because the piston displacement causes the chamber volume. Table 3 shows a list of parameters used in each element, which were measured in the test experimental apparatus.

5.2. 1D model for temperature prediction

For the temperature calculation, Eq. (3) is used in the cylinder, Eq. (1) in the pipes, Eq. (6) in the control valve, and Eq. (7) in the tank, respectively. Here, the temperature rise in the pump is not included.

A schematic diagram for 1D temperature analysis model is shown in Fig. 9, where an arrow in the figure shows heat flow direction. When the piston moves from the head side to the rod side, the heat flow calculation way is shown as follows. After the valve-ports, corresponding to R2 and R3 in Fig. 8, are closed, the working oil flows into the piston head chamber through Pipe 1, Valve 1, and Pipe 2 in Fig. 9. The heat of the inflow oil is exchanged with the cylinder housing, the temperature distribution of which is solved by 3D analysis, through Interface 1. On the other hand, the working oil in the piston rod chamber flows out after exchanging its heat with the housing through Interface 2. The outflow oil reaches the tank through Pipe 3 and Valve 4.

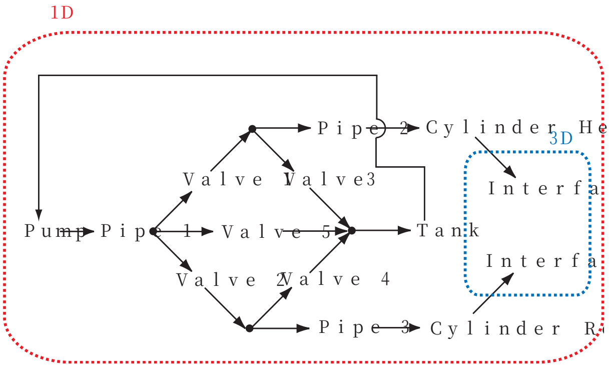


Figure 9: Heat Flow Model for Oil-Hydraulic System

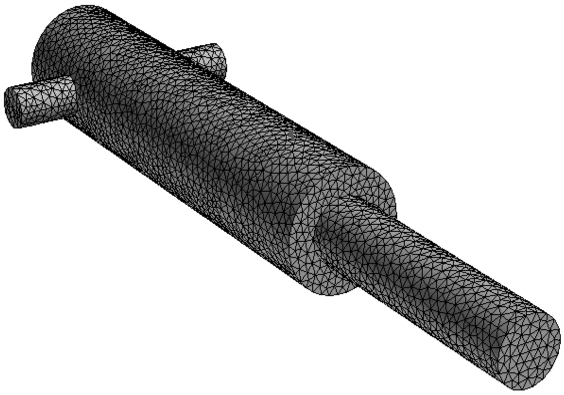


Figure 10: Numerical Grids for 3D Temperature Analysis

When the piston moves in the opposite direction, the inflow oil exchanges the heat through Interface 2 in the piston rod chamber and the outflow oil through Interface 1 in the piston head chamber, respectively. As just described, the working oil circulates in the circuit and the temperature rises at any location.

In this experiment, heat transfer coefficient was given as 10 W/m^{2}K from each component to the air and as 170 W/m^{2}K in the cooler, which was calculated reversely from the experiment.

5.3. 3D model for temperature analysis

Figure 10 shows the numerical grids for 3D calculation on temperature distributions of the cylinder. The number of numerical grid is 14000. Here, a calculating program was made according to a finite volume method and its reliability was verified by comparing with the result by ANSYS CFX-10.

Heat transfer coefficient in the cylinder head and rod chamber should be given previously before calculating heat exchange between 1D and 3D numerical analysis. Actually, because the flow rate from/to the cylinder varies with time, the heat transfer coefficient in the cylinder may also vary with time. It was difficult to determine the heat transfer coefficient

Table 4: Heat Transfer Coefficient at a *LoD*

LoD mm	h_{3D} W/m^{2}K	h_{1D} W/m^{2}K	h_r W/m^{2}K
94	118	75.4	76

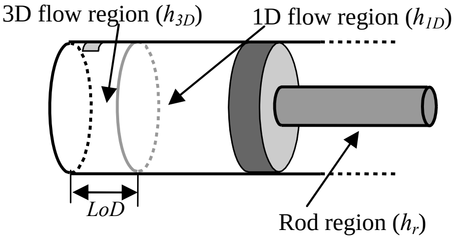


Figure 11: Definition of Heat Transfer Coefficient in Cylinder

Table 5: Analysis Conditions

Initial condition	Values measured in the experiment
Time step	3D : $dt = 0.1$ s, 1D : $dt = 1.0\times10^{-5}$ s
Physical property	Cylinder housing : $\rho = 7{,}850 [\mathrm{kg/m^3}]$, $c_p = 465 [\mathrm{J/kgK}]$, $\lambda = 43 [\mathrm{W/mK}]$, Working fluid : $c_p = 1{,}900 [\mathrm{J/kgK}]$, $K = 1.0\times10^{+9} [\mathrm{Pa}]$, $\lambda = 0.145 [\mathrm{W/mK}]$, $\rho = 860 [\mathrm{kg/m^3}]$, $\mu = 2.67\times10^{-2}\exp\left\{-2.9\times10^{-2}(T-40)\right\}$ [Pa.s], T [deg.]
Heat transfer	From housing to air : 10.0 W/m^{2}K
Ambient Air	$T = 23$ deg.

on the chamber inside wall, however, the coefficient was fixed to be constant in this study. Table 4 and Fig. 11 show the value of the heat transfer coefficients around the cylinder and their definitions, respectively. Table 5 shows the analysis conditions and values of the properties. Kinematic viscosity of the working oil is defined as a function of temperature.

6. RESULTS AND DISCUSSIONS

6.1. Temperature of Working Oil

Figure 12 shows the comparison on the temperature of the working oil with time at A- and B-port between in the experiment and in the calculation. In the figure, “TA” and ”TB” show the A-port and B-port, respectively. “Exp.” and “Cal.” show the experimental data and the calculated results obtained by the proposed prediction method.

Both in the experimental and the calculated results, the temperature at A-port was higher than that at B-port. This was because A-port was connected to the cylinder head chamber and B-port the cylinder rod chamber, respectively, and the cross-section of the head chamber was larger than that of the rod chamber. When the piston moved from the head to rod chamber, the flow rate into the head chamber became more than that out of the rod chamber and the pressure loss, defined in Eq. (6), became larger at A-port. In any case, the calculated results agreed well with the experimental results and the difference between both results was less than 1 deg. This fact showed that the temperature at A- and B-port behind the control valve was well predicted.

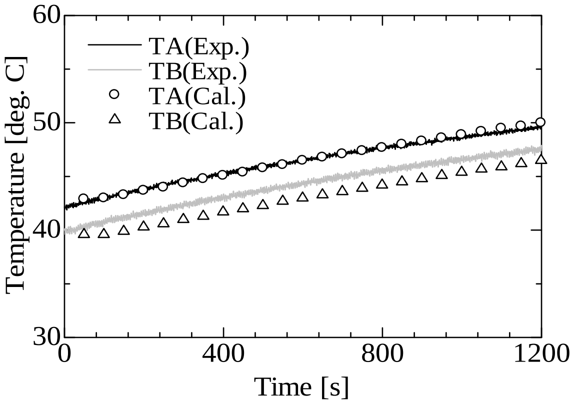


Figure 12: Temperature Change at A and B Port

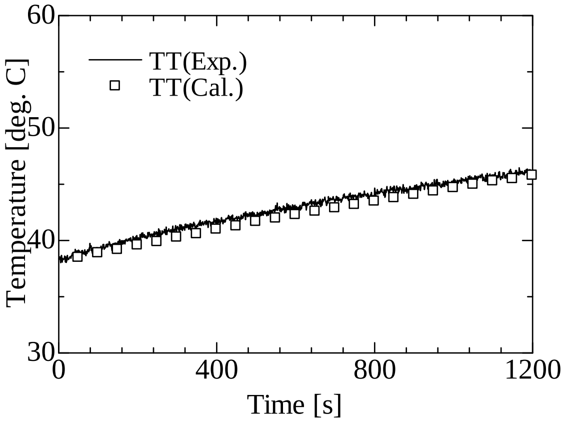


Figure 13: Temperature Change at Tank

Figure 13 shows the comparison on the temperature of the working oil in the tank with time between in the experiment and in the calculation. At t= 1200 sec., the tank temperature increased up to T= 45 deg., which was lower than the temperature at A-port, T= 50 deg. This was because the volume of the tank was much larger than that of the other component. The calculated results agreed well with the experimental results and it meant that it became possible to predict precisely the tank temperature if the heat transfer coefficient from the tank could be calculated in the experiment.

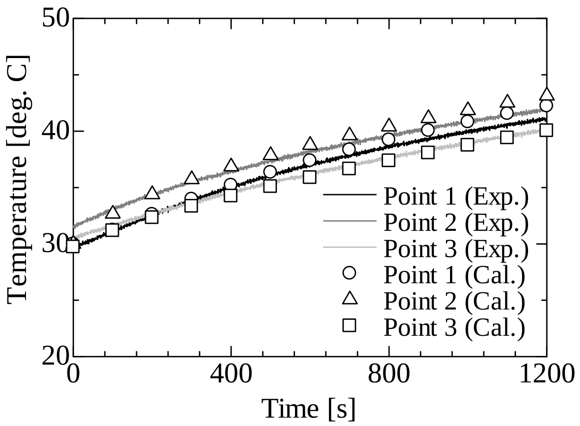


Figure 14: Temperature Change at Piston- Head Side

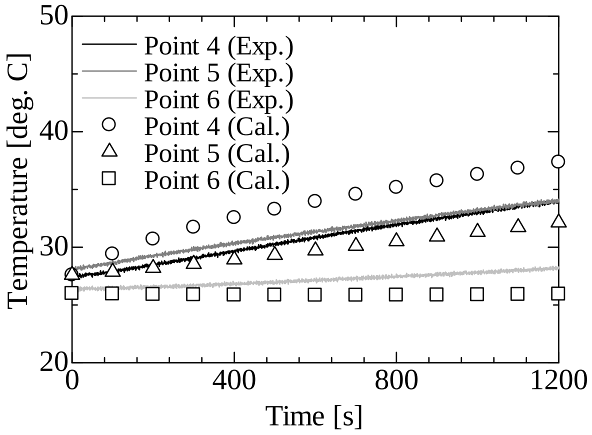


Figure 15: Temperature Change at Piston-Rod Side

6.2. Temperature of Cylinder Housing

Figure 14 shows the comparison on the temperature change at the measurement points, shown as 1, 2 and 3 in Fig. 7, on the cylinder surface in the head chamber side. The temperature increased with time at all points, because the temperature rise calculated according to 1D model was exchanged to the housing through Interface 1. The temperature at Point 2 was the highest among the three measurement points. This was because Point 2 was located just in the port inlet and the working oil with high temperature was continuously supplied there. The temperature at Point 1 was lower than that at Point2, because the wall thickness of the housing was different at Point 1, with the wall 25 mm thick, compared with that at Point 2, with the wall 10 mm thick. In any case, the calculated results agreed well with the experimental results and the difference between both results was less than 1 deg.

Figure 15 shows the comparison on the temperature change at the measurement points, shown as 4, 5 and 6 in Fig. 7, on the cylinder surface in the rod chamber side. Though the temperature at Point 4 was almost the same as that at Point 5 in the experiment, there were 5 deg. temperature differences between the two points in the calculation. One possible reason of the difference between the experimental and the calculated results might be that the contact thermal resistance between the piston rod and the cylinder, which existed actually in the experiments, was not considered in the calculation. Or, another possible reason might be that the rod chamber was modeled as one lumped parameter model because it was considered that the flow rate from/to the rod chamber was much less than that from/to the head chamber and the temperature in the rod chamber would be lower than that in the head chamber. The difference between the calculated and experimental results was less than 3 deg. at 1200 sec.

Table 6: Calculation Time

Exp. Time	Calculation Time	
1200 sec.	ANSYS-CFX (Estimation)	264 days
	Proposed method PC(Power of CPU: Pentium Xeon 3.6GHz)	92 min.

At Point 6, the temperature did not increase any more from the beginning. This was because the measurement location was far from the heat generating locations and the heat was transported merely by heat conduction in the piston rod.

6.3. Calculation time

Table 6 shows the comparison on the time for calculating temperature rise of the test hydraulic circuit between in the proposed method and in a commercial CFD code, ANSYS CFX-10. It took 92 min. for the proposed method and 264 days for the CFD code to calculate the dynamics and temperature changes of each component for the actual 1200 sec. In case of the CFD code, the time was estimated from the fact that it needed $1.9\times10^{+5}$ sec. to calculate temperature rise for 20 sec. in the actual time. The result showed that the proposed method gave very good performance.

7. CONCLUSION

In order to establish a method predicting temperature rise in an oil hydraulic system precisely and effectively, the proposed method was applied to an actual system. Dynamic characteristics of the system was analyzed by the bondgraphs and the temperature rise was calculated by 1D models, according to heat balance, coupled with 3D analysis to calculate temperature distributions in the housings. As a result, the followings become evident.

The experimental results and the calculated results on the temperature rise of the working oil were compared in the pipes, the valves and the tank, and the both results agreed well. It supported validity of the proposed method to predict temperature rise.

The calculated results on the temperature at the cylinder housing surface were compared with the experimental results. The both results agreed well at the piston head chamber side. On the other hand, there were differences between the both results at the piston rod chamber side. One possible reason of the differences might be the existence of the contact thermal resistance between the piston rod and the cylinder or the modeling method of the piston rod chamber.

The proposed method could predict the temperature rise in an oil-hydraulic circuit effectively with the calculation time much shorter than that by a commercial CFD code.

REFERENCES

ANSYS, 2005. *ANSYS CFX-Solver, Release 10.0.*

Dransfield, P., 1981. *Hydraulic Control Systems - Design and Analysis of Their Dynamics*, Berlin Heidelberg, Germany: Springer-Verlag.

Johansson, B., Anderson, J., and Krus, P., 2001. Thermal Modeling of an Electro-Hydrostatic Actuation System. *Proceedings of the International Conference on Recent Advantages in Aerospace Actuation Systems and Components*, June 13-15, Toulouse (France).

Karnopp, D. C., Margoils, D. L., and Rosenberg, R. C., 1975. *System Dynamics: A Unified Approach,* CA: John Wiley & Sons.

Navarro, R., 1997. Performance of an Electro-Hydrostatic Actuator on the F-18 Systems Research Aircraft. NASA/TM-97-206224.

Rito, G., Denti, E., and Galatolo, R., 2010. Object-oriented modeling of flight control actuation systems for power absorption assessment. *Proceedings of the 27th International Congress of the aeronautical science (ICAS2010)*, pp.1-11. Sept. 19-24, Nice (France).

Thoma, J. U., 1990. *Simulation by Bondgraphs,* Berlin Heidelberg, Germany: Springer-Verlag.

Takebayashi, W., Hara, Y., 2004. Thermal Design Tool for EHA. *Proceedings of the International Conference on Recent Advantages in Aerospace Actuation Systems and Components*, pp.15-20. November 24-26, Toulouse (France).

Tomioka, K., Tanaka, K. and Nagayama, K., 2005. Prediction method of heat generation and heat transfer in an oil-hydraulic system. *Proceedings of Int.Conf. on Integrated Modeling and Analysis in App-lied Control and Automation (IMAACA 2005),* pp. 89-94. Marseille (France).

Tanaka, K., Tomioka, K., Fuchiwaki, M., and Suzuki, K., 2011. New Concept of Modeling to Predict Temperature in Oil Hydraulic Cylinder Chamber Considering Internal Flow. *Proceedings of Int.Conf. on Integrated Modeling and Analysis in App-lied Control and Automation (IMAACA 2011),* pp.287-292, Sept. 12-14, Roma (Italy).

Yamamoto, K., Tanaka, K., Nakanishi, M. and Tarumi S., 1997. Analysis of Dynamic Behavior of a Hydraulic and Pneumatic Suspension Including Temperature Change Effect. *Proceedings of 5th Tri. Int. Symp. on Fluid Control, Measurement and Visualization (FLUCOME97),* pp.457-462. Vol. 1. Sept. 1-4, Hayama (Japan).

AUTHORS BIOGRAPHY

Kouki TOMIOKA,

He is a chief engineer in NITTO DENKO Co., Ltd. Engineering Center. He graduated from Tanaka's Lab. In Kyuhsu Insti. of Tech. He has wide interests from engineering research and development to music concert activity. He is also a good pianist.

Kazuhiro TANAKA,

He graduated from the university of Tokyo, got the first job as an assistant professor in Sophia Univeristy, and he moved back to the university of Tokyo. After then, he moved to Kyuhsu Insti. of Tech. and is a professor. He has interests for CFD analyses on flow patterns inside oi-hydraulic valves as well as Bondgraphs.

Hiroshi HIGO,

He graduated from the graduate school in Kyushu univeristy and he is a member of research at Prof. Tanaka's lab. He has interests to analyze and control the dynamics of oil-hydraulic and pneumatic actuating systems.

Fumio SHIMIZU

He graduated from the graduate school in Tokyo university of Agriculture and Technology, and he is an assistant professor. He has interests to analyze and design for the sound propagation system.

DYNAMICS OF TORQUE CONVERTER WITH LOCK-UP CLUTCH

Takeshi YAMAGUCHI $^{(a)}$**, Kazuhiro TANAKA**$^{(b)}$**, Katsuya SUZUKI**$^{(c)}$

$^{(a)}$Aisin AW Co., Ltd. Engineering Division, 10 Takane Fujii-cho Anjyo Aichi, 444-1192, Japan.
$^{(b)}$ $^{(c)}$Dept. of Mechanical Information Science and Technology, Kyushu Institute Technology,
680-4 Kawazu Iizuka, Fukuoka, 820-8502, Japan

$^{(a)}$ I23962_Yamaguchi@aisin-aw.co.jp, $^{(b)}$ kazuhiro@mse.kyutech.ac.jp, $^{(c)}$ kachandesu2002@yahoo.co.jp

ABSTRACT

The performance of a torque converter (T/C) has been one of important improvement for an automatic transmission equipped automobile. Improving its performance and efficiency is a key factor to saving fuel consumption. Moreover, the locking up operation or slipping control of an automatic transmission, which makes T/C efficiency higher, is another good opportunity for improving fuel economy. One-dimensional modeling of systems remains effective to recognize a mechanism of the system dynamics. Some bondgraphs models of T/C have been proposed and it becomes more effective for drivability and efficiency to make and add a model for dynamics of the lock-up clutch (L/C). A bondgraphs model of the L/C and a method of calculating parameters through CFD numerical analysis in the components inside T/C are proposed in the present study.

Keywords: torque converter, lock-up clutch, CFD, bondgraphs

1. INTRODUCTION

Figure 1 illustrates cut-away view of an automatic transmission, which consists of three major components; the torque converter (T/C), the gearing system and the hydraulic circuit called valve body. The T/C transfers power from the engine to the transmission gearing system and has been used in automatic transmissions for numerous applications such as passenger cars, trucks, buses and trains.

There are two important roles for the T/C. One is the reduction of vibration or noise from the engine by means of the automatic transmission fluid. Another important role is the multiplication of engine torque. Many studies have been carried out to predict hydrodynamic performance and to understand the flow field inside a T/C either experimentally or analytically using Computational Fluid Dynamics (CFD) ((Kim, Ha, Lim, and Cha 2008; Watanabe 1999; Yamaguchi 2002; Abe, Kondoh, Fukumura, and Kojima 1991; Brun and Flack 1995; By and Lakshminarayana 1995; Cigarini and Jonnavithula 1995). The analytical research to date regarding the flow field around the lock-up clutch (L/C) uses CFD simulations performed with steady-state conditions (Ejiri 2006; Abe, Maruyama, Hasegawa and Kondo 1999). A typical automotive T/C cross-section is shown in Fig. 2.

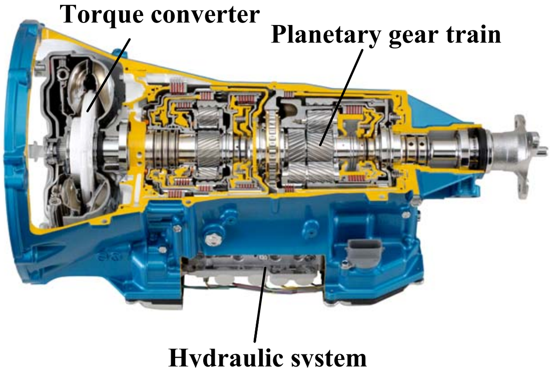


Figure 1: Automatic Transmission (T/C)

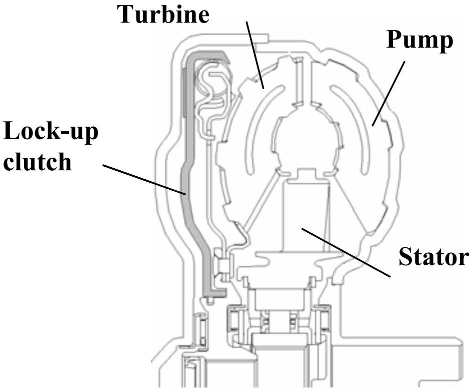


Figure 2: Torque Converter with the Lock-up Clutch

In order to improve the efficiency of an automatic transmission, it is often desired to engage the L/C as soon as possible to conserve the power flow from the engine. However, earlier engagement of the L/C is associated with a larger slip velocity against the input plate, making it increasingly important to manage the heat transfer on the friction paper. Moreover, sudden engagement of the L/C causes vibration and noise within the engine. For this reason, most research has focused either on heat management of the L/C or on the shudder mechanism. The T/C lock-up system is controlled by hydraulics. Understanding from a fluid dynamics perspective how the flow field around the L/C influences its behavior is a key factor in preventing shock and improving engagement time.

On the other hand, Ishihara introduced theoretical analysis of steady and unsteady characteristics of a T/C (Ishihara and Emori 1966; Hoshino and Ishihara 1990), providing the basic theory of the T/C and the design basis for its development. With the aim of increasing T/C transmission efficiency to meet the demand for higher fuel efficiency vehicles, recently many automobiles have a L/C inside the T/C, which is wet-typed clutch. By using this clutch, input torque is transmitted to both the clutch and fluid divided. If the clutch is working fully, all input torque is transferred through the clutch, and the fluid loss does not occur. Then the T/C efficiency is raised higher. Some researchers have studied the effect of L/C (Tsangarides and Tobler 1985) and slip control (Kono, Itoh, Nakamura, Yoshizawa and Osawa 1995; Hiramatsu, Akagi and Yoneda 1985) of T/Cs. These studies require determination of the dynamic characteristics of T/Cs.

The studies by Ishihara and Emori (1966) have provided the fundamental design method for T/Cs. In developing a practical T/C, however, design engineers apply the design method proposed by Ishihara, which is a complex and time-consuming task. Design engineers, who are expected to achieve higher fuel efficiency design, have been seeking an easier method of analyzing T/C dynamic characteristics. For this demand, a bondgraphs is considered as a better method.

Hrovat and Tobler (1985) applied the bondgraphs method to represent a T/C composed of fluid machinery elements (pump, turbine and stator) without the L/C, based on Ishihara's theory (Ishihara and Emori 1966) on the unsteady characteristics and the bondgraphs model was renewed by Suzuki and Tanaka (2003) with static characteristics of the L/C. By reference to these studies, a bondgraphs model representing the dynamics of the L/C is added to the above-mentioned models and the renewed model becomes a bondgraphs model of the T/C including the dynamic characteristics of the L/C. And CFD numerical analysis has been performed to study the flow structure in various conditions and the L/C engagement time.

2. MODELING BY BONDGRAPHS FOR TORQUE CONVERTER

Figure 3 is a bondgraphs representation of basic equations based on the reference (Ishihara and Emori 1966). This bondgraphs contains 4-port I-field, connecting hydraulic elements and mechanical elements via modulated gyrators (MGY). The multi-port I-field is decomposed into single-port I-elements following the procedure by Breedveld (1984), as shown in Fig. 4 (Hrovat and Tobler 1985; Suzuki and Tanaka 2003). The bondgraphs, which may look complex, is easy to understand in terms of correspondence to actual components. With the bondgraphs of Fig. 4, it is easy to find the correspondence between the bondgraphs elements for pump, turbine and stator and actual counterparts. Each symbol is suffixed with an identification number: 1, 2 and 3 represents pump, turbine and stator elements, respectively. All of these elements are connected via the 1-junction in the lower center of the bondgraphs, indicating that the flow rate in the pump, turbine and stator is the same at this 1-junction. Pressure loss caused by shock and friction is represented by R-element, indicating that the pressure loss can be obtained from the flow rate represented by 1-junction for each impeller. From the conservation law of angular momentum, the following three equations of motion can be obtained for the respective impellers,

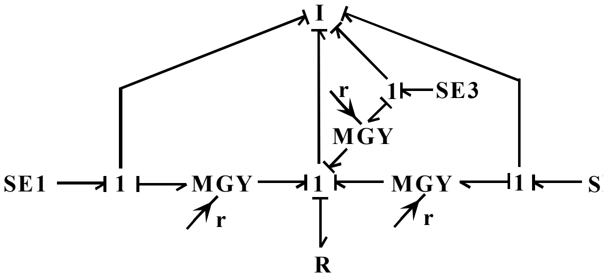


Figure 3: Bondgraphs for T/C

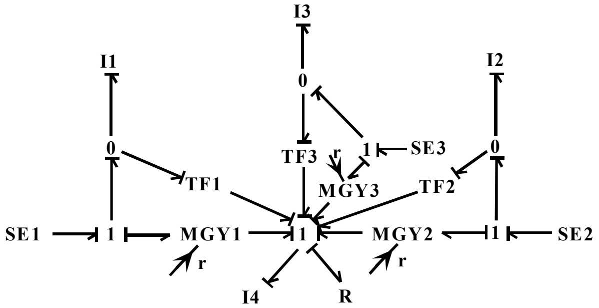


Figure 4: Decomposed Bondgraphs

For pump:

$$T_1 = I_1\dot{\omega}_1 - \rho S_1\dot{Q} + \rho Q\{(r_2\omega_1 - c_2\tan\alpha_{12})r_2 - (r_1\omega_3 - c_1\tan\alpha_{32})r_1\} \quad (1)$$

For turbine:

$$T_2 = I_2\dot{\omega}_2 - \rho S_2\dot{Q} + \rho Q\{(r_3\omega_2 - c_3\tan\alpha_{22})r_3 - (r_2\omega_1 - c_2\tan\alpha_{12})r_2\} \quad (2)$$

For stator:

$$T_3 = I_3\dot{\omega}_3 - \rho S_3\dot{Q} + \rho Q\{(r_1\omega_3 - c_1\tan\alpha_{32})r_1 - (r_3\omega_2 - c_3\tan\alpha_{22})r_3\} \quad (3)$$

Here, $\rho S_i \dot{Q}$ represents change in momentum of fluid existing in impeller; the dot indicates differential by time; S_i is a constant determined by the blade configuration (Ishihara and Emori 1966).

According to the conservation law of kinetic energy, the rate of kinetic energy increases $\dot{E}_k$ of fluid per unit time is represented by the following equation:

$$\dot{E}_k = T_1\omega_1 + T_2\omega_2 + T_3\omega_3 - \dot{E}_l \quad (4)$$

where $\dot{E}_l$ is energy loss of flow per unit time.

Therefore, another equation of motion holds as follows:

$$
\begin{aligned}
&\rho\Phi\dot{Q} - \rho(S_1\dot{\omega}_1 + S_2\dot{\omega}_2 + S_3\dot{\omega}_3) \\
&= \rho\{(r_2\omega_1 - c_2\tan\alpha_{12})r_2 - (r_1\omega_3 - c_1\tan\alpha_{32})r_1\}\omega_1 \\
&+\rho\{(r_3\omega_2 - c_3\tan\alpha_{22})r_3 - (r_2\omega_1 - c_2\tan\alpha_{12})r_2\}\omega_2 \\
&+\rho\{(r_1\omega_3 - c_1\tan\alpha_{32})r_1 - (r_3\omega_2 - c_3\tan\alpha_{22})r_3\}\omega_3 - P_L
\end{aligned} \tag{5}
$$

where Φ is a constant and can be defined as follows according to the blade configuration:

$$\Phi = \oint \sec^2 \alpha dl \tag{6}$$

P_L is a pressure loss defined by $\dot{E}_l = QP_L$, and is equal to the sum of shock loss and frictional loss, as expressed by Eq. (7). The coefficient of the shock loss is assumed to be 1.0 (Ishihara and Emori 1966).

$$
\begin{aligned}
P_L = 1/2\,\rho[&\{(r_1\omega_1 - c_1\tan\alpha_{11}) - (r_1\omega_3 - c_1\tan\alpha_{32})\}^2 \\
&+ \{(r_2\omega_2 - c_2\tan\alpha_{21}) - (r_2\omega_1 - c_2\tan\alpha_{12})\}^2 \\
&+ \{(r_3\omega_3 - c_3\tan\alpha_{31}) - (r_3\omega_2 - c_3\tan\alpha_{22})\}^2 + Lc_2^2]
\end{aligned} \tag{7}
$$

Here,

$$L = \sum_{i=1}^{3} \lambda_i \left\{1 + \left(\tan^2\alpha_{i1} + \tan^2\alpha_{i2}\right)/2\right\} \tag{8}$$

where λ_i is a loss coefficient.

In the above equations, the torque and angular velocity in a mechanical system component, and the pressure and flow rate in a hydraulic system, correspond to the effort and flow variables, respectively, of the bondgraphs. This correspondence can be expressed explicitly using a matrix notation as follows:

$$\mathbf{I}\dot{\mathbf{f}} = -\mathbf{G}(\mathbf{f})\mathbf{f} - \mathbf{R}(\mathbf{f}) + \mathbf{T} \tag{11}$$

where **f** is the matrix of the flow variables in the system bondgraphs, as expressed below:

$$\mathbf{f} = [\omega_1 \quad \omega_2 \quad \omega_3 \quad Q]^T \tag{12}$$

I is the matrix of inertial moment, as expressed below:

$$\mathbf{I} = \begin{bmatrix} I_1 & 0 & 0 & -\rho S_1 \\ 0 & I_2 & 0 & -\rho S_2 \\ 0 & 0 & I_3 & -\rho S_3 \\ -\rho S_1 & -\rho S_2 & -\rho S_3 & \rho\Phi \end{bmatrix} \tag{13}$$

$\mathbf{G} = \mathbf{G}(\mathbf{f})$ is the matrix of a modulated gyrator, as expressed below:

$$\mathbf{G} = \mathbf{G}(\mathbf{f}) = \begin{bmatrix} 0 & g^T \\ -g & 0 \end{bmatrix} \tag{14}$$

$$\mathbf{g} = \mathbf{g}(\mathbf{f}) = [\rho g_1 \quad \rho g_2 \quad \rho g_3] \tag{15}$$

where

$$g_1 = \{(r_2\omega_1 - c_2\tan\alpha_{12})r_2 - (r_1\omega_3 - c_1\tan\alpha_{32})r_1\} \tag{16}$$

$$g_2 = \{(r_3\omega_2 - c_3\tan\alpha_{22})r_3 - (r_2\omega_1 - c_2\tan\alpha_{12})r_2\} \tag{17}$$

$$g_3 = \{(r_1\omega_3 - c_1\tan\alpha_{32})r_1 - (r_3\omega_2 - c_3\tan\alpha_{22})r_3\} \tag{18}$$

$\mathbf{R} = \mathbf{R}(\mathbf{f})$ is the matrix of loss, as follows:

$$\mathbf{R} = \mathbf{R}(\mathbf{f}) = [0 \quad 0 \quad 0 \quad P_L]^T \tag{19}$$

T is the matrix of torque applied by each impeller to fluid, as expressed below:

$$\mathbf{T} = [T_1 \quad T_2 \quad T_3 \quad 0]^T \tag{20}$$

From the assumption that cross sectional area of flow passage is constant, $c_1 = c_2 = c_3 = Q/A$ in all of the above equations.

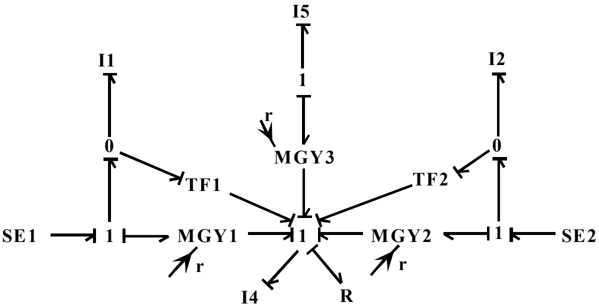


Figure 5: Bondgraphs for Converter Range

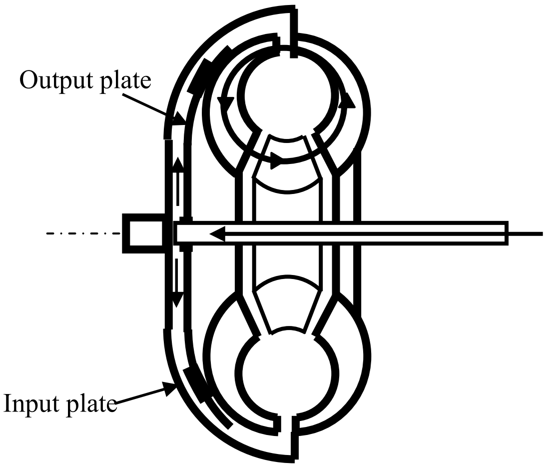


Figure 6: Torque Converter with a Lock-up Clutch

In the converter range, a stator is fixed and not rotating. However, since a stator is generally shaped as a blade, it produces a torque as the result of change in momentum, which is determined by the inlet and outlet blade angles. This torque is received by the transmission case. Fluid is then forced by the reaction torque to return along the blade surface to the pump inlet side. When these characteristics are represented by a bondgraphs, TF-element for a stator and the bond connecting to this element are omitted, and I3-element is combined with I-element for a transmission case to become I5-element, as shown in Fig. 5. Normally, when a transmission torque is transmitted, the corresponding torque reaction works on the casing, as shown in Fig. 5 (Suzuki and Tanaka 2003). As mentioned above, since torques work on the stator, stator-related bonds should be included in the bondgraphs to represent a T/C accurately.

3. LOCK-UP CLUTCH IN T/C

Many automobiles have the lock-up clutch (L/C) in a T/C shown in Fig. 6. The L/C is a wet-type friction clutch and consists of an input plate and an output plate. The input clutch is connected to the pump impeller and the output clutch is connected to the turbine runner. If the clutch is engaged, its revolution is the same as the pump and the turbine. Then a bond of the input plate is connected to 1-junction of the pump. Between input and output plate the oil flows. This flow is controlled by a regulator valve outside the T/C. If the regulator valve is closed, the clutch is working and it transmits the torque.

The entire input torque of T/C is transmitted only through this clutch to the output side not through the fluid. If the regulator valve is open, the clutch is disconnected and it does not transmit the torque. When the direct clutch loses its connection (this means that the difference between the revolution speed of the input and the output plate is small) that is regulated by the valve, transmission of input torque is separated from the clutch and the working fluid flows. Then the bondgraphs representation of the direct clutch controlled by the regulator valve is added as MR-element, shown in Fig.7 (Suzuki and Tanaka 2003).

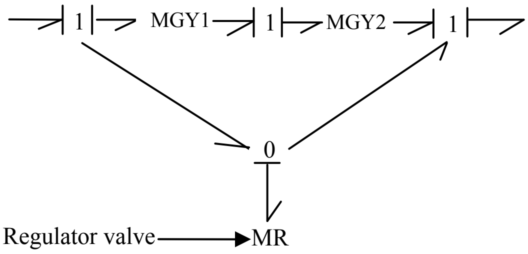


Figure 7: Lock-up Clutch Controlled by Regulator Valve

However, this model does not include the dynamic motion of the L/C. Bondgraphs sub-models representing the dynamics of the L/C are newly added instead of MR-element, as shown in Fig. 8. Figure 8 shows the effects on MR-element from the additional pump and motion effects of turbine and L/C. In Fig. 8, SE-element indicates the effect of oil pump pressure to control engagement of the L/C and turbine, TF1-element transformer from rotational system to the axial motion of system, and TF2-element transformer from the regulator valve system to the axial motion of system, , respectively. The sub-system between TF1- and TF2-element indicates the axial direction motion of the turbine and L/C. Another sub-system between SE- and TF2-element, lower rectangular part enclosed by the dotted line, indicates oil flow effects through the gap from the additional pump to the inlet of the turbine.

R-element, following SE-element, indicates the regulator valve resistance, the next R-element the leakage flow resistance through the gap while the L/C and turbine is moving. The motion of the turbine and L/C is affected by the pressure from the regulator valve and the axial direction spring equipped the turbine. The initial value of the axial gap is 1.3 mm. After the L/C begins to move and the gap length becomes less than 0.1 mm, the wet friction is activated between the input and output plate. These effects are shown by full arrows toward MR-element. CFD numerical analysis is used to calculate the values of each element.

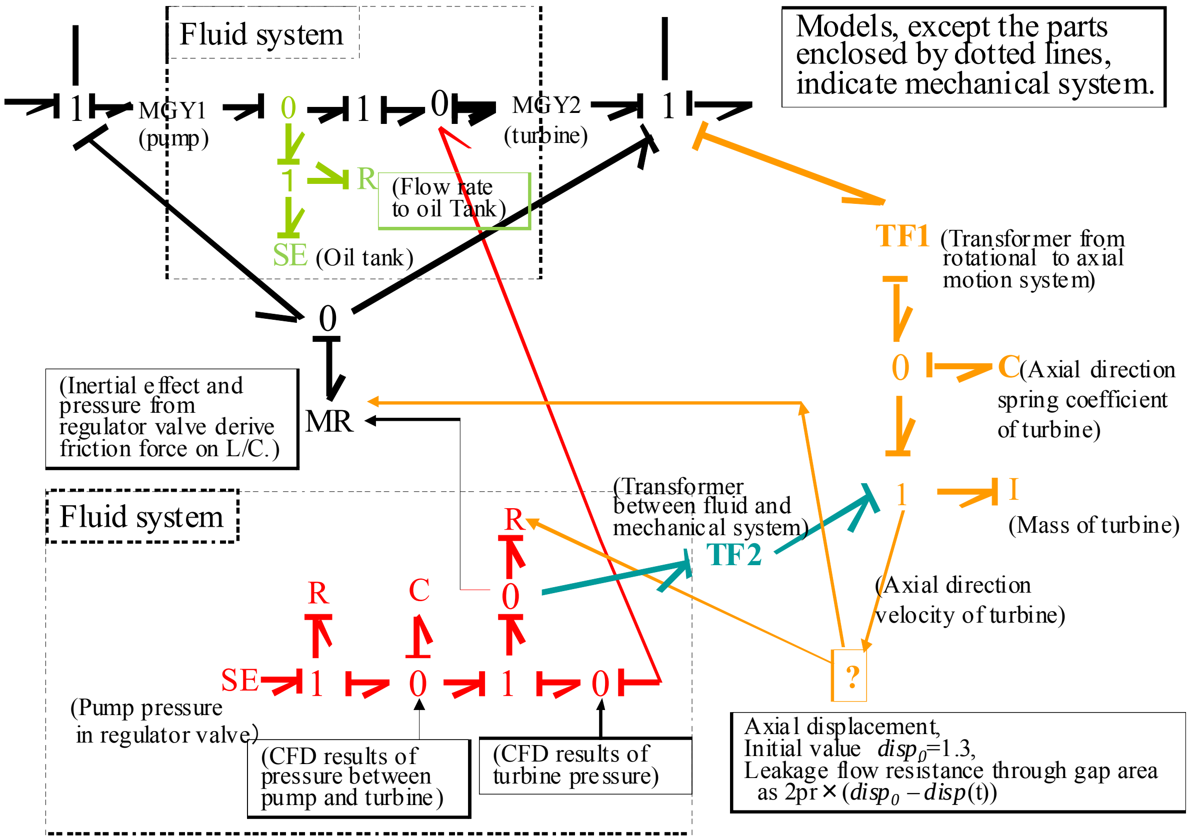


Figure 8: Bondgraphs model Including Dynamics of the Loch-up Clutch

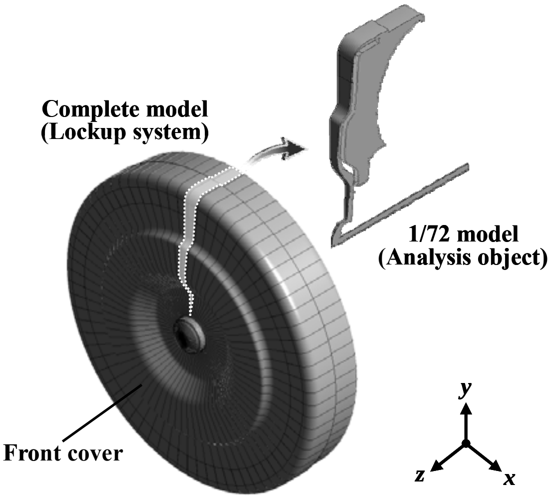


Figure 9: Analysis Object of the Lock-up System

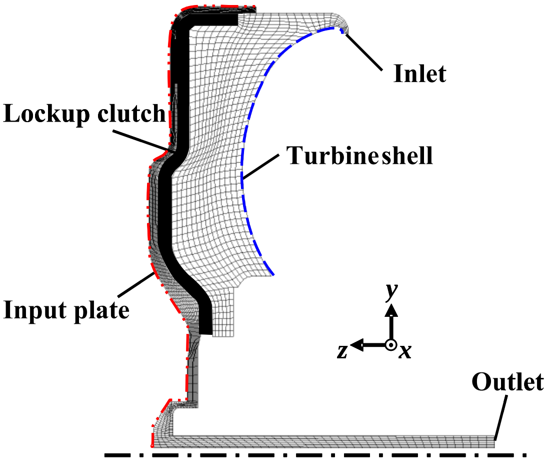


Figure 10: Computational Grids of the Lock-up Clutch

4. MODEL DESCRIPTION AND NUMERICAL ANALYSIS FOR L/C DYNAMICS

For the CFD analysis, the object of the lock-up system is shown in Fig. 9, and the CFD model of the L/C is shown in Fig. 10. Since this problem is axis-symmetric, only seventieth part of the L/C is modeled with periodical boundary conditions to reduce time spent on computation. Here, it is assumed that the flow field around the L/C is independent of the flow field inside the torus (i.e., pump, turbine, and stator blade area). Therefore, the torus was not included in this model. The speed ratio is defined as the turbine rotational speed divided by the pump rotational speed. The pump rotational speed is simulated by rotating the entire calculated domain, and the turbine rotational speed is simulated by rotating the walls of the model as defined by the L/C and the turbine shell.

Figure11 shows the boundary conditions for inlet and outlet pressure used for this simulation. At the beginning, the outlet pressure is higher than the inlet pressure. The resulting differential pressure prevents the L/C from moving. The inlet pressure then increases quadratically, and the outlet pressure decreases linearly. This variation in pressure is a typical lock-up control sequence used for actual automatic transmissions. The boundary conditions for this simulation are summarized in Table 1. Because the lock-up system begins to engage when the speed ratio, which indicates the ratio of turbine rotational speed to pump rotational speed and is represented by e, becomes more than 0.7 or 0.8, the simulations were performed in case of e ≧ 0.8.

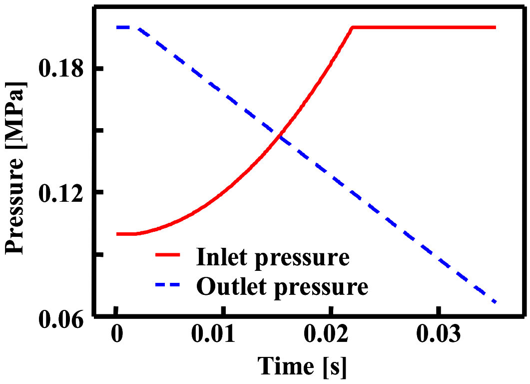


Figure 11: Inlet and Outlet Pressure Distribution

Table 1: Analysis Conditions for Transient State

Pump rotational speed	2000 rpm
Turbine rotational speed	1600, 1800, 2200, 2400 rpm
Speed ratio	0.8, 0.9, 1.1, 1.2
Inlet pressure	Solid line of Fig. 11
Outlet pressure	Dashed line of Fig. 11

ANSYS CFX Ver12.1 is used for the CFD solver. In this simulation, the lock-up motion is not known a priori and will be calculated based on the forces acting on the L/C according to Newton's Second Law. The motion equation is re-assembled and translated to CFX CEL (CFX Expression Language). The Navier–Stokes equations can then be solved and specified explicitly with this CEL. In reality, the friction paper on the L/C contacts the input plate to transfer torque. However, in the simulation, contact of the friction paper would violate the continuity equation and would either collapse or diminish the CFD mesh. In order to maintain fluid continuity and prevent the mesh from collapsing, a small gap is set between the L/C and the input plate. Since the friction paper is on the L/C, a minimum clearance of 0.1 mm is set between the friction paper and the input plate. The initial clearance is bounded at 1.3 mm, meaning the L/C travels from 1.3 mm to the 0.1 mm minimum gap.

Another new technique was introduced to keep the mesh quality constant. Instead of using a CAD rendering of the initial shape of the L/C to mesh the model, a CAD rendering of the final deflected shape at the minimum clearance of 0.1 mm (s = 0.1 mm) was meshed and then extended 1.3 mm (s = 1.3 mm) in the direction of movement to its initial position. This technique maintained good quality of the moving mesh and prevented divergence during the simulation. The number of numerical mesh is 11,000 elements, all of the hexahedral type.

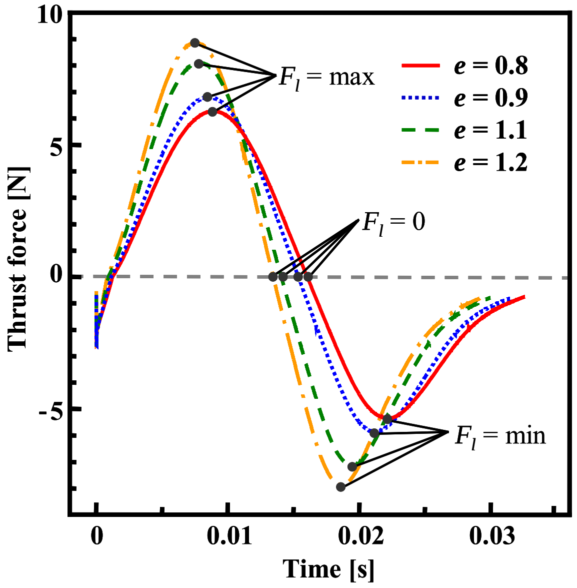


Figure 12: Thrust Force of the Lock-up Clutch

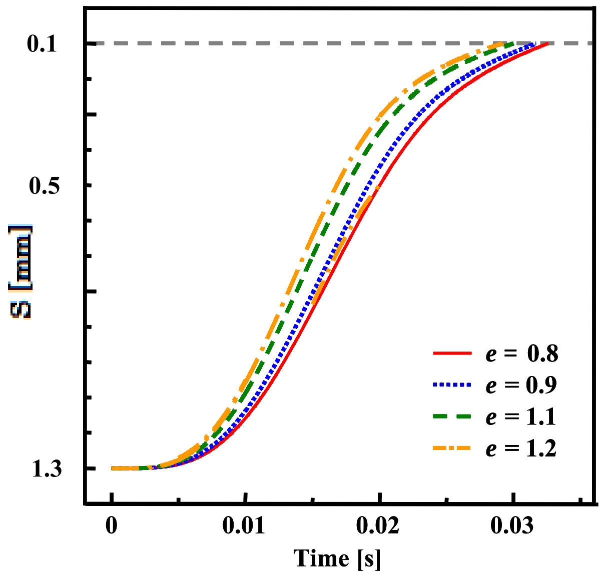


Figure 13: Lock-up Clutch Moving Distance

5. RESULTS AND DISCUSSIONS

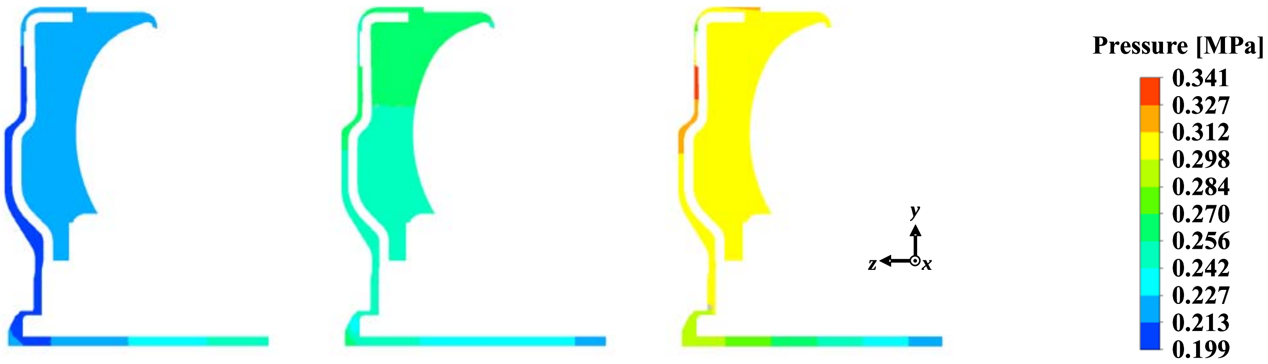
Figure 12 shows the thrust force acting on the L/C over time at various speed ratios. In this simulation, the pump rotational speed is fixed at 2000 rpm, and the turbine speed is varied to achieve the different speed ratios. Thus, the turbine speed is 1600 rpm for e = 0.8 and 2400 rpm for e = 1.2. Negative thrust force values act to separate the L/C from the input plate. Before the inlet pressure exceeds the outlet pressure, the thrust force acting on the clutch is negative, and the clutch holds steady. After engagement pressure is applied, the thrust force becomes positive and increases until it reaches a maximum inflection point. After passing this point, the thrust force gradually decreases to a minimum inflection point and then increases again. This trend is true at all speed ratios. The maximum inflection point, the minimum inflection point, and the point where the thrust force becomes zero are defined as F_l = max, F_l = min and F_l = 0, respectively.

Comparing the data in Fig. 12 at speed ratio e = 0.8 and 1.2, the maximum and minimum thrust forces have greater absolute values at the higher speed ratio than at the lower speed ratio. The shear stress induced in the working fluid by both the input plate wall and the L/C wall imparts momentum energy to the working fluid. Since the rotational speed of the input plate (pump rotational speed) is fixed, the difference in the momentum of the working fluid is due to the L/C speed (turbine speed).

Figure 13 shows the L/C movement from the initial condition (lock-up off position) with various speed ratios. The pump rotational speed is fixed at 2000 rpm. The L/C is in its initial position at s = 1.3 mm and is at its final engagement point at s = 0.1 mm. At each speed ratio, even though the inlet pressure continues to increase, the L/C does not begin to move until 0.005 sec. Once movement begins, the displacement grows quadratically until the L/C reaches its final engagement point. In addition, when the thrust force reaches minimum inflation point in Fig. 12, the inflation point also can be seen at the same time in Fig.13. Moreover, the higher speed ratios, the engagement time is reduced. At lower speed ratios, the slipping speed is higher, and more shear flow or more energy is created in the working fluid. Higher pressure is required to overcome the additional energy in the working fluid at lower speed ratios. However, since the inlet pressure is limited by the boundary conditions of the simulation (see Fig. 11), the engagement time required to overcome the additional energy at lower speed ratios is increased.

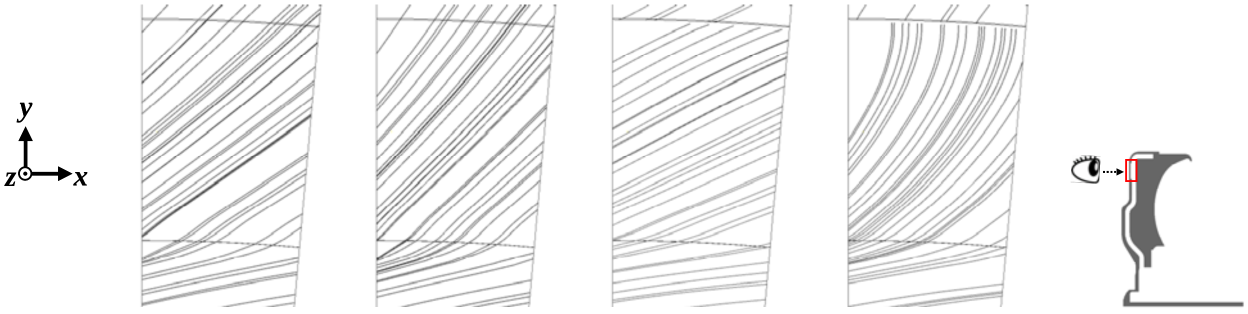
Figure 14 shows the pressure distribution in the meridian plane of the L/C for e=0.8 at three different values of thrust force: (a) F_l = max, (b) F_l = 0, and (c) F_l = min. In Fig. 14 (a), the outlet pressure is still higher than the inlet pressure due to the initial condition of the simulation. At the same time, the pressure increases in the radial direction due to centrifugal pressure. The pressure at the turbine shell is higher than the pressure at the input plate. This pressure difference, shown in Fig. 14 (a), creates a positive thrust force and acts to engage the L/C as previously discussed. Figure 14 (b) shows that the pressure distribution at the input plate and at the turbine shell is equivalent. Since there is no thrust force acting on the L/C, only inertia is driving the movement of the L/C. In Fig. 14 (c), the pressure distribution at the input plate becomes higher than at the turbine shell. The velocity of the L/C is reduced by the higher pressure at the input plate. The maximum and minimum pressure differences across the L/C coincide with the inflection points in Fig. 12.

Figure 15 shows the limiting streamlines on the input plate and the friction paper at speed ratio 0.8 and 1.2 with F_l = min condition. Over all, due to the inlet pressure driven radial flow and centrifugal force, the flow tends to move left-bottom side to right-top side. At e=0.8, the rotational velocity of the input plate is higher than that of the friction paper (the L/C). As a result, the limiting streamlines on the input plate tend to move



(a) F_l = max (b) F_l = 0 (c) F_l = min

Figure 14: Pressure Distributions in the Meridian Plane (e = 0.8)



e = 0.8 e = 1.2 e = 0.8 e = 1.2

(a) Input plate (b) Friction Paper

Figure 15: Streamline on the Input plate and the Friction Paper at F_l = min

more upward compared to the friction paper side. For the same reason, at e = 1.2, the limiting streamlines on the friction paper move more upward.

As mentioned above, the flow field around the L/C can be perfectly simulated. At this stage, it finally becomes possible to determine appropriately the value of every R-element shown in Fig. 8 which has relations with pressure loss in the flow passage around the L/C.

6. CONCLUSIONS

A proposed bondgraphs has been applied to represent the functions of a torque converter. This bondgraphs can accurately represent dynamics of the lock-up clutch and enables easy correspondence between the bondgraphs model for the torque converter and explanation of the physical phenomena occurring in the torque converter. By use of this bondgraphs it becomes possible to control the lock-up clutch in dynamic characteristics, especially in slip control of the torque converter.

The flow field around the lock-up clutch has been simulated with a commercial CFD code. It has been found that the motion of the lock-up clutch is not linear, and that a small gap between the output plate and the input plate has an enormous effect on the lock-up clutch engagement time and on the flow field. The effects of the speed ratio, the pressure distribution, the streamlines in the meridian plane and the streamlines on the input plate and the friction paper have been studied.

The flow structures in the meridian plane are generally the same at the time that the thrust force becomes maximum or minimum, even at different speed ratio.

As a result that the flow field around the lock-up clutch is simulated numerically, it finally becomes possible to determine appropriately the value of every R-element which has relations with pressure loss in the flow passage around the lock-up clutch. At this stage, the proposed new bondgraphs model enhances its reality.

REFERENCES

Abe, K., Kondoh, T., Fukumura, K., and Kojima, M., 1991. Three-Dimensional Simulation of the flow in a Torque Converter. *SAE Paper* No. 910800.

Abe, H., Maruyama, T., Hasegawa, K., and Kondo, N., 1999. Analyzing Responsiveness of lock-up Clutch by Applying Three Dimensional Computational Fluid Dynamics. *HONDA R&D Technical Review*, Vol. 11 No2. pp.67-72.

Breedveld, P.C., 1984. Decomposition of Multiport Elements in a Revised Multibondgraph Notation. *J. Franklin Inst*., Vol. 318: pp.253-273.

Brun, K., Flack, R. D., 1995. Laser Velocimeter Measurements in the Turbine of an Automotive Torque Converter: Part II-Unsteady Measurements. *ASME paper 95-GT-293.*

By, R. R., Lakshminarayana, B., 1995. Measurement and Analysis of Static Pressure Field in a Torque Converter Pump. *Trans. of ASME, J. Fluids Eng.*, Vol. 117: pp.109-115.

Cigarini, M., Jonnavithula, S., 1995. Fluid Flow in an Automotive Torque Converter: Comparison of Numerical Results with Measurements. *SAE Technical Paper* No. 950673.

Ejiri, E., 2006. Analysis of Flow in a Lock-up Clutch of an Automotive Torque Converter. *JSME International Journal*, Series B 49(1). pp.131-141.

Hiramatsu, T., Akagi, T., and Yoneda, H., 1985. Control Technology of Minimal Slip-type Torque Converter Clutch. *SAE Tech. Paper* No. 850460.

Hoshino, A., Ishihara, T., 1990. A Consideration on Performance Improvement of Hydraulic Torque Converters. *Trans. of ASME*, Vol. 100: pp.65-70.

Hrovat, D., Tobler, W. E., 1985. Bondgraph Modeling and Computer Simulation of Automotive Torque Converters. *J. of Franklin Inst.*, Vol. 319-1/2: pp.93-114.

Ishihara. T., Emori, R. I., 1966. Torque Converter as a Vibrator Damper and Its Transient Characteristics. *SAE Paper* No. 660368.

Kim B S., Ha, S. B., Lim, W. S., and Cha, S. W., 2008. Performance Estimation Model of a Torque Converter Part 1: Correlation between the Internal Flow Field and Energy Loss Coefficient. *International Journal of Automotive Technology* Vol. 9, No. 2: pp.141-148.

Kono, K., Itoh, H., Nakamura, S., Yoshizawa, K., and Osawa, M., 1995. Torque Converter Clutch Slip Control System. *SAE Tech. Paper* No. 950672.

Suzuki, K., Tanaka, K., 2003. Torque Converter with Lock-up Clutch by Bond Graphs. *Proceedings of ICBGM2003.* January 19-23, Orlando, Florida.

Tsangarides, M. C., Tobler, W. E., 1985. Dynamic Behavior of a Torque Converter with Centrifugal Bypass Clutch. *SAE Paper* No. 850461.

Watanabe, H., 1999. Flow Visualization and Measurement in the Stator of Torque Converter. *Toyota Central Labs R&D Review* Vol. 34 , No. 1. pp.57-66.

Yamaguchi, T., 2002. Automotive Torque Converter Performance Prediction Using a New CFD Approach. *Proceedings of The* 9^{th} *of International Symposium on Transport Phenomena and Dynamics of Rotating Machinery (ISROMAC-9),* February 10-14, Honolulu, Hawaii.

AUTHORS BIOGRAPHY

TAKESHI YAMAGUCHI,

He graduated from Tohoku University, and obtained M.S. in San Diego State University. After working at Tohoku University as an assistant professor, he became a chief engineer in Aisin AW and also a Ph.D. candidate now. He has wide interests in CFD analyses and development on an automatic transmission as well as the basic phenomena of fluid mechanics. He is also a good scuba diver.

KAZUHIRO TANAKA,

He graduated from the University of Tokyo, obtained the first job as an assistant professor in Sophia University, and moved back to the University of Tokyo. After then, he moved to Kyushu Institute of Technology and is a professor. He has interests in combining Bondgraphs modeling and simulation with CFD analyses on flow patterns inside various components and systems.

KATSUYA SUZUKI

He retired from Toyota Motor Corporation, where he was a project general manager of Drive Train Division. He obtained Ph.D. degree in Nagoya University. He is now a visiting professor in Kyushu Institute of Technology and enjoys a lively discussion of Bondgraphs modeling and simulation. He is a member of a chorus group singing the Ninth Symphony in Nagoya.

OPERATION STRATEGIES TO MINIMIZE METHANOL RECOVERY IN BATCH DISTILLATION OF HYDROALCOHOLIC MIXTURES

De Lucca V., F.(a), Munizaga-Miranda, R.(b), Gelmi, C.A.(c), Pérez-Correa, J.R.(d)

Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile,
Casilla 306, Santiago 22, Chile.

[(a)]fdelucca@puc.cl, [(b)]rhmuniza@uc.cl, [(c)]cgelmi@ing.puc.cl, [(d)]perez@ing.puc.cl

ABSTRACT

Methanol is a toxic compound produced during fermentation of fruit juices that normally contaminates a large variety of commercial distillates. This work focuses in applying dynamic simulation to explore operating conditions of a packed batch distillation column for reducing the concentration in the heart of methanol relative to ethanol, from a ternary mixture of water, ethanol and methanol. A non-equilibrium mass and energy balance model of the distillation process was derived and solved using the method of lines. The cooling rate in the partial condenser (time variable) and the head/heart cut time were the operating variables considered in this work. Simulations showed that, within the operating range studied, higher cooling rates and smaller cut times achieved lower methanol relative concentrations in the heart. Following a heuristic strategy, we reduced the relative concentration of methanol more than 15%, recovering 75% of the ethanol in the heart cut. Future work will consider an expanded operating range of the model, a more complete optimization problem and the application of global optimization techniques.

Keywords: spirits, toxic compounds, PDE, packed column

1. INTRODUCTION

The quality of a fruit wine distillate is defined by the concentration of a large number of minor compounds that give the product its delicate aroma. Nevertheless, often the spirit contains toxic compounds or off-flavors, which considerably reduce the quality of the product. Normally, fruit distillates are produced in batch distillation equipments, where three fractions or cuts are collected: head, body and tail. The body is the cut that gives rise to the commercial product.

Little attention has been given to modeling, simulation and optimal operation of packed batch columns (Mujtaba 2004). There are several reasons that may explain the apparent lack of interest/research in this area: (i) the models involve coupled partial differential and algebraic equations (PDAEDs) for the mass and heat transfer phenomena, which are hard to solve; (ii) lack of flexibility of most current packed column algorithms as they are typically process dependent with narrow operating regions; (iii) the PDAEDs systems are frequently very stiff due to wide ranges in relative volatilities or large differences in packing and reboiler holdups (Mujtaba 2004). Moreover, minoritary compounds represent a challenge in packed distillation columns modeling, because reliable and robust thermodynamic models are required.

This work focuses in exploring, through dynamic simulations, operating conditions that reduce the relative concentration of methanol to ethanol in the body of a wine batch distillation, without affecting the ethanol yield of the process. For this purpose, a dynamic non-equilibrium model for distilling a ternary mixture (water-ethanol-methanol) in a batch packed distillation column was developed.

2. METHODOLOGY

2.1. Modeling

We modeled a packed distillation column located in the Chemical and Bioprocess Engineering Department, at the Pontificia Universidad Católica de Chile. The stainless steel distillation apparatus consists of a stainless steel boiler (50 L), a packed column (0,3 m height and 8 cm i.d.) a partial and a total condenser, all thermally insulted. The boiler is equipped with two 2.4 kW electric heaters, the column is filled with glass (750 mL) and copper (500 mL), 5 mm Raschig rings, and the partial condenser consists of a copper coil (6 mm o.d.) with 140 cm^2 external surface mounted above the packing column. Most of the assumptions employed in this work were taken from a previous work (Carvallo et al. 2011). However, in our model we included the heating dynamics of the packing (Wang et al. 2003) to reproduce better the behavior of the distillation during the heart cut (first 4 to 20 min). In addition, we assumed that the liquid and the packing were in thermal equilibrium in the column. The main model equations are shown next.

2.1.1. Mass and Energy Balances in the Boiler

Mass and energy balances were defined assuming that the boiler is at thermodynamic equilibrium, resulting in the following set of ordinary differential equations:

Total mass balance:

$$\frac{dM_B}{dt} = L_1 - V_B \tag{1}$$

Ethanol mass balance:

$$\frac{dx_{1,B}}{dt} = \frac{1}{M_B} \cdot \left(L_1 \cdot \left(x_{1,1} - x_{1,B}\right) - V_B \cdot \left(y_{1,B} - x_{1,1}\right)\right) \tag{2}$$

Methanol mass balance:

$$\frac{dx_{2,B}}{dt} = \frac{1}{M_B} \cdot \left(L_1 \cdot \left(x_{2,1} - x_{2,B}\right) - V_B \cdot \left(y_{2,B} - y_{2,1}\right)\right) \tag{3}$$

Energy balance:

$$\frac{d\left(M_B \cdot h_B\right)}{dt} = L_1 \cdot h_1 - V_B \cdot H_B + Q_B - Q_p \tag{4}$$

2.1.2. Mass and Energy Balances in the Packing Column

Mass and energy balances were applied neglecting the dynamics of the vapor phase and assuming a liquid-vapor interface along the column height z as follows:

Total mass balance:

$$\frac{\partial M'}{\partial t} = \frac{\partial L}{\partial z} - \frac{\partial V}{\partial z} \tag{5}$$

Ethanol mass balance in the liquid phase:

$$\frac{\partial x_1}{\partial t} = \frac{1}{M'} \cdot \left(L \cdot \frac{\partial x_1}{\partial z} - V \cdot \frac{\partial y_1}{\partial z} + \left(x_1 - y_1\right) \cdot \frac{\partial V}{\partial z} \right) \tag{6}$$

Methanol mass balance in the liquid phase:

$$\frac{\partial x_2}{\partial t} = \frac{1}{M'} \cdot \left(L \cdot \frac{\partial x_2}{\partial z} - V \cdot \frac{\partial y_2}{\partial z} + \left(x_2 - y_2\right) \cdot \frac{\partial V}{\partial z} \right) \tag{7}$$

Ethanol mass balance in the vapor phase:

$$0 = -V \cdot \frac{\partial y_1}{\partial z} - y_1 \cdot \frac{\partial V}{\partial z} + S \cdot N'_1 \tag{8}$$

Methanol mass balance in the vapor phase:

$$0 = -V \cdot \frac{\partial y_2}{\partial z} - y_2 \cdot \frac{\partial V}{\partial z} + S \cdot N'_2 \tag{9}$$

Energy balance:

$$\frac{\partial T}{\partial t} = \left(M' \cdot \frac{\partial h}{\partial T} + m'_{pack} \cdot c_{p,pack} \right)^{-1} \cdot \ldots \\ \left(\begin{array}{l} L \cdot \frac{\partial h}{\partial z} - V \cdot \frac{\partial H}{\partial z} + \left(h - H\right) \cdot \frac{\partial V}{\partial z} - M' \cdot \frac{\partial h}{\partial x_1} \cdot \ldots \\ \left(L \cdot \frac{\partial x_1}{\partial z} - V \cdot \frac{\partial y_1}{\partial z} + \left(x_1 - y_1\right) \cdot \frac{\partial V}{\partial z} \right) \end{array} \right) \tag{10}$$

2.1.3. Mass and Energy Balances in the Partial Condenser

Mass and energy balances assumed that both phases were in thermodynamic equilibrium. Additionally, quasi-steady state was assumed, since its dynamics is much faster than the boiler and the packed column. The resulting equations are:

Total mass balance:

$$0 = V_N - L_C \cdot \left(1 + \frac{1}{R_D} \right) \tag{11}$$

Ethanol mass balance:

$$0 = V_N \cdot y_{1,N} - L_C \cdot \left(x_{1,C} - \frac{y_{1,D}}{R_D} \right) \tag{12}$$

Methanol mass balance:

$$0 = V_N \cdot y_{2,N} - L_C \cdot \left(x_{2,C} + \frac{y_{2,D}}{R_D} \right) \tag{13}$$

Energy balance:

$$0 = V_N \cdot H_N - L_C \cdot \left(h_C + \frac{H_D}{R_D} \right) - Q_C \tag{14}$$

The method of lines was applied to solve the PDE’s describing the packing column, considering 21 sections. Thus, each section was described by 4 ordinary differential equations; 3 mass balances plus the energy balance. In addition, the model of the boiler contains 4 ordinary differential equations and the model of the partial condenser contains 4 algebraic equations, since it is assumed pseudo-stationary. Finally, the gas phase in the packing column is assumed pseudo-stationary, contributing with two additional ordinary differential equations (mass balances of ethanol and methanol). The model also includes a set of constitutive algebraic equations describing the vapor/liquid equilibrium and the mass transfer. Thus, as a whole, the model contains 90 ordinary differential equations and 36 algebraic equations, which were solved in Simulink/Matlab®. This model was calibrated with data collected in a laboratory distillation column (Munizaga-Miranda 2011).

2.2. Exploring Operation Strategies

Due to the complexity of the problem, we used trial and error and intuition to first defined a feasibility operating range. This is a tricky task, because the bounds on one parameter depend on the value of the other operating parameters. The cooling rate in the partial condenser, QPC, and the head/heart cut time, t_{cut}, were the operating variables we explored. Since QPC is time variable and can take any form, to simplify we constrained the search to a specific time function containing few adjustable parameters. This function was used before in the optimization of a tray batch distillation operation (Osorio et al. 2005), since it generates smooth cooling rate trajectories that are easily attained in industrial stills with standard control systems. Hence, this exploratory study is reduced to find a limited number of operating parameters that achieve the lowest methanol relative concentrations. For the head stage, the operating parameters were the cooling rate at the end of the stage (Q_{head}) and the head/heart cut time (t_{cut}). On the other hand, for the heart stage, the operating parameters were the cooling rate (Q_{heart}) at a hypothetical distillation time (8 h) and the curvature α.

The distillation begins by heating the mixture in the boiler until the column reached total reflux in steady state. Then, the head stage starts by reducing the cooling rate from 633 W following a predefined function (eq. 15). The collection of the heart cut starts at t_{cut}, where the heating power in the boiler is reduced 30% and the cooling rate follows a trajectory defined by another predefined function (eq. 16). The heart cut finishes when 75% of the initial ethanol is recovered.

$$Q_{c\,head}(t) = Q_{head} - 11\left(Q_{head} - 633\right)\left(1 - 1.1^{\frac{t}{t_{cut}} - 1}\right) \qquad (15)$$

$$Q_b(t) = Q_{heart} - \frac{\alpha}{1-\alpha}\left(Q_{heart} - Q_{head}\right)\left(1 - \alpha^{\frac{t - t_{cut}}{8 - t_{cut}} - 1}\right) \qquad (16)$$

The following figure shows this operating strategy in both stages.

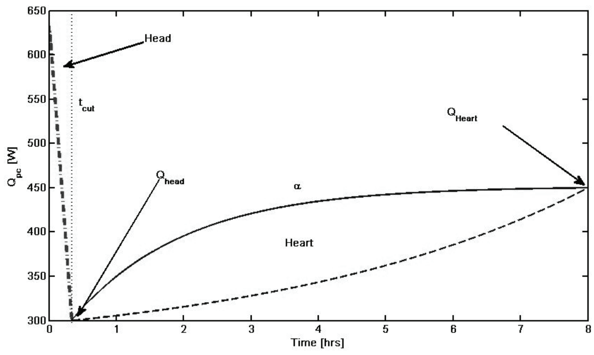


Figure 1: Operating strategy of the cooling rate in the partial condenser, during the head and heart Stages.

The cooling rate during the head stage (dot and dashes line on the left of Fig. 1) starts with total reflux (633 W, t = 0) and finishes at the head/heart cut time, when the collection of the heart cut begins. The cooling rate variation during the heart stage is defined by the values of parameters Q_{heart} and α in eq. 16. Q_{heart} corresponds to the hypothetical cooling rate that would be retired if the distillation lasted 8 h. However, the distillation finishes when 75% of the initial ethanol is recovered; hence, the distillation lasts less than 8 h. In addition, the curvature parameter, α, allows to define different trajectories, like for example the dotted line in Fig. 1.

Our aim is to reduce the concentration of methanol in the heart cut, in relation to the ethanol content in standard volume (20°C); i.e., relative to absolute alcohol (a.a.). This study considers 2 initial concentrations of methanol: low (2.3 g/L a.a.), which was considered before by Carvallo et al. (2011), and high (39.3 g/L a.a.). In each case, a hyper-grid of 4 operating parameters is assessed by simulations using 3 equidistant levels for Q_{head}, t_{cut} and Q_{heart}, and 3 logarithmically equidistant levels for α. Altogether, 162 simulations were performed. Initially, the average simulation time was 25 min each. However, after properly tuning the Max step size in the Simulink configuration parameters menu, the simulation time was reduced to 1 min 20 s. In addition, the ode15s stiff solver was used, which is a variable order solver based on the numerical differentiation formulas.

3. RESULTS

Considering several metrics, the model developed in this work outperformed the one developed by Carvallo et al. (2011). We believe that the better performance of our model can be explained by the inclusion of the packing heating dynamics. As an example, Fig. 2 shows the improvement in methanol recovery predictions compared to Carvallo's model.

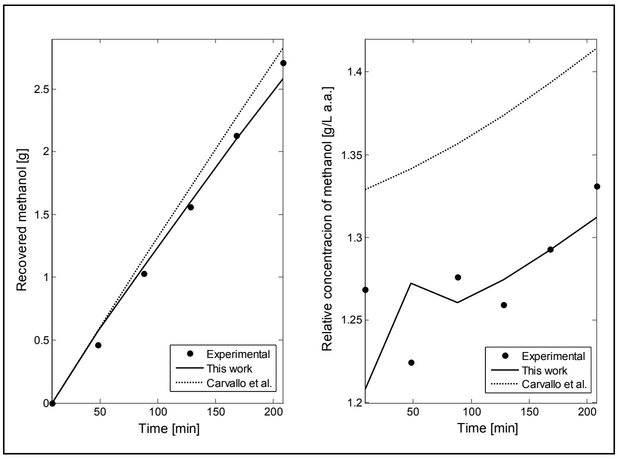


Figure 2: Evolution of total recovered methanol and relative methanol concentration in a given distillation.

It is important to highlight that the experimental points shown in Fig. 2 were used for model validation only, i.e., were not used for model calibration.

Our objective is to find by means of simulations, operating conditions for the batch distillation that reduce the relative concentration of methanol in the heart cut, for a fixed ethanol recovery of 75%. First, two feasible regions for the operating parameters of the distillation process were found. The first one considered the effect of Q_{head} and t_{cut}. Taking a single value of Q_{heart} and α, a range of Q_{head} and t_{cut} was assessed. This feasible operating region is:

$300 \text{ W} < Q_{head} < 550 \text{ W}$
$240 \text{ s} < t_{cut} < 1200 \text{ s}$
$Q_{heart} = 500\text{W}$
$\alpha = 0.01$

In turn, to assess the influence of Q_{heart} and α, Q_{head} and t_{cut} were fixed at 550 W and 1200 s, respectively. The resulting feasible region in this case is:

$Q_{head} = 550 \text{ W}$
$t_{cut} = 1200 \text{ s}$
$300 \text{ W} < Q_{heart} < 500 \text{ W}$
$0.01 < \alpha < 5$

For the low initial methanol case, with the best operating conditions found, it was possible to reduce 19.3% the concentration of methanol in the heart in relation to its concentration in the initial mixture. In turn, with the less effective operation only a 10.4% reduction was obtained. This difference in methanol recovery is in agreement with that observed in pear and kiwi distillates using traditional alembics and packing columns (García-Llobodanin et al., 2011; López-Vázquez et al., 2012). Still, the lowest methanol concentration achieved (1.85 g/L a.a.) for the low initial methanol case, is above the maximum allowed by the Chilean legislation (1.5 g/L a.a.).

For the high initial methanol case, with the best operating conditions found, it was possible to reduce 17.6% the methanol concentration in relation to the initial mixture. On the other hand, the less effective operation achieved a reduction of only 10.3%. Although, the minimum concentration achieved (32.4 g/L a.a.) is much higher than the legal limit in Chile.

Within the explored range, the most sensitive operating parameters on methanol concentration were the head/heart cut time (t_{cut}) and the cooling rate at the end of the head stage (Q_{head}); representing both a 70% of the total effect on the methanol concentration. In our study we considered a constant curvature in the cooling rate function in the head stage. Hence, given the high impact that this heat has in the methanol concentration, future works should consider a variable curvature.

As seen in Fig. 3, the relative concentration of methanol in the heart depends linearly on the cut time t_{cut}.

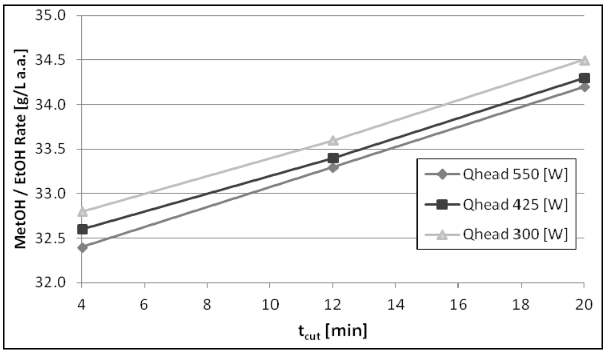


Figure 3: Effect of cut time in the relative concentration of methanol in the heart cut, for different values of Q_{head}, for Q_{heart} = 550 W and A = 0.01. High initial methanol concentration case.

It is evident from the figure above that the heart stage should be as short as possible, a result that is counter-intuitive. Moreover, it is better to operate with the highest possible rectification (high Q_{head}). Hence, both operating parameters are active constraint in our search problem. In addition, it was not observed a relevant interaction between Q_{head} and t_{cut} in the concentration of methanol in the heart.

Figure 4 shows the effect of Q_{heart} in the relative methanol concentration.

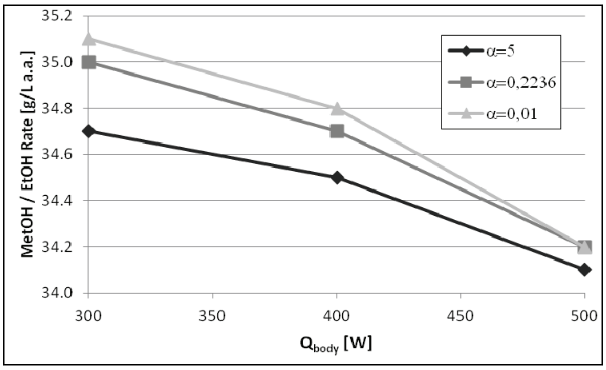


Figure 4: Effect of Q_{heart} in the relative concentration of methanol in the heart cut, for different curvatures A; Q_{head} = 550 W and t_{cut} = 20 min. High initial methanol concentration case.

Even though methanol concentrations varied in a nonlinear form with Q_{heart}, the best condition is also in the bound of the feasible range. Applying higher values of Q_{heart} (meaning high rectification) produce distillates close to the azeotrope or extremely low distillate flowrates, causing simulation problems.

In order to expand the feasible range of the model, operating parameters should be redefined or the cooling rate should be constrained during simulations, to avoid approaching the azeotrope in the distillate and ensure a minimum distillate flowrate.

In addition, Fig. 4 shows that while Q_{heart} increases the influence of the curvature parameter (α) in the methanol concentration falls. High values of α tend to increase the high cooling period during the heart stage, yielding a highly rectified distillate. Hence, by keeping the rectification high for long periods during distillation, we can reduce the relative concentration of methanol in the heart.

Figure 5 shows the impact of the main operating parameters on the distillation length.

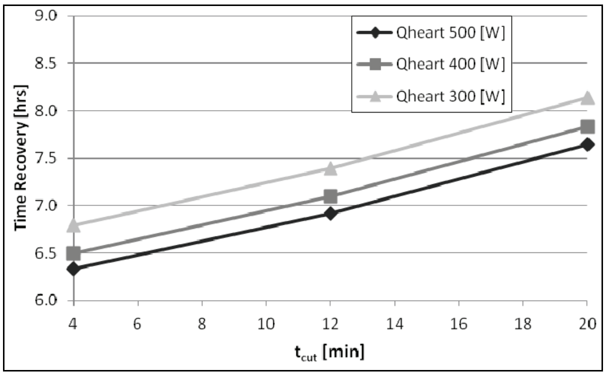


Figure 5: Effect of t_{cut} and Q_{heart} in the distillation length. High initial methanol concentration case.

According to the figure the distillation can last between 8.1 and 6.3 h, which correspond to the extreme values in the studied range of parameters t_{cut} and Q_{heart}. The former is the one that influences more the distillation length, accounting near 70% of the variation. This has important practical consequences, since the minimum value of t_{cut} reduces simultaneously the distillation length and the relative concentration of methanol in the heart. In turn, the other operating parameters consider in this study, either reduce the distillation length or reduce the methanol concentration. It is worthwhile mentioning, that our simulations indicate that the curvature parameter does not affect the duration of the distillation for values of Q_{heart} above 300 W.

4. CONCLUSIONS AND FUTURE WORK

Within the studied range of operating parameters, the defined optimization problem is reduced to evaluate the bounds of the feasible operating region. The difficulty is to find the feasible range, since the bounds on one parameter depend on the value of the other operating parameters. Using trial and error and intuition, we found the following set of operating parameters that achieved minimum relative methanol concentration in the heart cut: t_{cut} = 4 min, Q_{head} = 550 W, Q_{heart} = 500 W and α = 0,22. This set of values is independent of the methanol concentration in the initial mixture. Therefore, to find even better operating conditions, future work should focus in extending the feasible range of operation of the distillation simulation. Hence, a more robust model should be developed and better numerical algorithms should be applied. In addition, operating parameters can be redefined or cooling rate constraints can be included in the simulations, to avoid reaching the azeotrope in the distillate and ensure a minimum flowrate of distillate.

Furthermore, if a more difficult optimization problem arises, where the optimum operating parameters are not in the bounds of the feasible region, a global optimization code will be used to find the optimum.

Finally, since the legal limit for methanol was surpassed in both cases analyzed, it may be good option to change the objective function. For example, maximize the ethanol recovery in the heart, subjected to a maximum relative methanol concentration.

ACKNOWLEDGMENTS

This work was supported by project Fondecyt 1100357 from CONICYT.

REFERENCES

Carvallo, J., Labbe, M., Pérez-Correa, J.R., Zaror, C., Wisniak, J., 2011. Modeling methanol recovery in wine distillation stills with packing columns. *Food Control*, 22 (8), 1322–1332.

García-Llobodanina, L., Roca, J., López, J.R., Pérez-Correa, J.R., López, F., 2011. The lack of reproducibility of different distillation techniques and its impact on pear spirit composition. *International Journal of Food Science and Technology*, 46, 1956–1963.

Mujtaba, I.M, 2004. Batch distillation design and operation. Series on Chemical Engineering Vol. 3. London: Imperial College Press.

Munizaga-Miranda, R., 2011. *Modelación dinámica fenomenológica y simulación: destilación batch de una mezcla etanol-metanol-agua*. Undergradute thesis. Pontificia Universidad Católica de Chile. (In Spanish)

López-Vázquez, C., García-Llobodanin, L., Pérez-Correa, J.R., López, F., Blanco, P., Orriols, I., 2012. Aromatic Characterization of Pot Distilled Kiwi Spirits, *Journal of Agricultural and Food Chemistry,* 60, 2242 – 2247.

Osorio, D., Pérez-Correa, J.R., Biegler, L.T., Agosin, E., 2005. Wine distillates: practical operating recipe formulation for stills. *Journal of Agricultural and Food Chemistry*, 53 (16), 6326–6331.

Wang, L., Li, P., Wozny, G., Wang, S., 2003. A startup model for the simulation of batch distillation starting from a cold state. *Computers & Chemical Engineering*, 27 (10), 1485-1497.

NOMENCLATURE

c_p	Heat capacity, J/g K
c_t	Concentration, mol/mL
H	Enthalpy vapor phase, J/mol
h	Enthalpy liquid phase, J/mol
h_L	Liquid hold-up, m^3/m^{-3}
L	Molar flow in liquid phase, mol/s
M	Moles of retained liquid, mol
M'	Moles of retained liquid per meter of packing, mol/m
m'	Mass per meter of packing, g
Q	Heat, W
R_D	Reflux ratio
S	Cross section of the column, m^2
T	Temperature, K
t	Time, s
V	Molar flow in gas phase, mol/s
x	Molar fraction in liquid phase
y	Molar fraction in gas phase
z	Column height, m

AUTHORS BIOGRAPHY

Mr. De Lucca is a Chemical Engineer and postgraduate student at the Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile. His main interests are design, operation and control of distillation processes.

Mr. Munizaga-Miranda is a Chemical Engineer. He is currently working at CCU, a beverage company in Chile. His main interests are process modeling and optimization.

Dr. Gelmi is Assistant Professor at the Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile. His interest includes modeling and simulation of complex systems, numerical methods and optimization.

Dr. Pérez-Correa is Professor at the Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile. His interest includes modeling, optimization and control of chemical processes and bioprocesses, like distillation, fermentation and extraction of natural products.

CFD MODELING OF THERMAL DISTRIBUTION IN INDUSTRIAL SERVER CENTERS FOR CONFIGURATION OPTIMISATION AND CONTROL

Pierre-Luc Paradis[a,c], Drishtysingh Ramdenee[a], Adrian Ilinca[a], Hussein Ibrahim[a], Abderrazak El-Ouafi [a], Jean-Sébastien Deschênes[a], Alexandre Boudreau [a] Daniel Rousse [b]

[a] Université du Québec à Rimouski, Canada, G5L 3A1
[b] École de technologie supérieure, Canada, H3C 1K3

[c]pierre_lucparadis@hotmail.com

ABSTRACT

Industries and institutions are more and more prone to be equipped with servers to control all required computational and communication tasks. In many cases where the computational requirements of tasks or where data management are enormous, the need for proper configuration of the server cabinets is important to optimize cooling mechanisms and to improve energy efficiency. Specialized software are presently available to model new server centers with optimized configuration. However, there are no clear guidelines in the literature on the best way to build datacenter and on how you could increase efficiency of existing server center. In this paper, we present a Computational Fluid Dynamics (CFD) model made use to evaluate different solutions of cooling in data center and to write guidelines to help owners to improve efficiency of their buildings. Then a look is taken to control strategy that could be develop with those simulations.

Keywords: Flow Simulation, SolidWorks, Data center, Cooling optimization.

1. INTRODUCTION

Telecommunication services mainly those involving IP internet protocol have known enormous growth over the last few years and, thus, servers with enormous computational capacity have become very common. With them, challenges with regard to optimized cooling configuration, have surfaced. These mainly target already existing server centers as newly built ones are designed with specialized software using the cooling criteria as an intrinsic design parameter. Methods involving positioning of the server cabinets according to hot/cold corridor logic and their confinements at different levels exist and can lead to significant energy savings. However, in many cases, the non-uniformity of air inlet and outlet positioning and sizes bring additional challenges. Hence, in an attempt to advice upon proper ventilation to minimize air conditioning cost, a numerical model based on SolidWorks Flow Simulation CFD computational engine has been developed to characterize different situations and scenarios and then provide conclusions about the cooling mechanism in a server center. Furthermore, the model has been made generic as possible to be applicable to a large number of data centers. This paper, first, presents a literature review of this field and then focuses on the characteristics of the model. Following the model presentation, the article then ponders on the validation of the numerical model on a test bench case followed by results and conclusions obtained for an industrial server center case study. The article concludes by providing recommendations for optimization of the server arrangement and cooling mechanisms as well as an insight of future development in predictive control. A 2D lumped model is built for a simplified server centers and a state model is built in a temperature predictive algorithm control.

2. LITERATURE REVIEW AND CHOICE OF SOFTWARE

In order to better understand the intrinsic engineering analysis that are required into thermal modeling of server centers, the intrinsic mechanism of server centers and related cooling systems and to analyse improvement possibilities with regard to existing tools, a literature review has been performed. One of the essential conclusions of our research has been that simulation tools should be used for positioning optimisation (of server cabinets and cooling systems) of to be constructed data centers. In other words, the tools inscribe themselves as an architectural tool into the development of new datacenter. However, we need to develop a tool to provide guidelines in improving energy efficiency of already existing server centers where different constraints need to be respected. When it comes to modeling of distinct isolated thermal energy generating electronic devices, there have been recent developments in the availability of thermal modeling tools for architectural studies in the academic/research community. One such tool is HotSpot (K. Skadron and al, 2003) for microprocessors, which models temperature using thermal resistances and capacitances derived from the layout of micro-architectural structures that has been validated using finite element simulations. Rather than detailed thermal simulators for processors, quick estimation using convective energy dissipation techniques are used after calculating a processor's energy consumption using event counters in (F. Bellosa and al, 2003; A. Weissel and F. Bellosa, 2004). Such estimation has been used for developing temperature aware scheduling (A. Weissel and F. Bellosa, 2004). There have also been thermal modeling studies for individual disks (S. Gurumurthi, A. Sivasubramaniam, and V. Natarajan, 2005) and disk arrays (R. Huang and D. Chung, 2002), with the former providing a tool which also integrates with a disk performance simulator for architectural studies. These tools, which allow integrated performance and power/thermal studies, have been facilitating research contributions (K. Skadron and al, 2003; D. Brooks and M. Martonosi, 2001; M. Gomaa, M. D. Powel and T. N. Vijaykumar, 2004; L. Shang and al, 2004; S. Gurumurthi and al, 2003; Y. Kim and al, 2006; E. Pinheiro and R. Bianchini, 2004) in the architecture community for reducing power/temperature. All these tools are useful when studying and optimizing individual components. However, in this paper, we are interested in studying complete data center systems, where there could be interactions between different components. A recent tool (T. Heath and al, 2006) proposes using simple equations to calculate temperatures at very specific points in the server center system. While this approach suffices for certain simple "what-if" questions as suggested in (T. Heath and al, 2006), a CFD based model is needed for a more holistic examination of the system under a wider spectrum of static (e.g.: where to place components, computer room air conditioning (CRAC) units) and dynamic (e.g.: how long before the temperature reaches a threshold upon fan failure? what thermal management technique provides the best recourse upon emergency?). Fluid flows need to be modeled accurately for figuring out where components need to be placed and understanding complete system interactions. Thermal modeling of data centers: The importance of cooling high density datacenters/machine-rooms has attracted considerable interest recently (N. Rasmussen; N. Rasmussen; R. K. Sharma, C. E. Bash and C. D. Patel, 2002; C. E. Bash, C. D. Patel and P. K. Sharma, 2003; P. Rodgers and V. Eveloy, 2003; J. Moore, J. Chase and P. Ranganathan, 2006). Most of these studies (J. F. Karlsson and B. Moshfegh, 2005; C. D. Patel and al, 2002; C. D. Patel and al, 2001; C. D. Patel and al, 2003; M. H. Beitelmal and C. D. Patel, 2004; C. D. Patel and A. J. Shah, 2005; S. V. Patankar and K. C. Karki, 2004) have looked at this problem from an engineering perspective of designing CRAC and other cooling systems, placement of racks in machine rooms, etc., with many of them using CFD models. For instance, (R. Sharma and al, 2005) points out that heat recirculation is a limiting factor in existing cooling systems and proposes using fan/coil unit in the ceiling above the rack. Impact of CRAC failures on static provisioning has also been studied using CFD models (C. D. Patel and A. J. Shah, 2005). From the computer science/systems perspective, researchers are starting to use CFD models for workload placement (J. Moore and al, 2002; J. Moore and al, 2005; J. Mooreand al, 2006) across racks of a machine room, and balancing the temperature across these racks (R. Sharma and al, 2005). The different studies supported our choice for a CFD code. The most widely used code in such cases is « FloVENT simulation software » as it comprises of a specialised module in server air conditioning modeling. However, this license is not available at Univesity of Quebec at Rimouski (UQAR) and the choice was made among Comsol multiphysics, Ansys CFX, Ansys Fluent and SolidWorks Flow Simulation. For simplicity and functionality reasons SolidWorks would have been the software of choice for geometry modeling. Furthermore, the aim of the project was to define a tool that allows trend characterisation to advice on cooling, heat sources and heat sinks positioning rather than high accuracy point to point temperature definition. For these reasons, SolidWorks Flow Simulation was chosen.

3. CFD MODEL

The aim of this work is to build a numerical model of a server center such we can evaluate the impact of modifying the air flow, the addition of confinements, etc. on the temperature distribution. The model was built on SolidWorks commercial software and respects the actual geometrical parameters of the real server center. The boundary conditions for a given simulation were also defined according to measured data in the real physical center. The computational domain was bounded by the walls of the actual data centre. The CFD model was based on previous calibration works that evaluated the relative performance of turbulence models, transition models, mesh type and size and the size of the computational domain. The size of our domain in our case is the size of the room. The SolidWorks automatic mesh generator was used. The mesh was successively refined from a completely coarse one and a convergence study was performed. The mesh size relating to relative error of less than 3% between two consecutive simulations was taken. This consisted of 250 000 tetrahedrons. The used turbulence model was a k-episilon one with fully turbulent transition.

4. CFD VALIDATION

The SolidWorks based numerical model's capacities need to be tested and validated in order to be able to assume advices made to the industry viable. For such, the main server room of UQAR was used as a test bench. The geometric model of the UQAR server room was modeled on SolidWorks. For a specific working regime (assumed stationary as the variability is quite small over a short time period), the power output of the different server center components, the inlets and outlets air temperature and flow (pressure velocity and area of exit) as well as the power rating of other different heat sources were measured and entered as boundary conditions in the validation model. The computational domain was the data center room in real size as the simulation was a bounded one and the flow simulation was run as steady. The model mesh size and turbulence models were all configured according to the previously described methodology. Figure 1 illustrates the temperature distribution obtained for a specific situation.

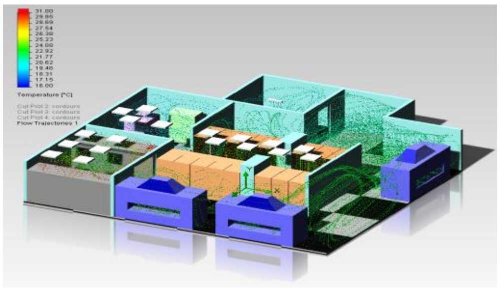


Figure 1: Results with Validation model

In this figure, the grey square represents the air inlet whereas the orange blocks represent the racks. The blue modules are the CRAC units. To validate the quality of the results, the data center was run at the same "regime" and the temperature in a corridor between the racks was measured on a central plane at middle altitude with respect to the two calculation columns lines. Table 1 illustrates the obtained simulated results compared to experimentally measured ones. It is to be noted that Windmate 200 instrument was used to measure both the volumetric flow rate and the air temperature. A windmate 200 is an instrument usually used in sea navigation to measure wind speed, wind direction, temperature, chill and other navigation pertinent parameters.

Table 1: Single Line Table Caption

x [m]	y [m]	z [m]	Fluid Temperature [°C]	Measured [°C]	Error [°C]
3.20	1.49	0.41	24.77	26.1	1.33
2.55	1.49	0.41	24.36	26.0	1.64
1.89	1.49	0.41	23.83	25.6	1.77
0.70	1.49	0.41	23.60	24.6	1.00
0.05	1.49	0.41	23.42	23.2	-0.22
-0.59	1.49	0.41	22.88	21.4	-1.48
-1.23	1.49	0.41	21.58	20.4	-1.18
-1.88	1.49	0.41	21.95	18.8	-3.15
		Mean	23.30	23.26	-0.23
		Standard deviation [%]	1.111	2.79	11.98

We note that the simulated and experimental trend of the temperature variation on the same crucial geometric reference was very similar. We note a mean error of -0.23198 ˚C and a standard deviation of the two sets of data of 12 %. These can be explained and justified as follows: The transport equation of any parameter, φ, in the computational code is defined as:

$$\frac{\partial \rho\varphi}{\partial t} + div(\rho\varphi\vec{u}) = div(\Gamma grad\varphi) + q_{\varphi} \qquad (1)$$

This equation defines only heat transport via source, convection and diffusive terms. These terms were neglected in our model as the convective factor was assumed much larger than the other thermal transport terms. Furthermore, analysis of simulated results allows us to believe that the flow was intermittently laminar and turbulent (in both space and time). Due to model flow limitations, the simulation was performed as purely turbulent. The k-epsilon model was used. This model works well in thin shear layers where the changes in the flow direction are always so slow that the turbulence can adjust itself to local conditions. In flows where convection and diffusion cause significant differences between production and destruction of turbulence, as it is here the case, a compact algebraic prescription for the mixing length is no longer feasible and the k-epsilon model is a justified choice. However, further analysis of the characterizing equations of the model lead us to believe that a K-omega Shear Stress Transport turbulence model (k- ω SST) would have been more appropriate. The study of the k-epsilon and the k- ω SST models in (D. Ramdenee and A. Ilinca, 2011) lead us to notice that the latter model includes terms that account for stress shear transport along boundaries defined as $\left[2\rho S_{ij}S_{ij} - \frac{2}{3}\rho\omega\frac{\partial U_i}{\partial x_j}\delta_{ij}\right] - \beta^{*}\rho\omega^{2}$. In the real server center, such phenomena do occur. Therefore, the modeling would have been more precise with the k- ω SST, which is, unfortunately not available in SolidWorks Flow Simulation. It is, also, interesting to note that the flow is intermittently laminar and turbulent, such that the use of a transition model including an intermittency factor would have improved precision. Apart from the modeling aspect, it is pertinent to note that the precision of the Windmate 200 is 1 ˚C for temperature and 3 % variability for volumetric flowrate. Finally, human presence during measurement is, also, considered to influence experimental results. Therefore, we can consider our model very accurate whilst recognizing improvement opportunities in future versions

5. INDUSTRIAL SIMULATION

The actual industrial data center that was simulated to recommend on thermal optimisation contained 36 cabinets and 4 cooling units. Figure 2 represents the configuration with respective power output and air flow at different data centre cabinets and cooling units.

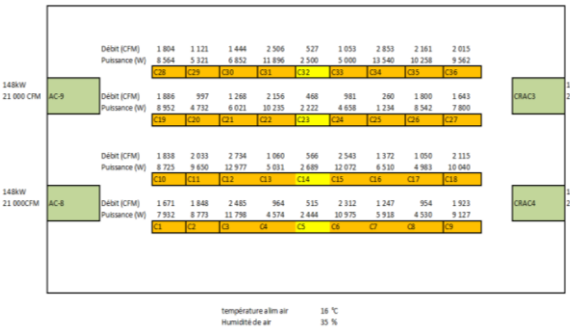


Figure 2: Real industrial model configuration

Following a meeting with the client, the different possible modifications that could be applied to the server center were established and the mandate was to evaluate the impact of the different modifications to advise upon the return rate of investment and energetic consumption impact. 7 different scenarios were identified. The scenarios were a combination of confinement possibilities of the cabinets and the air volumetric flow rate geometric positioning. In the complete study, all the different cases were exhaustively studied and recommendations were based on the analysis of around 6 different simulations. With reference to figure 3, 4 and 5 where the geometries of the different scenarios are illustrated, in this article, we will emphasize on the 1) effect of complete confinement, 2) effect of cooling airflow from the ceiling (down flow) or the floor (up flow), 3) effect of removing all confinements and 4) the effect of only confining the cabinets top with applying down flow cooling effect (from the ceiling). The choice of these results have been based on the need to define scenario(s) that amplify temperature fall and standardization, effect of standard cooling flow direction, recirculation and vortex analysis and the need to characterize the most energy efficient scenario.

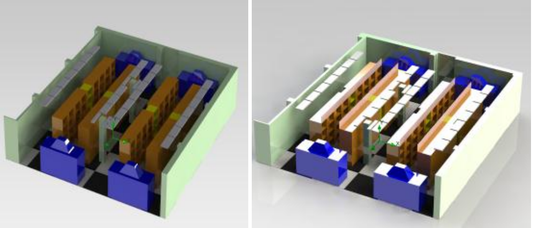


Figure 3: Scenarios representing cool air flow from raised floor and exit in ceiling without confinement and confinement over cabinet only

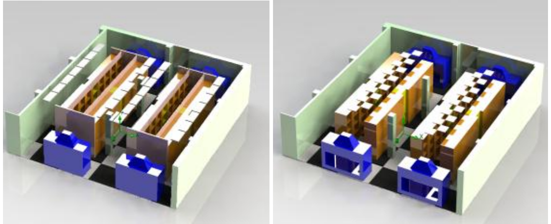


Figure 4: Scenarios representing cool air flow from raised floor and exit in ceiling with full confinement and cool air down flow with exit through air conditioners without any confinement

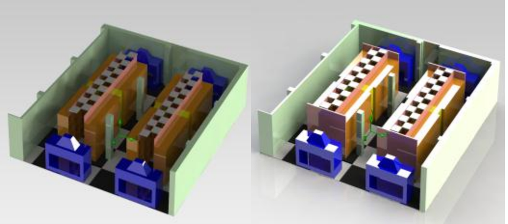


Figure 5: Scenarios representing cool air down flow with exit through air conditioners with confinement only above the cabinets and full confinement respectively

6. RESULTS

6.1. Effect of total confinement

When the confinement is complete, the mean temperature in the hot corridor tends to become uniform and falls by several degrees. We notice from figures 6, 7, 8 and 9 that the temperature range gets reduced from a maximum of 31^{0}C to 27^{0}C. This method seemingly better controls the hot air exits and enable more effective channeling of the cool air. The confinement is in fact a vertical lid that connects and bounds any cabinet or wall with the ceiling.

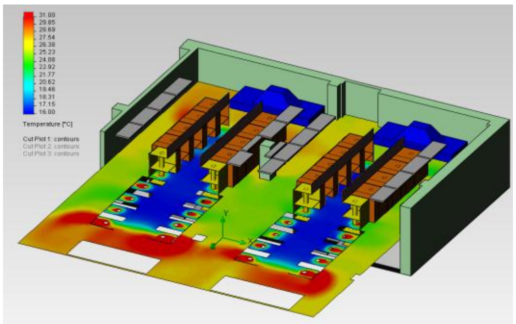


Figure 6: Horizontal temperature distribution for second scenario illustrated in figure 3

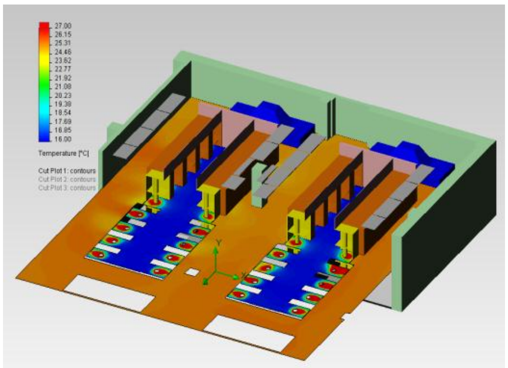


Figure 7: Horizontal temperature distribution for first scenario in figure 4

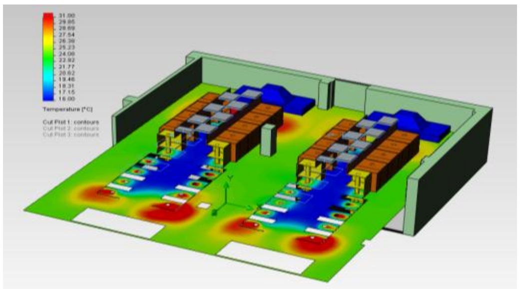


Figure 8: Horizontal temperature distribution for first scenario in figure 5

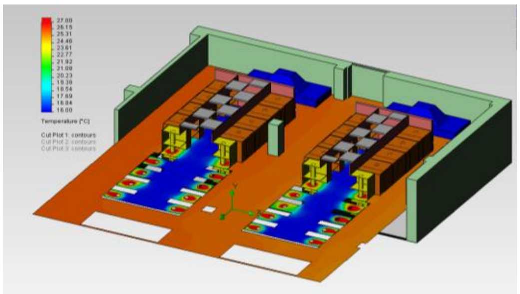


Figure 9: Horizontal temperature distribution for second scenario in figure 5

Figures 6 to 9 illustrate temperature distribution results on the same horizontal plane for diverse cool air flow direction for totally confined and semi confined geometries. It is clearly noticeable that for figures 7 and 9 where the simulations are for different air flow but complete confinement, the temperature is very uniform. Furthermore, analysis of the temperature color range allows seeing rapid temperature drop.

6.2. Effect of up flow or down flow cooling

This section ponders on the impact of the flow being from raised floor to ceiling or vice versa or into air conditioners directly. Being given that the previous section showed marked advantage of total confinement, we will evaluate the importance of cooling scheme with total confinement only. Figures 7, 9, 10 and 11 illustrate temperature uniformity in the cold corridor which is cooled by airflow at 16 °C. We can, thus, conclude on the fact that when the cold corridor is completely confined, there is no significant difference in the temperature distribution by up flow or down flow cooling.

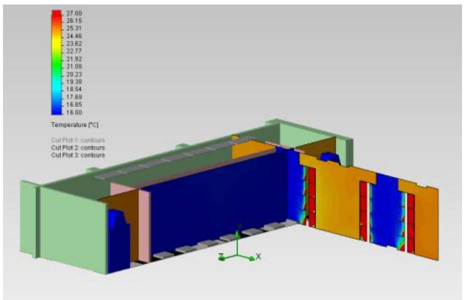


Figure 10: Vertical temperature distribution cool air down flow with exit through air conditioners without any confinement

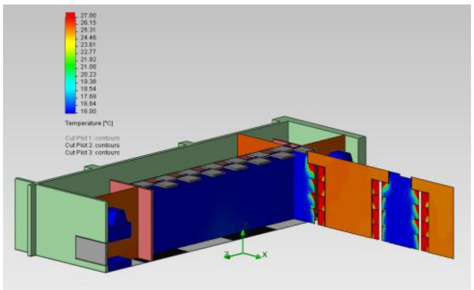


Figure 11: Vertical temperature distribution for second scenario in figure 5

Analysis of figure 7, 9, 10 and 11 allow us to see appreciable homogeneity and comparable low profile temperature range. These two advantages have been already associated to total confinement modification. It was, thus, concluded, that the two cooling sequences (simulated in these 2 cases) do not have any major consequence on the results.

6.3. Effect of removing all confinements

Previous simulations allowed us to see that confinement appreciably improve homogeneity and lower average temperature distribution. Simulation without any confinements has been performed to evaluate the presence of recirculation zones and vortex shedding that reduces homogeneity and adds to zones of "thermal uncertainty". Such is visible in figures 6 and 8 whereby circular high temperature zones are visible. These have been explained by hot air recirculation which is noticeable in figures 12 and 13.

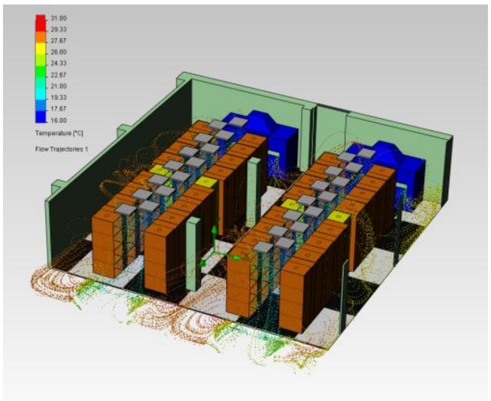


Figure 12: Air flow simulation for second scenario 1 in figure 3

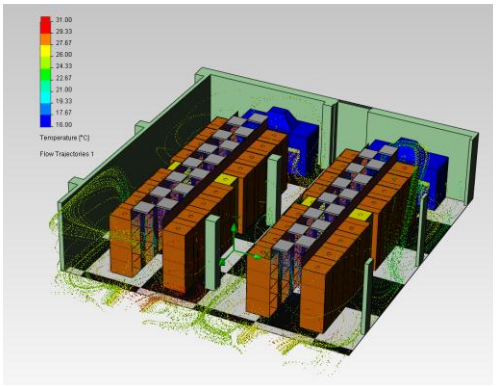


Figure 13: Air flow simulation for second scenario 2 in figure 3

Close analysis of figure 12 allows us to see more high temperature zones due to hot air recirculation as compared to figure 13. This can be explained by the absence of confinements in simulation of figure 12. The confinements in simulation of figure 13 are only over the cabinets.

6.4. High efficiency gain

It has been noticed through several previous simulations that major advantages are achieved with adding confinements above the cabinets. Successive simulations with different flow direction and mechanisms with these confinements have leaded us to advice on the significant energetic gain of down flow cooling from ceiling with confinements above the cabinets. This can be illustrated by comparison of figure 8 with figure 14, and comparison of figure 15 with figure 16.

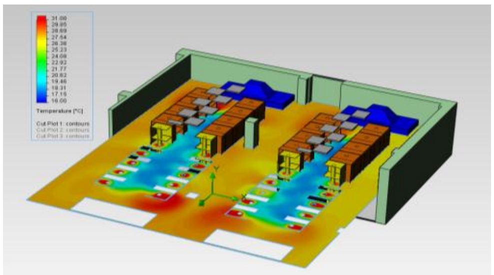


Figure 14: Horizontal temperature distribution for second scenario in figure 4

Comparison of figure 8 and 14 let us notice, that on a horizontal plane for ceiling to air conditioners air flow in both cases, the simulation with confinements (figure 8) propose cooler charged air in the cabinets. Similar conclusions can be made from figures 15 and 16.

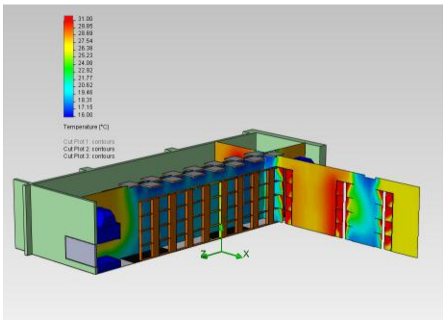


Figure 15: Vertical temperature distribution for cool air down flow with exit through air conditioners without any confinement

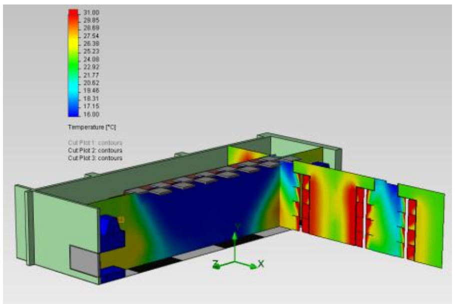


Figure 15: Vertical temperature distribution for cool air down flow with exit through air conditioners with confinement only above the cabinets

7. FUTURE OF THE PROJECT

It is essential to understand that most industrial data centers we have worked upon are two-lines one and the temperature sensors are placed far from the cold corridor found between the server centers lines. This is quite contradictory because the cold central corridor is the most crucial place for temperature control. In this section of the report, we propose an outline of the work being conducted in order to define a refined temperature control algorithm. In the article we develop differential equations and a control algorithm for a 2D lumped model of the cold corridor. This part of the article only describes a methodology that is yet to be modified to the server context. This will be part of the future of this presented project. In this case, we model 4 walls and a thermal power flux as generic as an energy $Q_{ser,}$ This model does perfectly represent the server center model but is used as starting point into applying predictive control to a future 3D model of the server room. Furthermore, we acknowledge that definition of perpendicular heat transport might be an important factor since the server room is not a confined one. In the future, effort will be made on the 2D model to compare predicted results with experimentally measured ones, for the same situations and improve the accuracy of the model as a result. Examination of the complete air conditioning system allows us to see that we can control the alimentation air temperature via valves controlling chilled water flow in the CRAC unit. The volumetric flow of the air can also be controlled. It is essential to keep in mind that, for any control mechanism, the norms of air content must be respected and be a primary condition in the algorithm code. An outline of a typical air conditioning system in a building is illustrated in figure 16, which is an excerpt from (Tabeli and al).

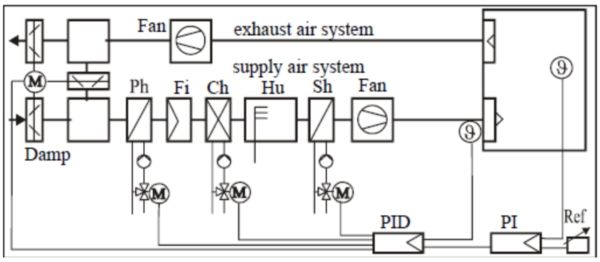


Figure 16: Typical Air conditioning system from (Talebi and al)

The air conditioning system with its control loops for a temperature cascade control is used for education and research on building energy management systems. In its structure it is equivalent to common industrial applications explaining the choice of this example. The supply air system consists of motorized dampers (Damp), a preheated exchanger (Ph), a chilled water exchanger (Ch) and a secondary heating exchanger (Sh) to condition the air temperature. The steam humidifier (Hu) and filter (Fi) are not taken into account for the temperature control (Talebi, R. Daryani and H. Plass, 1998; Talebi, R. Daryani, and H. Plass, 1998). In order to predicatively evaluate the temperature distribution and accordingly adjust alimentation temperature, a primary 2D model has been built. Despite the advent of higher accuracy techniques, a lumped method has as advantage the implementation of a simple model that allows real time simulation to propose predictive control mechanisms.

Predictive control will be applied on a 2D model expressed via state equations defined by differential equations. In this section, we will describe the equations for a simple four walled room with a given disturbance due to the data centers Q_{server}. Equations are only developed for one partition. In the future, such will be improved and adjusted to the real server centre room. Let T represent the temperature, Q, the heat flux, A, the surface are, X the thermal transmittance in $Wm^{-2}k^{-1}$, the subindex "i" representing interior, "o" representing exterior and "v" representing ventilation. Furthermore, "w" represents wall with subindexes 1, 2, 3 and 4 expressing the four different walls. "f" refers to the floor, "c" refers to the ceiling and "p" to a partition. "C" refers to capacitance. Finally, "t_{iair}" and "t_{oair}" represent the internal and external air temperature. The different temperatures shall be measured by different sensors. Application of state equations with known data center disturbance will allow for calculation of temperature gradient and hence predict future temperature. Henceforth, ventilation flaps and valves will be accordingly controlled to keep optimized temperature. The definition of state equations will be implemented in Simulink Linear library model. Derivatives will be integrated using a 4th order Runge Kutta method with a 0.1 hour integration interval. (This is assumed sufficient considering the macro variability scheduled in the data centre system). The simple four walls, one partition model is first schematized in an equivalent electrical diagram illustrated in figure 17 for simplicity reasons.

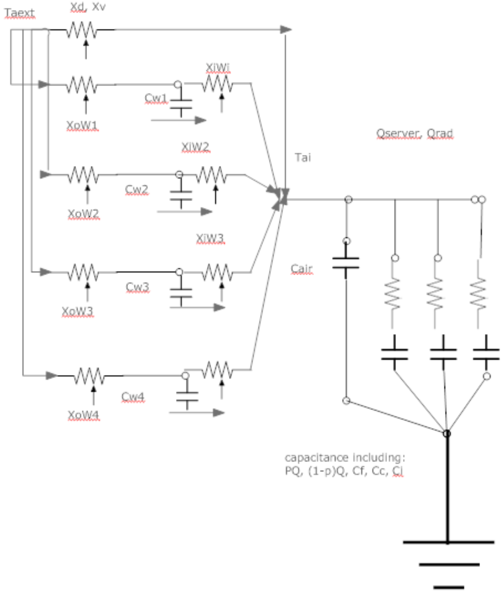


Figure 17: Equivalent Electrical schematic of thermal loads

Following equations describe the different thermal transfers at the different parts of the system: wall, partitions, etc.

Wall 1:

$$\frac{dT_{w1}}{dt} = \frac{Aw_1}{Cw_1}[X_i w_1(\mathrm{t_{iair}} - T_{w1}) + X_0 w_1(\mathrm{t_{oair}} - -T_{w1})] \tag{1}$$

Wall 2:

$$\frac{dT_{w2}}{dt} = \frac{Aw_2}{Cw_2}[X_i w_2(\mathrm{t_{iair}} - T_{w2}) + X_0 w_2(\mathrm{t_{oair}} - -T_{w2})] \tag{2}$$

Wall 3:

$$\frac{dT_{w3}}{dt} = \frac{Aw_3}{Cw_3}[X_i w_3(\mathrm{t_{iair}} - T_{w3}) + X_0 w_3(\mathrm{t_{oair}} - -T_{w3})] \tag{3}$$

Wall 4:

$$\frac{dT_{w4}}{dt} = \frac{Aw_4}{Cw_4}[X_i w_4(\mathrm{t_{iair}} - T_{w4}) + X_0 w_4(\mathrm{t_{oair}} - -T_{w4})] \tag{4}$$

Floor:

$$\frac{dT_f}{dt} = \frac{A_f}{C_f}\left[{\frac{PQ_{ext}}{A_f}}_i + X_f(\mathrm{t_{iair}} - -T_f)\right] \tag{5}$$

Ceiling:

$$\frac{dT_c}{dt} = \frac{A_c}{C_c}[X_c(\mathrm{t_{iair}} - T_c)] \tag{6}$$

Single partition:

$$\frac{dT_{ip}}{dt} = \frac{A_{ip}}{C_{ip}}\left[{\frac{(1-P)Q_{ext}}{A_{ip}}}_i + X_{ip}(\mathrm{t_{iair}} - -T_{ip})\right] \tag{7}$$

Internal Air:

$$\frac{dt_{iair}}{dt} = \frac{1}{C_a}\Big[Q_p + Q_{ext} + \left(A_g X_g\right) + (\mathrm{t}_{\mathrm{oair}} - \mathrm{t}_{\mathrm{iair}}) \\ + A_{w1}X_{iw1}(T_{w1} - \mathrm{t}_{\mathrm{iair}}) \\ + A_{w2}X_{iw2}(T_{w2} - \mathrm{t}_{\mathrm{iair}}) \\ + A_{w3}X_{iw3}(T_{w3} - \mathrm{t}_{\mathrm{iair}}) \\ + A_{w4}X_{iw4}(T_{w4} - \mathrm{t}_{\mathrm{iair}}) \\ + A_f X_f\left(T_f - \mathrm{t}_{\mathrm{iair}}\right) + A_c X_c(T_c - \mathrm{t}_{\mathrm{iair}}) \\ + A_{ip}X_{ip}\big(T_{ip} \\ - \mathrm{t}_{\mathrm{iair}}\big)\Big] \qquad (8)$$

The differential equations are then written in matrix form of the type:

$$\dot{T} = AT + Bi \qquad (9)$$

$\dot{T}$ is a vector of derivatives, A and B are matrices of coefficients, T is a vector of states and i is a vector of inputs. This leads us to the following matrix equation.

$$\dot{T} = \begin{bmatrix} \dot{T}_{w1} \\ \dot{T}_{w2} \\ \dot{T}_{w3} \\ \dot{T}_{w4} \\ \dot{T}_f \\ \dot{T}_c \\ \dot{T}_{ip} \\ \dot{t}_{iair} \end{bmatrix} \qquad (10)$$

$$A = \begin{bmatrix} \frac{-A_{w1}}{C_{w1}}[X_{iw1+}X_{ow1}] & 0 & 0 & 0 & 0 & 0 & 0 & \frac{A_{w1}X_{w1}}{C_{w1}} \\ 0 & -\frac{A_{w2}}{C_{w2}}[X_{iw2+}X_{ow2}] & 0 & 0 & 0 & 0 & 0 & \frac{A_{w2}X_{w2}}{C_{w2}} \\ 0 & 0 & -\frac{A_{w3}}{C_{w3}}[X_{iw3+}X_{ow3}] & 0 & 0 & 0 & 0 & \frac{A_{w3}X_{w3}}{C_{w3}} \\ 0 & 0 & 0 & -\frac{A_{w4}}{C_{w4}}[X_{iw4+}X_{ow4}] & 0 & 0 & 0 & \frac{A_{w4}X_{w4}}{C_{w4}} \\ 0 & 0 & 0 & 0 & -\frac{A_f}{C_f} & 0 & 0 & \frac{A_f X_f}{C_f} \\ 0 & 0 & 0 & 0 & 0 & \frac{-A_c X_c}{C_c} & 0 & \frac{A_c X_c}{C_c} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-A_p X_{ip}}{C_{ip}} & \frac{A_{ip}X_{ip}}{C_{ip}} \\ \frac{-A_{w1}x_{w1}}{C_a} & \frac{-A_{w2}x_{w2}}{C_a} & \frac{-A_{w3}x_{w3}}{C_a} & \frac{-A_{w4}x_{w4}}{C_a} & \frac{A_f X_f}{C_a} & \frac{A_c X_c}{C_a} & \frac{A_p X_{ip}}{C_a} & \frac{-1}{C_a}[A_g X_g + X_v + A_{w1}X_{v1} + A_2X_2 + A_3X_3 + A_4X_4 + A_f X_f + A_c X_c + A_{ip}X_{ip}] \end{bmatrix} \qquad (11)$$

$$T = \begin{bmatrix} T_{w1} \\ T_{w2} \\ T_{w3} \\ T_{w4} \\ T_f \\ T_c \\ T_{ip} \\ t_{iair} \end{bmatrix} \qquad (12)$$

$$B = \begin{bmatrix} 0 & 0 & 0 & \frac{A_{w1}X_{0w1}}{C_{w1}} \\ 0 & 0 & 0 & \frac{A_{w2}X_{0w2}}{C_{w2}} \\ 0 & 0 & 0 & \frac{A_{w3}X_{0w3}}{C_{w3}} \\ 0 & 0 & 0 & \frac{A_{w4}X_{0w4}}{C_{w4}} \\ 0 & 0 & \frac{p}{c_f} & 0 \\ 0 & 0 & \frac{1-p}{c_{ip}} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{C_a} & \frac{1}{C_a} & 0 & \frac{A_g X_g + X_v}{C_a} \end{bmatrix} \qquad (13)$$

$$i = \begin{bmatrix} Q_d \\ Q_{ser} \\ Q_s \\ T_{a0} \end{bmatrix} \qquad (14)$$

These equations make up the state equation. Inputs i are entered via WPPT like systems including historical and probabilistic data of data centre operation as well as sensor measured external data. This allows the differential equations inscribed in the matrices to predict the temperature gradient and temperature distribution. Henceforth, a controller (as a modification of the state equations) can be added to the system that will accordingly vary the CRAC unit commands to optimise temperature distribution. The *algorithmic* layout of this control algorithm is presented in figure 18.

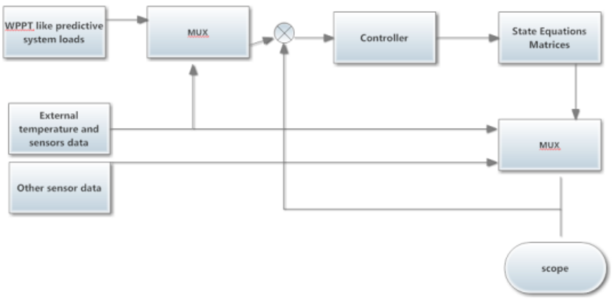


Figure 18: Preliminary layout of control algorithm

However, this model does not allow for direct control since the variable that we wish to control is the mean temperature of the cold air used to cool the server centers in the cold corridor. In practice, in order to take action on this temperature, two parameters are usually altered: the opening of the water valve in the heat exchanger coil of the cooling unit and the frequency of the CRAC engine. Since, $\dot{Q} = \dot{m} C_p \Delta T$, we can control the mean temperature in the cold corridor by acting on $\dot{m}$ which is the volumetric flow rate and the input cold air temperature to vary the cooling level of the data centers. The control loop is established with respect to a temperature sensor that measures the air temperature of the air returning to the cooling system or with respect to a thermostat situated in the cold corridor. Therefore, the control algorithm will make use of a large number of steady state simulations with three varying parameters. These are: the temperature of the input cold air, the volumetric flow of the cold air and the power output of the server centres. Analysis of the different simulations will allow us to define within a multi variable choice algorithm a correlation between the temperatures measured by the sensor to propose a control via correction from results obtained from the different simulations. Furthermore, the inclusion of statistical model providing information on the server centers power output and a model allowing quantification of energy and cost efficiency by varying the volumetric flow rate of the cold air, in the simulation-sensor control model, will allow real time optimisation according to different criteria. The state model can be used as a gradient matrix to verify simulation results at chosen simulation time steps.

8. CONCLUSION

In this paper we have presented work performed on an industrial data center in an attempt to define, according to given inputs, the temperature distribution in the cold and hot corridors. Emphasis has been laid in the modeling of the cold corridors. Moreover, different scenarios, including flow direction and partition have been evaluated using CFD methods. The CFD model has been calibrated as far as possible and the standard deviation between the measured and CFD predicted results have been explained and analyzed from both experimental (measurement errors, etc.) and numerical (turbulence and transport models) point of view. Different results have been illustrated to propose and enable the reader to evaluate the pertinence of such tools in data center and room temperature distribution modeling. Finally, the authors have pondered upon control algorithms and preliminary work based on a simplified 2D lumped model has been illustrated in this paper. In the future, we project to work on an optimized meshed model and propose a 3D predictive control model.

REFERENCES

A. Weissel and F. Bellosa, June 2004, Dynamic Thermal Management for Distributed Systems. In *Proceedings of the Workshop on Temperature-Aware Computer Systems (TACS)*.

C. D. Patel and A. J. Shah, 2005, Cost Model for Planning, Developement and Operation of a Data Center. *HPL 2005-107 (R.1), HP Lab. Technical Report*.

C. D. Patel, C. E. Bash, and M. Beitelmal, July 2003, Smart Cooling of Data Centers. In *Proceedings of the Pacific RIM/ASME International Electronics Packaging Technical Conference and Exhibition (IPACK)*.

C. D. Patel, C. E. Bash, C. Belady, L. Stahl, and D. Sullivan, July 2001, Computational Fluid Dynamics Modeling of High Compute Density Data Centers to Assure System Inlet Air Specifications. In *Proceedings of ASME International Electronic Packaging Technical Conference and Exhibition (IPACK)*.

C. D. Patel, R. Sharma, C. E. Bash, and A. Beitelmal, May 2002, Thermal Considerations in Cooling Large Scale High Compute Density Data Centers. In *Proceedings of Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*.

C. E. Bash, C. D. Patel, and P. K. Sharma, 2003, Efficient Thermal Management of Data Center - Immediate and Long-Term Research Needs. *Intl. J. Heat, Ventilating, Air-Conditioning and Refrigeration Research Needs*, 9(2):137–152.

D. Brooks, M. Martonosi, January 2001, Dynamic thermal management for High-Performance Microprocessors. In *Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA)*, pages 171–182.

D. Ramdenee & A. Ilinca 'Computational fluid dynamics introduction' Internal report, UQAR. 2011

E. Pinheiro and R. Bianchini, June 2004, Energy Conservation Techniques for Disk Array-Based Servers. In *Proceedings of the International Conference on Supercomputing (ICS)*.

F. Bellosa, S. Kellner, M. Waitz, and A. Weissel, September 2003, Event driven energy accounting of dynamic thermal management. In Proceedings of the Workshop on Compilers and Operating Systems for Low Power.

J. F. Karlsson and B. Moshfegh, 2005, Investigation of indoor climate and power usage in a data center. *Energy and Buildings*, 37:1075–1083.

J. Moore, J. Chase, , and P. Ranganathan, June 2006, ConSil: Lowcost Thermal Mapping of Data Centers. In *Proceedings of the Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML)*.

J. Moore, J. Chase, , and P. Ranganathan, June 2006, Weatherman: Automated, Online, and Predictive Thermal Mapping and Management for Data Centers. In *Proceedings* *of the International Conference on Autonomic Computing (ICAC)*.

J. Moore, J. Chase, P. Ranganathan, and R. Sharma, April 2005, Making Scheduling Cool: Temperature-Aware Workload Placement in Data Centers. In *Proceedings of the USENIX Annual Technical Conference*.

J. Moore, R. Sharma, R. Shih, J. Chase, C. D. Patel, and P. Ranganathan, May 2002, Going beyond CPUs: The Potential of Temperature-Aware Solutions for the Data Center. In *Proceedings of the Workshop of Temperature-Aware Computer Systems (TACS-1) held at ISCA.*

K. Skadron, M. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan and D. Tarjan, June 2003, Temperature-Aware Microarchitecture. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, pages 1–13

L. Shang, L.-S. Peh, A. Kumar, and N. Jha, December 2004, Thermal Modeling, Characterization and Management of On-Chip Networks. In *Proceedings of the International Symposium on Microarchitecture (MICRO)*, pages 67–78.

M. Gomaa, M. D. Powel, and T. N. Vijaykumar, 2004, Heat-and-Run: Leveraging SMT and CMP to Manage Power Density Through the Operating System. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 260–270.

M. H. Beitelmal and C. D. Patel, 2004, Thermo-Fluids Provisioning of a High Performance High Density Data Center. *HPL 2004-146 (R.1), HP Lab. Technical Report*.

N. Rasmussen. Cooling Strategies for Ultra-High Density Racks and Blade Servers. In *APC White Paper 46, http://www.apcc.com/prod docs/results.cfm?class=wp&allpapers=1*.

N. Rasmussen. Guidelines for Specification of Data Center Power Density. In *APC White Paper 120 http://www.apcc.com/prod docs/results.cfm?class=wp&allpapers=1*.

P. Rodgers and V. Eveloy, 2003, Prediction of Microelectronics Thermal Behavior in Electronic Equipment: Status, Challenges and Future Requirements. In *Proceedings of the International Conference on Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems*.

R. Huang and D. Chung, May 2002, Thermal Design of a Disk-Array System. In *Proceedings of the InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, pages 106–112.

R. K. Sharma, C. E. Bash, and C. D. Patel, June 2002, Dimensionless parameters for evaluation of thermal design and performance of large-scale data centers. In *8th ASME/AIAA Joint Thermophysics and Heat Transfer Conf*.

R. Sharma, C. Bash, C. Patel, R. Friedrich, and J. Chase, January 2005, Balance of Power: Dynamic Thermal Management for Internet Data Centers. *IEEE Internet Computing*, 9(1):42–49.

S. Gurumurthi, A. Sivasubramaniam, and V. Natarajan, June 2005, Disk Drive Roadmap from the Thermal Perspective: A Case for Dynamic Thermal Management. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, pages 38–49.

S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke, June 2003, DRPM: Dynamic Speed Control for Power Management in Server Class Disks. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, pages 169–179.

S. V. Patankar and K. C. Karki, 2004, Distribution of Cooling Airflow in a Raised-Floor Data Center. In *American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE)*.

T. Heath, A. P. Centeno, P. George, Y. Jaluria, and R. Bianchini, October 2006, Mercury and Freon: Temperature Emulation and Management in Server Systems. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*.

Talebi et al. 'Application of fuzzy Logic for control and Energy management of Air conditioning, Chilling and heating'' reference book.

Talebi, R. Daryani and H. Plass, May 1998, Application of fuzzy control for intelligent building part I: fuzzy control for an AC system, WAC'98 WAC'98 World Automation Congress, Anchorage, Alaska CD-ROM–ISBN–1-889335-07-X.

Talebi, R. Daryani and H. Plass, 1998, Application of fuzzy control for intelligent building part I: fuzzy control for an AC system, intelligent automation and control, Proceedings of the WAC'98, p.745 – 750, TSI press Series, ISBN 0-9627451-7-0, Albuquerque, NM,USA.

Y. Kim, S. Gurumurthi, and A. Sivasubramaniam, Febuary 2006. Understanding the Performance-Temperature Interactions in Disk I/O of Server Workloads. In *Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA)*, pages 179–189.

NEURAL NETWORK SIMULATION OF SIMPLE BIOLOGICAL SYSTEMS INCLUDING OPTIMAL CONTROL PROBLEMS

Tibor Kmet[a], Maria Kmetova[b]

[a]Constantine the Philosopher University in Nitra, Department of Informatics, Tr. A. Hlinku 1, 949 74 Nitra, Slovakia
[b]Constantine the Philosopher University in Nitra, Department of Mathematics, Tr. A. Hlinku 1, 949 74 Nitra, Slovakia

[a]tkmet@ukf.sk, [b]mkmetova@ukf.sk

ABSTRACT

The mechanistic model of the phytoplankton photosynthesis-light intensity relationship and nitrogen transformation cycle are investigated. Assuming that phytoplankton regulates its photosynthetic production rate with certain strategy which maximizes photosynthetic and biomass production, respectively two such possible strategies were examined using a neural network based optimal control synthesis for solving fixed and free final time optimal control problems with control and state constraints. The optimal control problem is transcribed into nonlinear programming problem which is implemented with adaptive critic neural network. Results show that adaptive critic based systematic approach holds promise for obtaining the fixed and free final time optimal control with control and state constraints.

Keywords: optimal control problem, state and control constraints, free terminal time, neural network simulation, phytoplankton photosynthesis, nitrogen transformation cycle, optimal photosynthetic production

1. INTRODUCTION

Optimal control of nonlinear systems is one of the most active subjects in control theory. There is rarely an analytical solution although several numerical computation approaches have been proposed (for example, see (Kirk, 1989), (Polak, 1997). The most of the literature dealing with numerical methods for the solution of general optimal control problems focuses on algorithms for solving discretized problems. The basic idea of these methods is to apply nonlinear programming techniques to the resulting finite dimensional optimization problem (Buskens, Maurer, 2000). When Euler integration methods are used, the recursive structure of the resulting discrete time dynamic can be exploited in computing first-order necessary condition.

In the recent years the neural networks are used for obtaining numerical solutions to optimal control problem (Padhi, Unnikrishnan, Wang, and Balakrishnan, 2001), (Padhi, Balakrishnan and Randoltph, 2006). For the network, a feed forward network with one hidden layer, a steepest descent error backpropagation rule, a hyperbolic tangent sigmoid transfer function and a linear transfer function were used.

The paper presented extends adaptive critic neural network architecture proposed by Padhi, Unnikrishnan, Wang and Balakrishnan (2001) to the optimal control problems with control and state constraints. The organization of the paper is as follows. In Section 2 optimal control problems with control and state constraints are being introduced. We summarize necessary optimality conditions and give a short overview on basic result including iterative numerical methods and discussed discretization methods for given optimal control problem and a form of resulting nonlinear programming problems. Section 3 presented a short description of adaptive critic neural network synthesis for optimal problem with state and control constraints. Section 4 consists of a mechanistic model of phytoplankton photosynthesis. We prove the existence of unique globally asymptotically stable equilibrium depending on light intensity. Using optimal control theory we maximize photosynthetic production rate for fixed and free final time. Section 5 presented a nitrogen transformation cycle. We investigate a preferential utilization of nitrogen compounds by phytoplankton using adaptive critic neural network. We examine short and long-term strategy of utilization. Conclusions are being presented in Section 6.

2. OPTIMAL CONTROL PROBLEM

We consider nonlinear control problem subject to control and state constraints. Let $x(t) \in R^n$ denote the state of a system and $u(t) \in R^m$ the control in a given time interval $[t_0, t_f]$.

Optimal control problem is to minimize

$$J(x,u) = g\left(x(t_f)\right) + \int_{t_0}^{t_f} f_0\big(x(t),u(t)\big)dt \qquad (1)$$

subject to

$$\dot{x}(t) = f\big(x(t),u(t)\big),$$

$$x(t_0) = x_0,$$

$$\psi(x(t_f)) = 0,$$

$$c\big(x(t),u(t)\big) \le 0,\ \ t \in \big[t_0,t_f\big].$$

The functions $g: R^n \to R,\ f_0: R^{n+m} \to R,$ $f: R^{n+m} \to R^n, c: R^{n+m} \to R^q$ and $\psi: R^{n+m} \to R^r, 0 \le r \le n$ are assumed to be sufficiently smooth on appropriate open sets. The theory of necessary conditions for optimal control problem of form (1) is well developed (Kirk, 1989), (Pontryagin, Boltyanskii, Gamkrelidze and Mischenko, 1983).

We introduce an additional state variable

$$x_0(t) = \int_0^t f_0\big(x(s),u(s)\big)ds$$

defined by the

$$\dot{x}(t) = f_0\big(x(t),u(t)\big), x_0(0) = 0.$$

Then the augmented Hamiltonian function for problem (1) is

$$H(x,\lambda,\mu,u) = \sum_{j=0}^{n} \lambda_j f_j(x,u) + \sum_{j=0}^{q} \mu_j c_j(x,u), \qquad (2)$$

where $\lambda \in R^{n+1}$ is the adjoint variable and $\mu \in R^q$ is a multiplier associated to the inequality constraints. Let $(\hat{x},\hat{u})$ be an optimal solution for (1) then the necessary condition for (1) (Kirk, 1989), (Pontryagin, Boltyanskij, Gamkrelidze and Mischenko,1983) implies that there exist a piecewise continuous and piecewise continuously differentiable *adjoint function* $\lambda: \big[t_0,t_f\big] \to R^q,\ \mu(t) \ge 0$ and a multiplier $\sigma \in R^r$ satisfying

$$\dot{\lambda}_j(t) = -\frac{\partial H}{\partial x_j}(\hat{x}(t),\lambda(t),\mu(t),\hat{u}(t))$$

$$\lambda_j\big(t_f\big) = g_{x_j}\Big(\hat{x}\big(t_f\big)\Big) + \sigma\psi_{x_j}\Big(\hat{x}\big(t_f\big)\Big), j = 0,\dots,n \quad (3)$$

$$\dot{\lambda}_0(t) = 0$$

$$0 = \frac{\partial H}{\partial u}\big(\hat{x}(t),\lambda(t),\mu(t),\hat{u}(t)\big).$$

For free terminal time t_f, an additional condition needs to be satisfied:

$$H\big(t_f\big) = \Big(\sum_{j=0}^{n} \lambda_j f_j(x,u) + \sum_{j=0}^{q} \mu_j c_j(x,u)\Big)\big|_{t_f} = 0.$$

Furthermore, the complementary conditions hold i.e. in $t \in \big[t_0,t_f\big],\ \mu \ge 0,\ c(x,u) \le 0$ and $\mu\, c(x,u) = 0$. Herein, the subscript x or u denotes the partial derivative with respect to x or u.

2.1. Discretization of optimal control problem

Direct optimization methods for solving the optimal control problem are based on a suitable discretization of (1). Choose a natural number N and let $t_i \in \big[t_0,t_f\big], i = 1,\dots,N-1$, be an equidistant mesh point with $t_i = t_0 + ih,\ i = 1,\dots,N,$ where h is time step and $t_f = Nh + t_0$. Let the vectors $x^i \in R^{n+1}, u^i \in R^m,\ i = 1,\dots,N,$ be approximation of state variable and control variable $x(t_i)$, $u(t_i)$, respectively at the mesh point. *Euler's* approximation applied to the differential equations yields

$$x^{i+1} = x^i + hf(x^i,u^i), \qquad i = 0,\dots,N-1.$$

Choosing the optimal variable $z := (x^0,x^1,\dots,x^{N-1},u^0,\dots,u^{N-1}) \in R^{N_s}, N_s = (n+1+m)N$, the optimal control problem is replaced by the following discretized control problem in the form of nonlinear programming problem with inequality constraints:

$$min\, J(z) = G(x^N),$$

where

$$G(x^N) = g((x_1,\dots,x_n)^N,t_N) + x_0^N, \qquad (4)$$

subject to

$$-x^{i+1} + x^i + hf(x^i,u^i) = 0,$$

$$x^0 = x(t_0),$$

$$\psi(x^N) = 0,$$

$$c(x^i,u^i) \le 0,\ i = 0,\dots,N-1.$$

In a discrete-time formulation we want to find an admissible control which minimizes object function (4). Let us introduce the Lagrangian function for the nonlinear optimization problem (4):

$$L(z,\lambda,\sigma,\mu,h) = \sum_{i=0}^{N-1} \lambda^{i+1}(-x^{i+1} + x^i + f(x^i,u^i)) + $$
$$+G(x^N,t_N) + \sum_{i=0}^{N-1} \mu^i c(x^i,u^i) + \sigma\psi(x^N,t_N). \qquad (5)$$

and define $H(i)$ and Φ as a follows:

$$H(i) = \lambda(i+1)(x^i + hf(x^i,u^i),$$
$$\Phi = G + \sigma\psi.$$

The first order optimality conditions of Karush-Kuhn-Tucker (Polak, 1997) for the problem (4) are:

$$\begin{aligned} 0 = L_{x^i}(s,\lambda,\mu,h) &\\ &= \lambda^{i+1} + h\lambda^{i+1} f_{x^i}(x^i,u^i) - \lambda^i \\ &+ \mu^i c_{x^i}(x^i,u^i), i = 0,\dots,N-1, \quad (6) \end{aligned}$$

$$0 = L_{x^N}(s,\lambda,\mu,h) = G_{x^N}(x^N) + \sigma\psi_{x^N}(x^N,t_N) - \lambda^N \quad (7)$$

$$0 = L_{u^i}(s,\lambda,\mu,h) = h\lambda^{i+1} f_{u^i}(x^i,u^i) + \mu^i c_{u^i}(x^i,u^i),$$
$$i = 0,\dots,N-1, \quad (8)$$

$$0 = L_h(s,\lambda,\mu,h) = \Phi_h + \sum_{i=0}^{N-1} H_h(i) + \sum_{i=0}^{N-1} \mu^i c_h(x^i,u^i), \quad (9)$$

Eq. (6-9) represents the discrete version of necessary condition (3) for optimal control problem (1).

3. ADAPTIVE CRITIC NEURAL NETWORK FOR OPTIMAL CONTROL PROBLEM WITH CONTROL AND STATE CONSTRAINTS AND FREE TERMINAL CONDITION

It is well known that a neural network can be used to approximate smooth time-invariant functions and uniformly time-varying function (Hornik, Stichcombe and White, 1989), (Sandberg, 1998). Neurons are grouped into distinct layers and interconnected according to a given architecture (Figure 1). Each connection between two neurons has a weight coefficient attached to it. The standard network structure for an approximation function is the multiple-layer perceptron (or feed forward network). The feed forward network often has one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons.

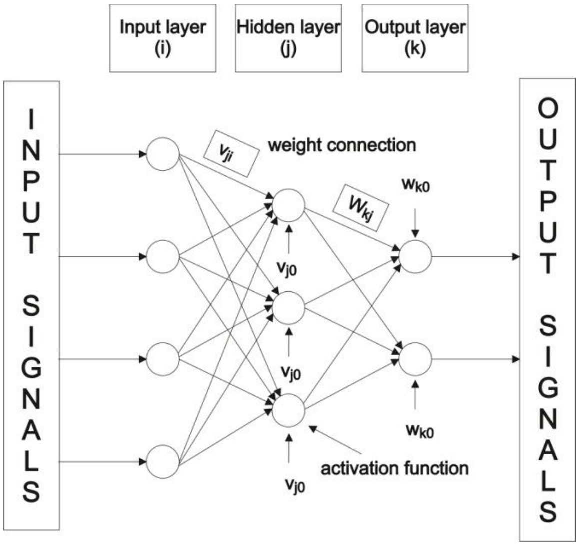


Figure 1: Feed Forward Neural Network Topology with One Hidden Layer (v_{ji}; w_{kj} are values of connection weights, v_{j0}; w_{k0} are values of bias)

Figure 1 shows a feed forward neural network with n_i inputs nodes one layer of n_{hl} hidden units and n_o output units. Let $in = [in_1,\dots,in_{n_i}]$ and $out = [out_1,\dots,out_{n_o}]$ be the input and output vectors of the network, respectively. Let $V = [v_1,\dots,v_{n_{hl}}]$ be the matrix of synaptic weights between the input nodes and the hidden units, where $v_j = [v_{j0}, v_{j1},\dots,v_{jn_i}]$ and v_{j0} is the bias of the jth hidden unit.

Let also $\mathrm{W} = [\mathrm{w}_1,\dots,\mathrm{w}_{\mathrm{n_o}}]$ be the matrix of synaptic weights between the hidden and output units, where $\mathrm{w_k} = [\mathrm{w_{k0}}, \mathrm{w_{k1}},\dots,w_{kn_o}]$ and w_{k0} is the bias of the kth output unit, w_{kj} is the weight that connects the jth hidden units to the kth output unit.
The response of the jth hidden unit is given by

$$hl_j = tanh\left(\sum_{i=0}^{n_i} v_{ji} in_k\right),$$

where $tanh(.)$ is the activation function for the hidden units. The response of the kth output unit is given by

$$out_k = \sum_{j=0}^{n_{hl}} w_{kj} hl_j.$$

Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. The number of neurons in the input and output layers is given, respectively, by the number of input and output variables in the process under investigation.

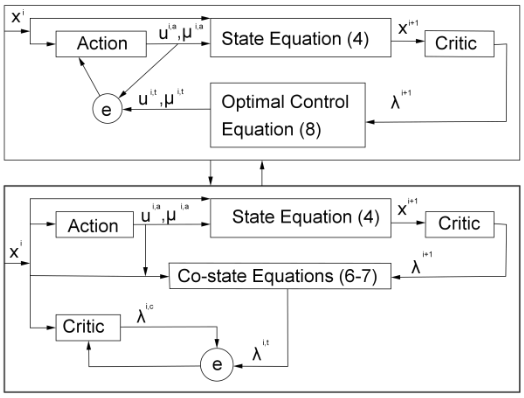


Figure 2: Architecture of Adaptive Critic Network Synthesis

The multi-layered feed forward network shown in Figure 2 is training using the steepest descent error backpropagation rule. Basically, it is a gradient descent, parallel distributed optimization technique to minimise the error between the network and the target output (Rumelhart, Hinton and Wiliams, 1987).

In the Pontryagin's maximum principle for deriving an optimal control law, the interdependence of the state, costate and control dynamics is made clear. Indeed, the optimal control $\hat{u}$ and multiplier $\hat{\mu}$ is given by Eq. (8), while the costate Eqs. (6-7) evolves backward in time and depends on the state and control. The adaptive critic neural network is based on this relationship. It consists of two networks at each node: an action network the inputs of which are the current states and outputs are the corresponding control $\hat{u}$ and multiplier $\hat{\mu}$, and the critic network for which the current states are inputs and current costates are outputs for normalizing the inputs and targets (zero mean and standard deviations). For detail explanation see (Rumelhart, Hinton and Wiliams, 1987), (Kmet, 2011).

From free terminal condition ($\psi(x) \equiv 0$) and from Eqs. (6-7) we obtain that $\lambda_0^i = -1$ for $i = N,\ldots,0$ and $\lambda_j^N = 0$ for $j = 1,\ldots, n$. We use this observation before proceeding to the actual training of the adaptive critic neural network. The steps for training the action network are as follows:

1) Generate set S. For all $x^k \in S$,
follow the steps below:
(1.i) Input x^k to the action network to obtain
$u^{k;a}$ and $\mu^{k;a}$.
(1.ii) Using x^k and $u^{k;a}$ solve state
equation (4) to get x^{k+1}.
(1.iii) Input x^{k+1} to the critic network
to obtain μ^{k+1}.
(1.iv) Using x^k and μ^{k+1} solve (8)
to calculate $u^{k;t}$ and $\mu^{k;t}$.

When

$$\|(u^{k,a},\mu^{k,a}) - (u^{k,t},\mu^{k,t})\| \,/\, \|(u^{k,t},\mu^{k,t})\| < \epsilon_a,$$

the convergence criterion for the action network training is met.

The training procedure for the critic network which expresses the relation between x^k and λ^k is as follows:

1) Generate set S. For all $x^k \in S$,
follow the steps below:
(1.i) Input x^k to the action network to obtain
$u^{k;a}$ and $\mu^{k;a}$.
(1.ii) Using x^k and $u^{k;a}$ solve state
equation (4) to get x^{k+1}.
(1.iii) Input x^{k+1} to the critic network
to obtain λ^{k+1}.
(1.iv) Using x^k , $u^{k,a}$, $\mu^{k,a}$ and λ^{k+1} solve (6)
to calculate $\lambda^{k,t}$.
(1.v) Input x^k to the critic network
to obtain $\lambda^{k,c}$.

When

$$\|\lambda^{k,c} - \lambda^{k,t})\| \,/\, \|\lambda^{k,t}\| < \epsilon_c,$$

the convergence criterion for the action network training is met.

Further discussion and detail explanation of this adaptive critic method can be found in (Hornik, Stichcombe, White, 1989), (Padhi, Unnikrishnan, Wang and Balakrishnan, 2001), (Padhi, Balakrishnan and Randoltph, 2006), (Werbos, 1992),.

4. A MECHANISTIC MODEL OF PHYTOPLANKTON PHOTOSYNTHESIS

Mathematical models of photosynthesis in bioreactors are important for both basic science and the bioprocess industry (Garcia-Camacho et al., 2012). There is a class of models based on the concept of the "photosynthetic factories" developed by Eilers and Peeters (1988). The dynamic behaviour of the model has also been discussed in (Eilers and Peeters, 1993), (Kmet, Straskraba and Mauersberger 1996), (Papacek, Celikovsky, Rehak and Stys, 2010), (Wu and Merchuk, 2001) Assuming that phytoplankton regulates its photosynthetic production rate with a certain strategy which maximize production, two such possible strategies is examined, i.e. instantaneous and the integral maximal production.

4.1. Description of the Model

Basic for the following consideration is the mechanistic model of phytoplankton photosynthesis. It is based on unit processes concerning the cellular reaction centres called photo-synthetic-factories - PSF. It is known from algal physiology (Eilers, Peeters, 1988) that three states of a PSF are possible: x_1- resting, x_2 - activated and x_3 - inhibited. Transitions between states depend both on light intensity and time. The probabilities of the PSF being in the state x_1, x_2 or x_3, are given as p_1, p_2 and p_3, respectively. Transitions between states can be expressed as follows:

$$\begin{aligned}\dot{p}_1 &= -\alpha I p_1 + \gamma p_2 + \delta p_3 \qquad (10)\\ \dot{p}_2 &= \alpha I p_1 - (\beta I + \gamma) p_2\\ \dot{p}_3 &= \beta I p_2 - \delta p_3.\end{aligned}$$

The parameters α, β, γ and δ occurring in this model are positive constants and I is a light intensity.

Let $p(t, p^0)$ be a solution of (10) with the initial condition $p(0, p^0) = p^0$, where $p_1^0 + p_2^0 + p_3^0 = 1$. Note that solutions of the system (10) exist for all $t \geq 0$. By adding up the right-hand side of (10) we get

$$\dot{p}_1 + \dot{p}_2 + \dot{p}_3 = 0,$$

i.e. $\sum_{i=1}^{3} p_i(t, p^0) = 1$ for all $t \geq 0$. Of course these equations are considered in

$$S = \{p \in R_+^3 : p_1 + p_2 + p_3 = 1\}.$$

4.2. Global behaviour under constant condition

Proposition. *Let parameters $\alpha, \beta, \gamma, \delta$ and I be positive, then there exists an unique positive equilibrium $\bar{p}$ which is globally asymptotically stable on S.*

Proof: Vector $\bar{p}$ is the solution of the following linear equation system:

$$P \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad (11)$$

where

$$P = \begin{pmatrix} -\alpha I & \gamma & \delta \\ \alpha I & -\beta I - \gamma & 0 \\ 0 & \beta I & -\delta \end{pmatrix}.$$

Let us consider the matrix $D = P + \rho E$, where $\rho = max|P_{ii}|$ and E is the unit matrix. D is an irreducible nonnegative matrix and so the apparatus of the Perron-Frobenius theory of a nonnegative matrix applies. Since $1P = 0$, also $1D = \rho 1$, where 1 denotes the unit vector. By the Perron- Frobenius theory there is a unique positive right eigenvector r associated with the eigenvalue ρ and we can normalize to get

$$\bar{p}_i = {r_i}/{\sum_{i=1}^{3} r_i}.$$

Since $(P + \rho E)\bar{p} = \rho\bar{p}$, then $P\bar{p} = 0$ and $\bar{p} > 0$. By the Perron-Frobenius theorem we get, (see Akin (1979) that the matrix P is a stable matrix on S, i.e. P has one zero eigenvalue and the other eigenvalues have negative real parts. This proves the statement of the proposition. It follows that system (10) has a unique positive equilibrium $\bar{p}$ with entries

$$\begin{aligned} \bar{p}_1(I) &= \frac{\beta\delta I + \gamma\delta}{F} \qquad (12) \\ \bar{p}_2(I) &= \frac{\alpha\delta I}{F} \\ \bar{p}_3(I) &= \frac{\alpha\beta I^2}{F}, \end{aligned}$$

where $F = \alpha\beta I^2 + (\alpha + \beta)\delta I + \gamma\delta$.

The simplex S is positively invariant. System (10) for all $I \geq 0$ has a unique positive equilibrium $\bar{p}(I)$ with entries given by (12).
The equilibrium $\bar{p}(I)$ is globally asymptotically stable on S, that means for fixed light intensity I all solutions with initial condition $p(0) \in S$ converge to $\bar{p}(I)$.

4.3. Optimization of Photosynthetic Production

Let us assume that phytoplankton regulates its photosynthetic production rate (FP) with a certain strategy which maximizes production. The rate of the photosynthetic production FP is proportional to the number of transitions from x_2 to x_1. Let us investigate the optimal values of light intensity $I(t)$, for which the photosynthetic production $FP = \gamma p_2(t)$ is maximal under constraints $I \in [I_{min}, I_{max}]$.
We will examine two strategies:
1. Instantaneous maximal photosynthetic production with respect to I, (local optimality), i.e.

$$\dot{p}_2 = f_2(p, I, t) \rightarrow max$$

for all t, under the constraints $I \in [I_{min}, I_{max}]$.
2. Integral maximal biomass (global optimality), i.e.

$$J(I) = \int_0^{t_f} \gamma p_2(t) dt,$$

under the constraints $I \in [I_{min}, I_{max}]$.

4.4. Local Optimality

In the case of strategy 1, we maximize the following function:

$$J\big(I(t)\big) = I(t)\alpha p_1(t) - I(t)\beta p_2(t)$$

under the constraints $I \in [I_{min}, I_{max}]$. For $p(t) = \bar{p}(I)$ we examine the following function:

$$FP(I) = \frac{\alpha\delta I\gamma}{\delta\gamma + (\beta\delta + \alpha\delta)I + \alpha\beta I^2}.$$

By straightforward calculation we get that the optimal light intensity is given by

$$I^* = \sqrt{\frac{\gamma\delta}{\alpha\beta}}.$$

4.5. Global Optimality

In case of strategy 2, we have the following optimal control problem: Find a function $\hat{I}(t)$ for which the goal function

$$J(I) = \int_0^{t_f} \gamma p2(t)\, dt$$

attains its maximum, where t_f is fixed. We introduce an additional state variable

$$p_0(t) = \int_0^{t} \gamma p_2(s) ds \qquad (13)$$

defined by

$$\dot{p}_0(t) = \gamma p_2(t), \qquad p_0(0) = 0.$$

We are led to the following optimal control problems:

$$Maximize\; p_0\big(t_f\big) \qquad (14)$$

under the constraints

$$c_1(p, I) = I_{min} - I \leq 0$$
$$c_2(p, I) = I - I_{max} \leq 0.$$

Discretization of Eqs. (10, 13, 14) using Eqs. (6, 7, 8) and state equation (4) leads to

$$Minimize\; (-p_0^N)$$

subject to
$$p^{i+1} = p^i + hf(p^i, I^i), \qquad i = 0, \ldots, N-1,$$

$$\lambda_i = \lambda^{i+1} + h\lambda^{i+1} f_{x^i}(x^i, u^i) + \mu^i c_{x^i}(x^i, u^i),$$
$$i = N-1, \dots, 0,$$

$$\lambda_0^i = -1, \qquad i = 0, \dots, N-1,$$

$$\lambda^N = (-1,0,0,0),$$

$$0 = h\lambda^{i+1} f_{I^i}(p^i, I^i) + \mu^i c_{I^i}(p^i, I^i),$$

where the vector function

$$F(p, I) = (-\gamma p_2, f_1(p, I), \dots, f_3(p, I))$$

is given by Eq. (13) and by right-hand side of Eq. (10). In the adaptive critic synthesis, the critic and action network were selected such that they consist of three and two subnetworks, respectively, each having 3-18-1 structure (i.e. three neurons in the input layer, eighteen neurons in the hidden layer and one neuron in the output layer).

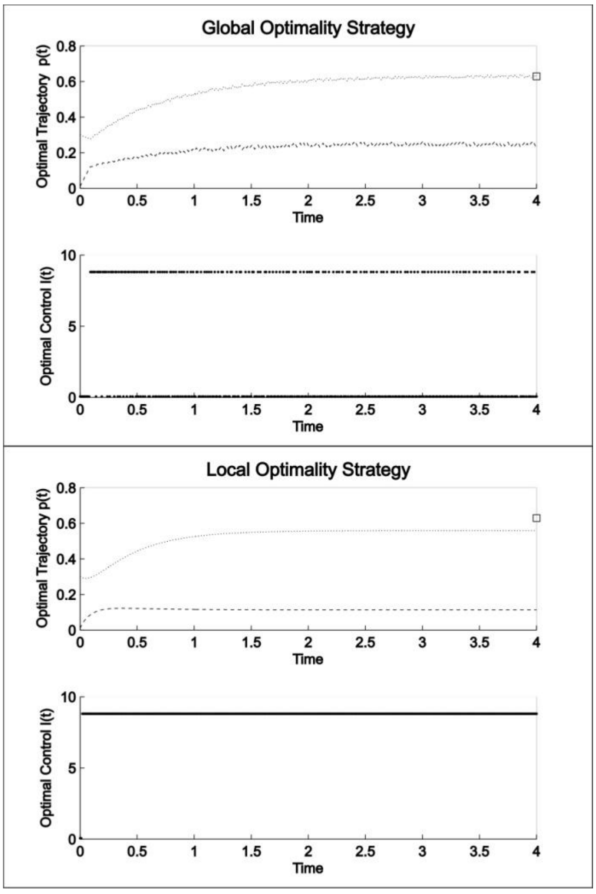


Figure 3: Adaptive Critic neural Network Simulation of Optimal Control $\hat{I}(t)$ and $\tilde{I}(t)$ for Global and Local Strategies, respectively with Fixed Final Time, dotted line $\hat{p}_1(t)$, $\tilde{p}_1(t)$, dashed line $\hat{p}_2(t)$, $\tilde{p}_2(t)$

The results of numerical solutions (Figure 3) have shown that the optimal strategies $\tilde{I}(t)$ and $\hat{I}(t)$ based on short or long-term perspective, respectively, have different time trajectory, for a given initial condition $\tilde{I}(t) = I_{max}$ and optimal control $\hat{I}(t)$ for long-term strategies obtains extreme values of the control set, i.e. optimal control is bang-bang. For long-term strategy optimal trajectory $\hat{p}_2(t)$ converges to $\bar{p}_2(I^*)$. Therefore let us consider the following free final time optimal control problems

$$J(t_f, I) = \int_0^{t_f} \gamma p_2(t) dt$$

and

$$J(t_f, I) = \int_0^{t_f} dt$$

with final condition $\psi(p(t_f)) = 0$, where $\psi(p) = (p_1 - \bar{p}_1(I^*), p_2 - \bar{p}_2(I^*))$. Results of adaptive-critic simulations are shown in Figure 4.

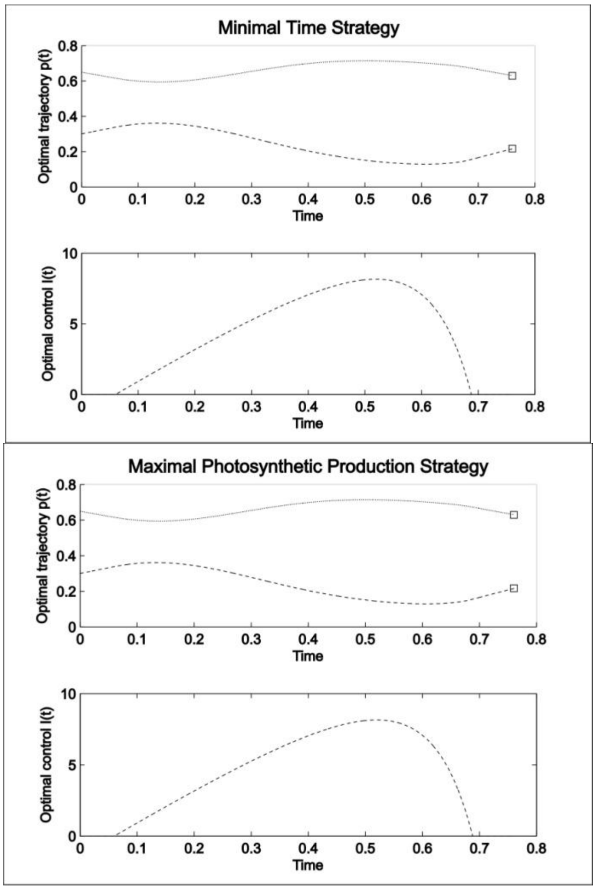


Figure 4: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{I}(t)$ for Maximal Photosynthetic Production and Minimal Time, respectively to a Point $\bar{p}_2(I^*)$ with Initial Condition $x(0) = (0.3;\ 0.65;\ 0.05)$ (dotted line $\hat{p}_1(t)$, dashed line $\hat{p}_2(t)$)

The results of numerical calculations have shown that the proposed adaptive critic neural network is able to meet the convergence tolerance values that we choose, which led to satisfactory simulation results. Simulations, using MATLAB show that proposed neural network is able to solve nonlinear free final time optimal control problem with state and control constraints.

5. NITROGEN TRANSFORMATION CYCLE

5.1. Description of the Model

The aerobic transformation of nitrogen compounds (Kmet, 1996), (Kmet, 2009) includes:

- the decomposition of complex organic substances into simpler compounds, ammonium being the final nitrogen product,
- ammonium and nitrate oxidation,
- the assimilation of nitrates.

Specific groups of microorganisms participate in these processes. Heterotrophic bacteria (x_1) assimilates and decomposes the soluble organic nitrogen compounds DON x_6 derived from detritus x_5. Ammonium x_7, one of the final decomposition products undergoes a biological transformation into nitrate x_9. This is carried out by aerobic chemoautotrophic bacteria in two stages: ammonia is first oxidized by nitrifying bacteria from the genus Nitrosomonas x_2 into nitrites x_8 that serve as an energy source for nitrating bacteria mainly from the genus Nitrobacter x_3. The resulting nitrates may be assimilated together with ammonia and soluble organic forms of nitrogen by the phytoplankton x_4, whereby the aerobic transformation cycle of nitrogen compounds is formed (Figure 5).

The individual variables $x_1,\dots,x_9$ represent nitrogen concentrations contained in the organic as well as in inorganic substances and living organisms presented in a model.

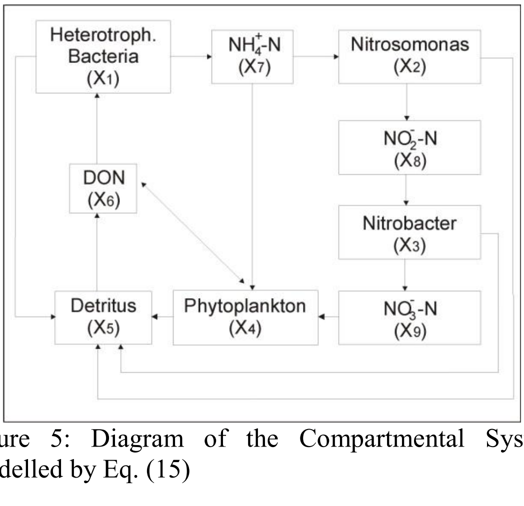


Figure 5: Diagram of the Compartmental System Modelled by Eq. (15)

The following system of ordinary differential equations is proposed as a model for the nitrogen transformation cycle:

$$\dot{x}_i = x_i(U_i(x) - E_i(x) - M_i(x))$$
$$\dot{x}_5 = \sum\nolimits_{i=1}^{4} x_i M_i(x) - K_5 x_5$$
$$\dot{x}_6 = K_5 x_5 - x_1 U_1(x) + x_4 E_4(x) - x_4 P_6(x) \qquad (15)$$
$$\dot{x}_7 = x_1 E_1(x) - x_2 U_2(x) - x_4 P_7(x)$$
$$\dot{x}_8 = x_2 E_2(x) - x_3 U_3(x)$$
$$\dot{x}_9 = x_3 E_3(x) - x_4 P_9(x)$$

where x_i are the concentration of the recycling matter in microorganisms, the available nutrients and detritus, respectively (15).

$$U_i(x) = \frac{K_i x_{i+5}}{1 + g_i x_{i+5}} \quad \text{for } i = 1,2,3,4$$
$$p = u_1 x_6 + u_2 x_7 + u_3 x_9$$

$$U_4(x) = \frac{K_4 p}{1 + g_4 p}\ , \quad U - \text{uptake rate}$$

$$L_i(x) = \frac{a_{2i-1} U_i(x)}{1 + a_{2i} U_i(x)} + 1 - \frac{a_{2i-1}}{a_{2i}},$$
$$L - \text{excretion activity}$$

$$M_i(x) = g_{2i+3} + g_{2i+4} L_i(x), \qquad M - \text{mortality rate}$$

$$E_i(x) = U_i(x) L_i(x) \quad \text{for } i = 1,2,3,4,$$
$$E - \text{excretion rate}$$

$$P_i(x) = \frac{K_4 u_i x_i}{1 + g_4 p} \quad \text{for } i = 6,7,9.$$

5.2. Optimal Biomass Production

The variables $u = (u_1, u_2, u_3)$ express the preference coefficients for update of x_6, x_7, x_9. It can be expected that the phytoplankton will employ control mechanisms in such a way as to maximize its biomass over a given period *T* of time:

$$J(u) = \int_0^{t_f} x_4(t)dt \ \rightarrow max$$

under the constraints

$$c(x,u) := b_1 U_4(x,u) + b_2 P_6(x,u) + b_3 P_9(x,u) + b_4 E_4(x,u) \le W(I),$$

$u_i \in [0, u_{i\,max}]$ for $i = 1,2,3.$

The last inequality expresses the fact that amount of energy used for ”living expenses” (synthesis, reduction and excretion of nutrients) by phytoplankton cannot exceed a certain value *W*(*I*) which depends on light intensity *I* (for detail explanation see (Kmet, 1996)). We introduce an additional state variable

$$x_0(t) = \int_0^t x_4(s)ds \qquad (16)$$

defined by

$$\dot{x}_0(t) = x_4(t), x_0(0).$$

We are led to the following optimal control problems:

1) long-therm strategy:

$$Maximize\ x_0\left(t_f\right) \tag{17}$$

under the constrains

$$c(x,u) \le W(I),\ \ u_i \in [0, u_{i\,max}]\, for\ i = 1{,}2{,}3.$$

2) short-therm strategy:

$$Maximize\ f_4(x,u) = U_4(x,u) - E_4(x,u) - M_4(x,u)$$

for all $t \in \left[t_0, t_f\right]$ under the constrains

$$c(x,u) \le W(I),\ \ u_i \in [0, u_{i\,max}]\, for\ i = 1{,}2{,}3.$$

Discretization of Eqs. (15 - 17) using Eqs. (6 - 8) and state equation (4) leads to

$$Minimize\ (-x_0^N) \tag{19}$$

subject to

$$x^{i+1} = x^i + hF(x^i,u^i), \qquad i = 0, \dots, N-1,$$

$$\lambda_i = \lambda^{i+1} + h\lambda^{i+1}F_{x^i}(x^i,u^i) + \mu^i c_{x^i}(x^i,u^i),$$
$$i = N-1, \dots, 0,$$

$$\lambda_0^i = -1, \qquad i = 0, \dots, N-1,$$

$$\lambda^N = (-1{,}0{,}0{,}0{,}0{,}0{,}0{,}0{,}0{,}0),$$

$$0 = h\lambda^{i+1}F_{u^i}(x^i,u^i) + \mu^i c_{u^i}(x^i,u^i),$$

where the vector function

$$F(x,u) = (-x_6, F_1(x,u), \dots, F_9(x,u))$$

is given by Eq. (16) and by right-hand side of Eq. (15).

The solution of optimal control with state and control constraints using adaptive critic neural network and NLP methods is displayed in Figs. 6 - 10 for different initial conditions $x(0)$ and different values of reduction coefficients b_2 and b_3. We used values of coefficients given in Table 1 for numerical calculation.

Table 1: Values of the Constants Used in the Model

$a_1 = 0.007$	$u_7 = 0.03$
$a_2 = 0.0182$	$u_9 = 0.2$
$a_3 = 0.5$	$g_1 = 0.14$
$a_4 = 0.67$	$g_2 = 1.5$
$a_5 = 1$	$g_3 = 2.0$
$a_6 = 1.39$	$g_4 = 1.5$
$a_7 = 0.66$	$g_5 = 0.8$
$a_8 = 0.67$	$g_6 = 0.4$
$K_1 = 19.3$	$g_7 = 0.2$
$K_2 = 8.17$	$g_8 = 0$
$K_3 = 71.28$	$g_9 = 0.15$
$K_4 = 3.4323$	$g_{10} = 0$
$K_5 = 0.62$	$g_{11} = 0.1$
$u_6 = 0.01$	$g_{12} = 0$

In the adaptive critic synthesis, the critic and adaptive network were selected such that they consist of nine and four subnetworks, respectively, each having 9-27-1 structure (i.e. nine neurons in the input layer, twenty-seven neurons in the hidden layer and one neuron in the output layer). The proposed adaptive critic neural network is able to meet the convergence tolerance values that we choose, which led to satisfactory simulation results. Simulation using MATLAB shows that there is a very good agreement between short-term and long-term strategy and proposed neural network is able to solve nonlinear optimal control problem with state and control constraints.

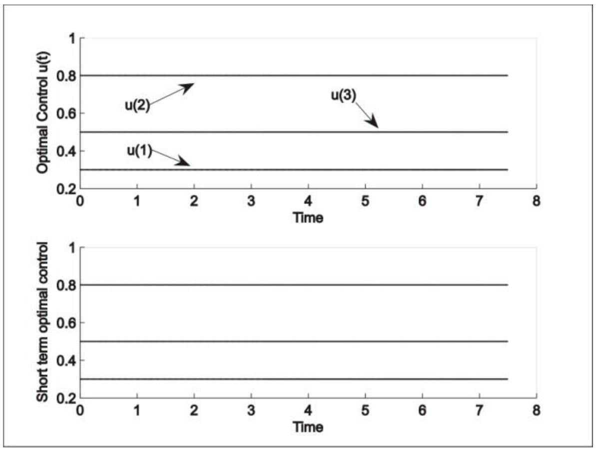


Figure 6: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{u}(t)$ for Initial Condition $x(0) = (0.01{,}0.01{,}0.02{,}0.001{,}0.04{,}0.001{,}0.001{,}0.07{,}0.01)$ and $b_2 < b_3$ or $b_3 < b_2$

The optimal strategy is the following. In the presence of high ammonium concentration, the uptake of DON and nitrate is stopped. If the concentration of ammonium drops below a certain limit value, phytoplankton and long-term strategy and proposed neural network is able to solve nonlinear optimal control problem with state and control constraints. The optimal strategy is the following. In the presence of high ammonium concentration, the uptake of DON and nitrate is stopped. If the concentration of ammonium drops below a certain limit value, phytoplankton starts to assimilate DON or nitrate dependently on values b_2, b_3.

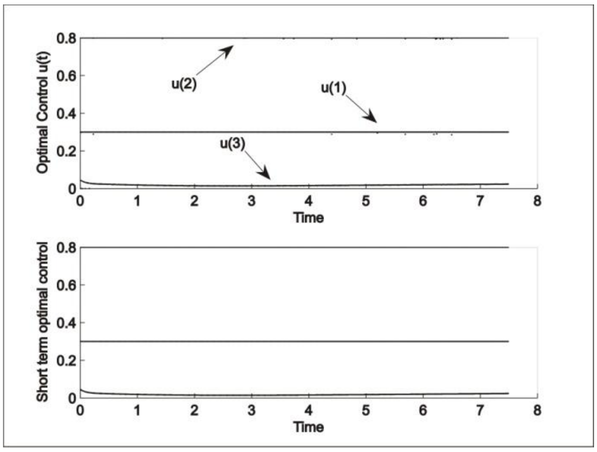


Figure 7: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{u}(t)$ for Initial Condition $x(0) = (0.1,0.1,0.2,0.8,0.4,0.001,0.001,0.7,1)$ and $b_2 < b_3$

If the concentration of all three forms of nitrogen are low all of them are assimilated by phytoplankton at the maximal possible rate, i.e. $\hat{u}_i(t) = u_{imax}$ for all $t \in [t_0, t_f]$ (Figure 6). Our results are quite similar to those obtained by (Kmet, 1996).

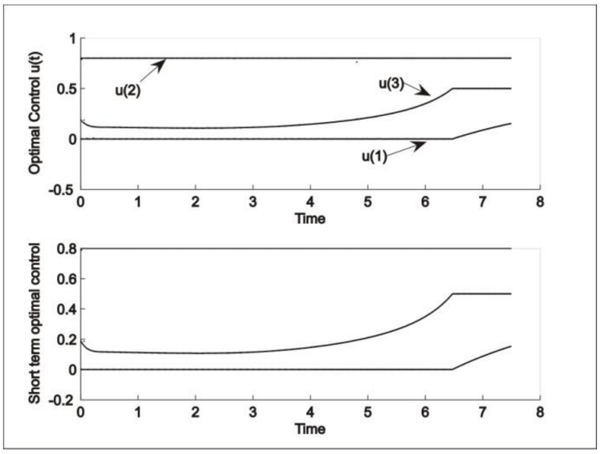


Figure 8: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{u}(t)$ for Initial Condition $x(0) = (0.1,0.1,0.2,0.8,0.4,0.001,0.001,0.7,1)$ and $b_3 < b_2$

6. CONCLUSION

A single network adaptive critic approach is presented for optimal control synthesis with control and state constraints. We have formulated, analysed and solved an optimal control problems related to optimal photosynthetic production and optimal biomass production, respectively. Using MATLAB, a simple simulation model based on adaptive critic neural network was constructed. Numerical simulations have shown that adaptive critic neural network is able to solve nonlinear optimal control problem with control and state constraints and fixed and free final time.

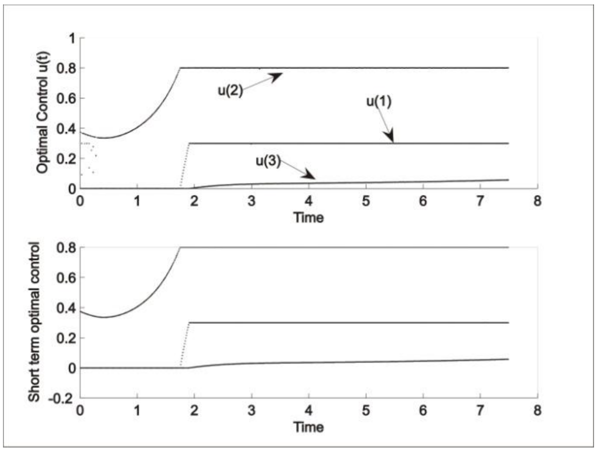


Figure 9: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{u}(t)$ for Initial Condition $x(0) = (0.1,0.1,0.2,0.8,0.4,0.5,0.6,0.7,1)$ and $b_2 < b_3$

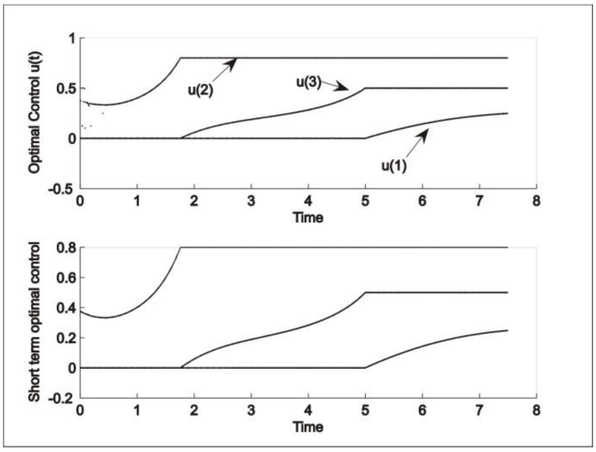


Figure 10: Adaptive Critic Neural Network Simulation of Optimal Control $\hat{u}(t)$ for Initial Condition $x(0) = (0.1,0.1,0.2,0.8,0.4,0.5,0.6,0.7,1)$ and $b_3 < b_2$

ACKNOWLEDGMENTS

Authors are grateful to the referee for his valuable suggestions. The paper was worked out as a part of solution of the scientific project number KEGA 04UJS-4/2011.

REFERENCES

Akin, E., 1979. *The Geometry of Population Genetics*. Springer-Verlag, Berlin.

Bryson, Jr., A. E., 1999. *Dynamic Optimization*. Addison-Wesley Longman Inc.

Buskens, Ch., Maurer, H., 2000. SQP-methods for solving optimal control problems with control and state constraints: adjoint variable, sensitivity analysis and real-time control. *Jour. Comp. Appl. Math*. 120: 85–108.

Eilers, P. H. C., Peeters, J. C. H., 1988. A model for relationship between light intensity and the rate of photosynthesis in phytoplankton. *Ecol. Modelling* 42: 199–215.

Eilers, P. H. C., Peeters, J. C. H., 1993. Dynamic behaviour of a model for photosynthesis and photoinhibition. *Ecol. Modelling* 69: 113–133.

Garcia-Camacho, F., Sanchez-Miron, A., Molina-Grima, E., Camacho-Rubio, F., Merchuck, J. C., 2012. A mechanistic model of photosynthesis in microalgal including photoacclimation dynamics. *J. Theor. Biol*. 304: 1–15.

Hornik, M., Stichcombe, M., White, H., 1989. Multilayer feed forward networks are universal approximators. *Neural Networks* 3: 256–366.

Kirk, D. E., 1989. *Optimal Control Theory: An Introduction.* Dover Publications, Inc. Mineola New York.

Kmet, T., 1996. Material recycling in a closed aquatic ecosystem. I. Nitrogen transformation cycle and preferential utilization of ammonium to nitrate by phytoplankton as an optimal control problem. *Bull. Math. Biol*. 58: 957 – 982.

Kmet, T., Straskraba, M., Mauersberger P., 1993. A mechanistic model of the adaptation of phytoplankton photosynthesis. *Bull. Math. Biol*. 55: 259–275.

Kmet, T., 2009. Neural network simulation of nitrogen transformation cycle. In: *ECMS 2009 - European Conference on Modelling and Simulation*. Madrid, June 9th - 12th, 2009: 352–358.

Kmet, T., 2011. Neural network simulation of optimal control problem with control and state constraints. In: *Honkela T. et al. Artificial Neural Networks and Machine Learning* – ICANN 2011 LNCS 6792: 261–268

Padhi, R., Unnikrishnan, N., Wang, X., and Balakrishnan, S. N., 2011. Adaptive-critic based optimal control synthesis for distributed parameter systems. *Automatica* 37: 1223–1234.

Padhi, R., Balakrishnan, S. N., Randoltph T., 2006. A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems. *Neural Networks* 19: 1648–1660.

Papacek S., Celikovsky, S., Rehak, B. and Stys, D., 2010. Experimental design for parameter estimation of two time-scale model of photosynthesis and photoinhibition in microalgae. *Math. Comp. Sim*. 80: 1302–1309.

Polak, E., 1997. *Optimization Algorithms and Consistent Approximation.* Springer Verlag, New York, Berlin Heidelberg.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., Mischenko, E. F., 1983. *The Mathematical Theory of Optimal Process.* Moscow, Nauka (in Russian).

Rumelhart, D. F., Hinton, G. E., Wiliams, R. J., 1987. Learning internal representation by error propagation. In: *Rumelhart, D. E., McClelland, D. E., and PDP Research Group. Parallel Distributed Processing* Vol. 1: Foundation. The MIT Press, Cambridge MA 1: 318–362.

Sandberg, E., W., 1998. Notes on uniform approximation of time-varying systems on finite time intervals. *IEEE Transactions on Circuits and Systems-1: Fundamental Theory and Applications*, 45(8).

Wu, X., Merchuk, J. C., 2001. A model integrating fluid dynamics in photosynthesis and photoinhibition. *Chem. Ing. Scien*. 56: 3527–3538.

AUTHORS BIOGRAPHY

Tibor Kmet is a full professor of applied informatics at Constantine the Philosopher University in Nitra, Slovakia. He received a PhD in applied mathematics from Faculty of Mathematics Physics and Informatics, Comenius University in Bratislava, Slovakia in 1987. His research interests are optimization of biological processes and neural network simulation of optimal control problem.

Maria Kmetova is an assoc. professor of mathematics at Constantine the Philosopher University in Nitra, Slovakia. She graduated from Comenius University in Bratislava. She received a PhD. degree in Geometry and Topology from Faculty of Mathematics Physics and Informatics, Comenius University in Bratislava in 2001. Her research interest focuses on the geometric design and visualization.

JACOBIAN BASED CONTROL OF WALKING ROBOT WITH COMPLIANT LEGS

M M Gor[(a)], P M Pathak[(b)], A K Samantaray[(c)], Jung-Ming Yang[(d)], Seong Woo Kwak[(e)]

[(a),(b)]Robotics & Control Lab, Indian Institute of Technology Roorkee, India
[(c)]Indian Institute of Technology, Kharagpur, India
[(d)]Catholic University of Daegu, South Korea
[(e)]Keimyung University, South Korea

[(a)]mehulmgor@rediffmail.com, [(b)]pushpfme@iitr.ernet.in, [(c)]samantaray@mech.iitkgp.ernet.in, [(d)]jmyang@cu.ac.kr,
[(e)]ksw@kmu.ac.kr

ABSTRACT

Jacobian based control of walking robot is difficult as compared to terrestrial manipulator as walking robot's each leg can be considered as a terrestrial manipulator. Moreover each leg is not always in motion as it has to pass through stand and motion phase depending upon the gait pattern. This paper proposed an approach where first gait pattern is assumed and for this gait pattern joint motions are assumed. For these joint motions the leg tip velocity is simulated. For pattern generation this leg tip velocity is assumed as the reference velocity and given to leg tip during the motion of that particular leg. The actual velocity of leg is calculated from the joint velocity of leg. The error in leg tip velocity is evaluated in PID controller, calculates the corrective voltage required for the joint motors. For the said analysis, the bond graph has been adopted as the modeling tool.

Keywords: Walking Robot, Workspace, Jacobian, Bond Graph

1. INTRODUCTION

Research in legged robots started almost four decades ago with an attempt to realize rigid legged locomotion. Researchers have used either analytical based approaches or biologically inspired approaches for the accomplishment of legged robot locomotion control.

Research pertaining to four legged robots has started almost two decades ago. But the research has not yet matured enough to produce commercially exploitable robotic system.

Legged robot have great problem of trajectory planning and stability which demands for good kinematics and dynamics model of the system.

Bernardi and Da Cruz (2007) presented a kinematical and dynamical analysis of a quadruped robot in which robots behavior was considered as parallel chain during pushing stage of platform and then serial chain during leg motion seeking for the new point of grasping.

Angeles (2007) showed quadruped walking which includes kinematic loops that open when a leg takes off and open chains that close when a leg touches the ground. This fact implies time-varying in a degree of freedom. (Pfeiffer 1995).

Kolter (2008) presented the controller consisting of a high-level planner that plans a set of footsteps across the terrain, a low-level planner that plans trajectories for the robot's feet and center of gravity and a low-level controller that tracks these desired trajectories using a set of closed-loop mechanisms. Many researchers (Zhao 2002; Aclan 2009) used PID controller to control bipedal robot locomotion in which predefined trajectories are taken as input.

Research in the direction of use of compliance in legged robots started with a motivation to realize faster and efficient locomotion. Research in this field can be broadly categorized as studies focused on understanding the inherent passive dynamics phenomena associated with the compliance (in the form of leg muscles and tendons, at joints in the form of cartilages, viscous fluid etc.) in animals and human beings and the implementation of the principles through analytical and experimental models in legged robots ranging from one, four to six legged robots. Krishnan (2010) shows simulation study of compliant legged quadruped in joint space. Yet from literature review, it can be found that there are several issues to be addressed so that useful legged robots can be made. Control of walking robot in workspace is one of them.

The main idea of this article is to study and simulate four legged walking robot in workspace, where reference tip velocity pattern is generated which will be compared with the actual tip velocity of leg by PID controller. Using jacobian and calculated error by PID controller, corrective voltage is found for the joints.

2. MODELING OF COMPLIANT LEG WALKING ROBOT

It is assumed that the walking robot is performing locomotion through the bounding gait. Modeling of the quadruped robot with compliant leg in sagittal plane consists of modeling of translational and angular dynamics of robot legs and body. Each leg of the robot has been modeled as an open chain manipulator comprising of one compliant link and one rigid link connected through revolute joints (Krishnan 2011). As shown in Figure 1, frame {A} is an inertial frame of reference and {V} is the body frame.

A coordinate frame is also attached to each link. The link frames are named by number according to the link to which they are attached i.e. frame {i} is rigidly attached to link i. The joint between links i and i+1 is numbered as i+1. Frame {0} and {1} are coincident, where frame {0} is body fixed frame at the hip joint. The rotational inertia of a body is defined about a frame fixed at the centre of gravity (CG) of the body. The axis of the CG frame is fixed along the principal directions of the body. r_F and r_R is the distance of {0_F} and {0_R} frame respectively from {V} frame, ϕ is body angular displacement, θ_{1F} and θ_{2F} are the joint angles of first and second joint of front leg, θ_{1R} and θ_{2R} are the joint angles of first and second joint of rear leg, l_1 and l_2 are the link lengths of the first and second links of front and rear leg.

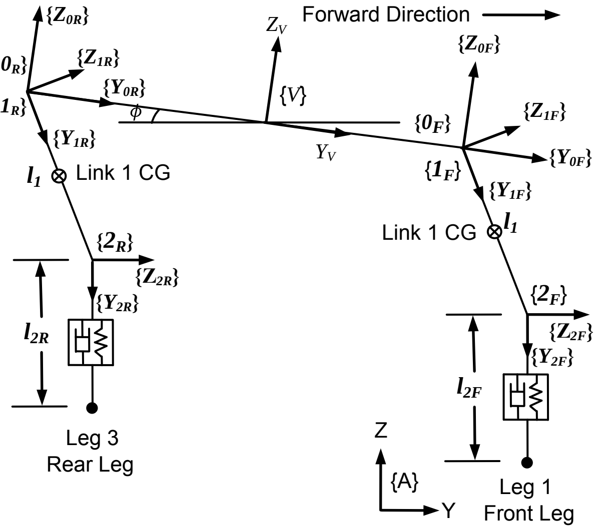


Figure 1: Compliant legged quadruped robot in Sagittal plane presentation

The position coordinate of the front leg tip can be expressed with respect to inertial frame as

$$\begin{bmatrix} Y_{Ftip} \\ Z_{Ftip} \end{bmatrix} = \begin{bmatrix} \left\{ \begin{matrix} Y_{CG} + r_F cos\phi \\ + l_1 \cos(\phi + \theta_{1F}) \\ + l_2 \cos(\phi + \theta_{1F} + \theta_{2F}) \end{matrix} \right\} \\ \left\{ \begin{matrix} Z_{CG} + r_F sin\phi \\ + l_1 \sin(\phi + \theta_{1F}) \\ + l_2 \sin(\phi + \theta_{1F} + \theta_{2F}) \end{matrix} \right\} \end{bmatrix} \tag{1}$$

The compliance in lower links of each leg is modeled on the basis of the following equations.

$$l_{2F}^2 = \left(Y_{Ftip} - Y_{2F}\right)^2 + \left(Z_{Ftip} - Z_{2F}\right)^2 \tag{2}$$

$$l_{2R}^2 = \left(Y_{Rtip} - Y_{2R}\right)^2 + \left(Z_{Rtip} - Z_{2R}\right)^2 \tag{3}$$

In above equations l_{2F} and l_{2R} are instantaneous lengths of link 2 of the front and rear leg respectively. Y_{Ftip} and Z_{Ftip} are the coordinate of toe of front leg and Y_{Rtip} and Z_{Rtip} are the coordinate of the rear leg. Y_{2F} and Z_{2F} are the coordinate of the frame {2_F}, and Y_{2R} and Z_{2R} are the coordinate of the frame {2_R}.

Taking derivative of equations (2) and (3)

$$\dot{l}_{2F} = \frac{(Y_{Ftip} - Y_{2F})}{l_{2F}} \left(\dot{Y}_{Ftip} - \dot{Y}_{2F}\right) + \frac{(Z_{Ftip} - Z_{2F})}{l_{2F}} \left(\dot{Z}_{Ftip} - \dot{Z}_{2F}\right) \tag{4}$$

$$\dot{l}_{2R} = \frac{(Y_{Rtip} - Y_{2R})}{l_{2R}} \left(\dot{Y}_{Rtip} - \dot{Y}_{2R}\right) + \frac{(Z_{Rtip} - Z_{2R})}{l_{2R}} \left(\dot{Z}_{Rtip} - \dot{Z}_{2R}\right) \tag{5}$$

Considering compliance in the lower links of the leg the velocity of the front leg tip can be given as

$$\begin{bmatrix} \dot{Y}_{Ftip} \\ \dot{Z}_{Ftip} \end{bmatrix} = \begin{bmatrix} \left\{ \begin{matrix} \dot{Y}_{CG} - (r_F \sin\phi)\dot{\phi} \\ -l_1 \sin(\phi + \theta_{1F})(\dot{\phi} + \dot{\theta_{1F}}) \\ - l_2 \sin(\phi + \theta_{1F} + \theta_{2F})(\dot{\phi} + \dot{\theta_{1F}} + \dot{\theta}_{2F}) \\ + \dot{l_{2F}} \cos(\phi + \theta_{1F} + \theta_{2F}) \end{matrix} \right\} \\ \left\{ \begin{matrix} \dot{Z}_{CG} + (r_F cos\phi)\dot{\phi} \\ +l_1 \cos(\phi + \theta_{1F})(\dot{\phi} + \dot{\theta_{1F}}) \\ + l_2 \cos(\phi + \theta_{1F} + \theta_{2F}) \left(\dot{\phi} + \dot{\theta_{1F}} + \dot{\theta}_{2F}\right) \\ + \dot{l_{2F}} \sin(\phi + \theta_{1F} + \theta_{2F}) \end{matrix} \right\} \end{bmatrix} \tag{6}$$

Similarly Y_{Rtip} and Z_{Rtip} velocity equations for rear leg tip can be derived as

$$\begin{bmatrix} \dot{Y}_{Rtip} \\ \dot{Z}_{Rtip} \end{bmatrix} = \begin{bmatrix} \left\{ \begin{matrix} \dot{Y}_{CG} - r_R sin(\pi + \phi)\dot{\phi} \\ -l_1 \sin(\pi + \phi + \theta_{1R})(\dot{\phi} + \dot{\theta_{1R}}) \\ - l_2 \sin(\pi + \phi + \theta_{1R} + \theta_{2R})(\dot{\phi} + \dot{\theta_{1R}} + \dot{\theta}_{2R}) \\ + \dot{l_{2R}} \cos(\pi + \phi + \theta_{1R} + \theta_{2R}) \end{matrix} \right\} \\ \left\{ \begin{matrix} \dot{Z}_{CG} + r_R cos(\pi + \phi)\dot{\phi} \\ +l_1 \cos(\pi + \phi + \theta_{1R})(\dot{\phi} + \dot{\theta_{1R}}) + \\ l_2 \cos(\pi + \phi + \theta_{1R} + \theta_{2R}) \left(\dot{\phi} + \dot{\theta_{1R}} + \dot{\theta}_{2R}\right) \\ + \dot{l_{2R}} \sin(\pi + \phi + \theta_{1R} + \theta_{2R}) \end{matrix} \right\} \end{bmatrix} \tag{7}$$

The bounding gait selected for locomotion enables us to study the locomotion dynamics in a sagittal plane. Using equations (6) and (7) bond graph model of four legged walking robot in sagittal plane can be drawn which is shown in Figure 2.

A soft pad is used to avoid differential causality. Soft pads are artificial compliances/lumped flexibilities (Ghosh 1991, Pathak 2005) that can be used in bond graph. In particular a soft pad is used instead of a pad in order to avoid algebraic loop while deriving equations. In bondgraph model, all bonds are not numbered to bring clarity in the figure. Transformer moduli for finding velocities at various points is shown in Table 1.

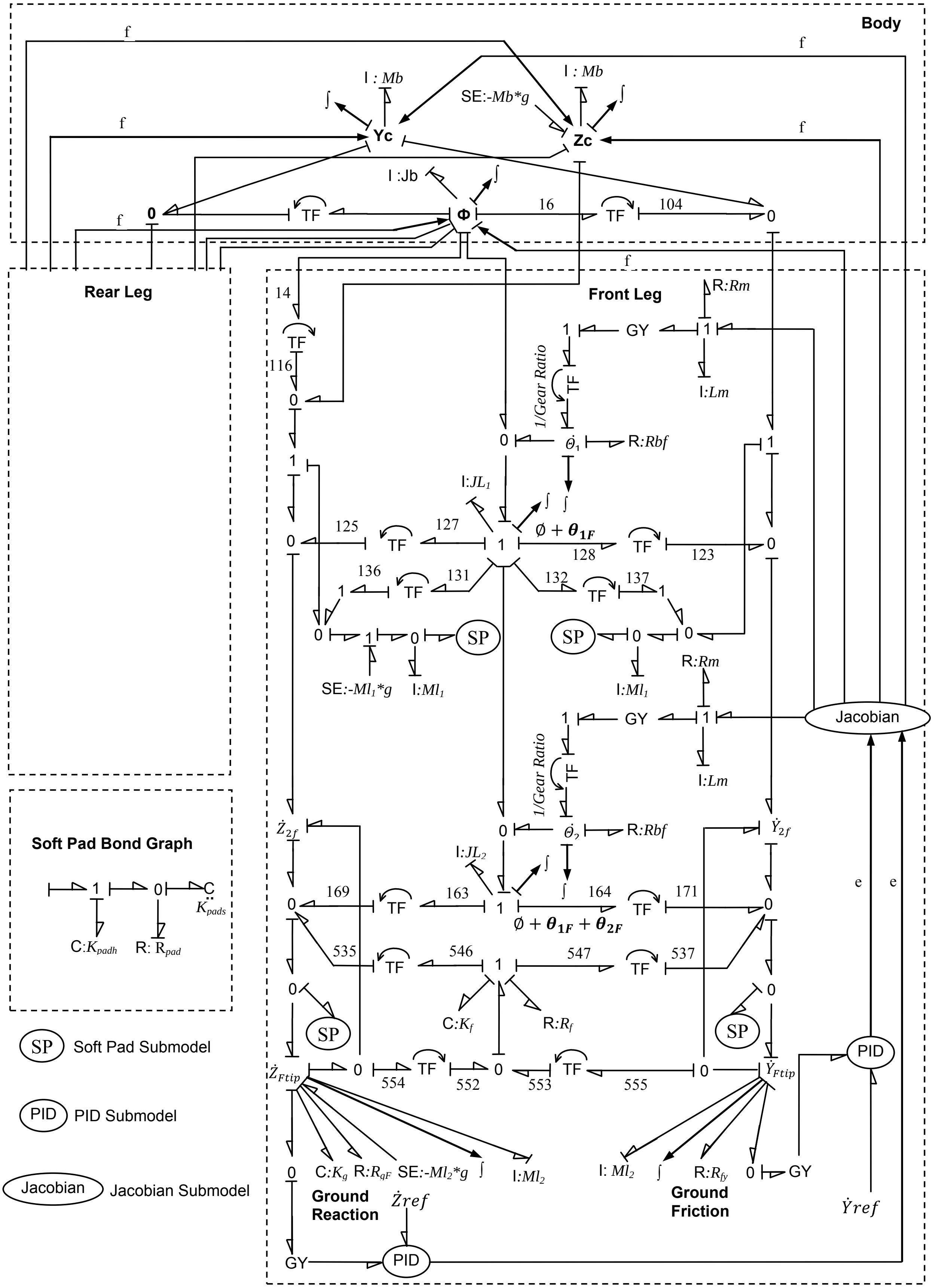


Figure 2: Bond Graph Model of Four Legged Walking Robot

Table 1: Transformer moduli for finding velocities at various points

	Y Direction	Z Direction
Platform	TF16-104=$-(r_F \sin\phi)$	TF14-16=$(r_F cos\phi)$
First Link Tip	TF128-123=$-l_1 \sin(\phi + \theta_{1F})$	TF127-125=$l_1 \cos(\phi + \theta_{1F})$
First Link CM	TF132-137=$-0.5 * l_1 \sin(\phi + \theta_{1F})$	TF131-136=$0.5 * l_1 \cos(\phi + \theta_{1F})$
Second Link Tip	TF164-171=$-\ l_2 \sin(\phi + \theta_{1F} + \theta_{2F})$ TF547-537=$\cos(\phi + \theta_{1F} + \theta_{2F})$ TF555-553=$\frac{(Y_{Ftip} - Y_{2F})}{l_{2F}}$	TF163-169=$l_2 \cos(\phi + \theta_{1F} + \theta_{2F})$ TF546-535=$\sin(\phi + \theta_{1F} + \theta_{2F})$ TF554-552=$\frac{(Z_{Ftip} - Z_{2F})}{l_{2F}}$

3. CONTROL IN WORKSPACE

3.1. Reference Velocity Pattern Generator

When compliant legged walking robot model works in joint space, it is found that foot tip velocity follows nearby sinusoidal curve. So here reference velocity pattern is generated using the same concept. It is assumed that robot will complete its walking cycle in 2.8 s and thus here all data are presented for 10 cycles i.e. for 28 s. Phase difference between front leg and rear leg reference velocity is kept as 0.2 s. Figure 3 and Figure 4 shows reference velocity of front leg and rear leg respectively.

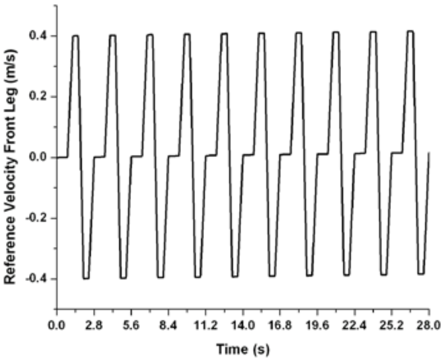


Figure 3: Reference Velocity of Front Leg

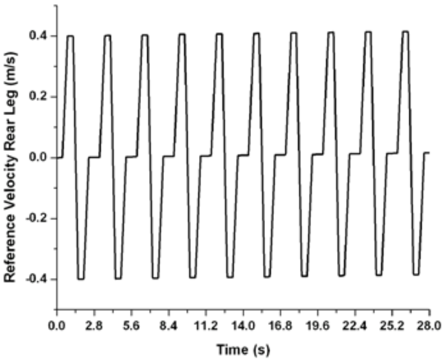


Figure 4: Reference Velocity of Rear Leg

3.2. PID Controller

The purpose of the use of PID controller is to make the leg movement in a desired way in terms of the reference pattern. PID controller has no offset error and reduces the tendency for oscillations. Controller is presented using bond-graph. The advantage of working in the bond graph domain is that a clear physical interpretation can be given to each controller coefficient. The PID controller discussed in the standard textbook (like Bolton 2007) gives a control signal

$$I_{out} = K_{\mathrm{p}} e + K_I \int e\, dt + K_V \frac{de}{dt} + I_o \tag{8}$$

Where I_{out} is the output from the controller when there is an error e which is changing with time t, I_o is the set point output when there is no error, K_p is the proportionality constant, K_I is the integral constant and K_V is the derivative constant. Bond graph model of such controller can be represented as shown by Mukherjee (2006). Figure 5 shows the submodel of PID controller developed for this work.

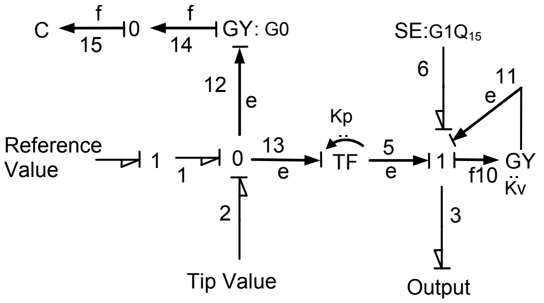


Figure 5: Bondgraph Submodel of PID Controller

3.3. Jacobian

In the bond graph modeling of the walking robot working in workspace with a PID controller, one needs to evaluate the jacobian of the manipulator. Jacobian is a linear mapping from velocities in joint space to Cartesian space. The inverse problem, where the joint velocities are to be determined for a given tip velocity, is practical importance and requires the inverse of the Jacobian.

The kinematic relations for the front leg tip displacements Y_{tip} and Z_{tip} in the Y and Z directions given with respect to inertial reference frame {A} is shown in equation (1). For jacobian model assuming link2 as rigid and differentiating equation (1), front leg tip velocity component can be derived as

$$\begin{bmatrix}\dot{Y}_{Ftip}\\ \dot{Z}_{Ftip}\end{bmatrix}= \begin{bmatrix}\left\{\begin{matrix}\dot{Y}_{CG}-(r_F\sin\phi)\dot{\phi}\\ -l_1\sin(\phi+\theta_{1F})(\dot{\phi}+\dot{\theta_{1F}})\\ -\,l_2\sin(\phi+\theta_{1F}+\theta_{2F})(\dot{\phi}+\dot{\theta_{1F}}+\dot{\theta}_{2F})\end{matrix}\right\}\\ \left\{\begin{matrix}\dot{Z}_{CG}+(r_F cos\phi)\dot{\phi}\\ +l_1\cos(\phi+\theta_{1F})(\dot{\phi}+\dot{\theta_{1F}})\\ +\,l_2\cos(\phi+\theta_{1F}+\theta_{2F})\,(\dot{\phi}+\dot{\theta_{1F}}+\dot{\theta}_{2F})\end{matrix}\right\}\end{bmatrix} \tag{9}$$

Similarly, velocity component of rear leg can be derived as

$$\begin{bmatrix}\dot{Y}_{Rtip}\\ \dot{Z}_{Rtip}\end{bmatrix}= \begin{bmatrix}\left\{\begin{matrix}\dot{Y}_{CG}-r_R sin(\pi+\phi)\dot{\phi}\\ -l_1\sin(\pi+\phi+\theta_{1R})(\dot{\phi}+\dot{\theta_{1R}})\\ -\,l_2\sin(\pi+\phi+\theta_{1R}+\theta_{2R})(\dot{\phi}+\dot{\theta_{1R}}+\dot{\theta}_{2R})\end{matrix}\right\}\\ \left\{\begin{matrix}\dot{Z}_{CG}+r_R cos(\pi+\phi)\dot{\phi}\\ +l_1\cos(\pi+\phi+\theta_{1R})(\dot{\phi}+\dot{\theta_{1R}})+\\ l_2\cos(\pi+\phi+\theta_{1R}+\theta_{2R})\,(\dot{\phi}+\dot{\theta_{1R}}+\dot{\theta}_{2R})\end{matrix}\right\}\end{bmatrix} \tag{10}$$

Bondgraph presentation of the jacobian for front leg is shown in Figure 6. Similarly jacobian for rear leg is also developed.

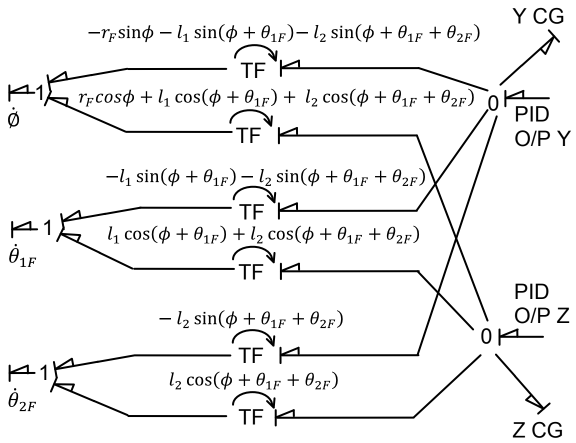


Figure 6: Bondgraph Submodel of Front Leg Jacobian

PID controller discussed in section 3.2 is used for trajectory control of walking robot. Generated reference flow velocity pattern is provided to the controller. The tip velocity of the leg is supplied to the controller. This input is compared with the reference flow input pattern and the error is processed in the controller, which in turn provides the corrective voltage as output of controller. This corrective voltage is fed to the Jacobian block, which evaluates the required voltage at the joints which will be supplied to the joints.

4. SIMULATION & RESULTS

To validate control strategy discussed above, the model has been simulated. Input parameters are listed in Table 2. Each cycle takes 2.8 s. Simulation is carried out for 10 complete cycle which takes 28 s. Animation result is shown in Figure 7. Figure 8 and 9 shows body CG displacement in Y and Z direction respectively. It can be seen that body propagates in positive Y direction. Figure 10 and 11 shows front leg tip movement in Y and Z direction respectively. Figure 12 and 13 shows rear leg tip movement in Y and Z direction respectively.

Table 2: Input parameters

Parameters	**Value**
Arm Parameters	
Front & Rear Leg Link 1 Length(l1)	0.065 m
Front & Rear Leg Link 2 Length(l2)	0.05 m
Front & Rear Leg Link 1 Mass(Ml1)	0.075 kg
Front & Rear Leg Link 2 Mass(Ml2)	0.065 kg
Location of Leg from body CG(rF/rR)	0.042 m
Compliant link 2 stiffness(Kf)	400 N/m
Compliant link 2 damping(Rf)	4 Ns/m
Inertia of Link 1 (Jl1)	0.00004 kgm^2
Inertia of Link 2 (Jl2)	0.00003 kgm^2
Common Parameter	
Mass of body(Mb)	0.28 kg
Inertia of body(Jb)	0.002 kgm^2
Ground Frictional Resistance(Rfy)	10 Ns/m
Ground Stiffness(Kg)	10000 N/m
Ground Damping(Rg)	10 Ns/m
Controller Parameter	
Proportional Gain of controller(Kp)	1.2
Derivative Gain of Controller(Kv)	-0.01
Integral Gain of Controller(G1)	5
Pad Parameter	
Stiffness of Spring (Kpads)	2000 N/m
Stiffness of Spring(Kpadh)	900 N/m
Damping Resistance(Rpad)	1 Ns/m
Actuator Parameter	
Motor Constant (Km)	0.01Nm/A
Motor Resistance(Rm)	0.1 Ohms
Motor Inductance(Lm)	0.001 H
Bearing Resistance(Rbf)	0.1 Nm/r/s
Gear Ratio	230

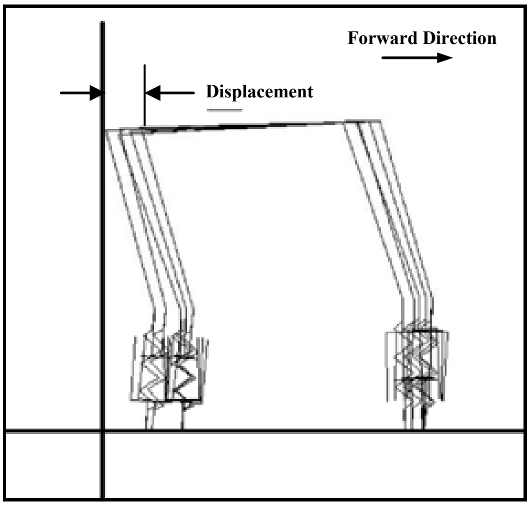


Figure 7: Animation result of Walking Robot

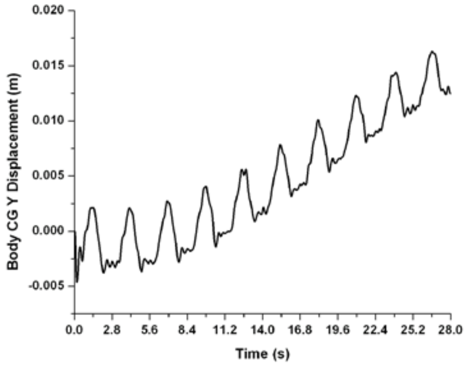


Figure 8: Body CG Y Displacement

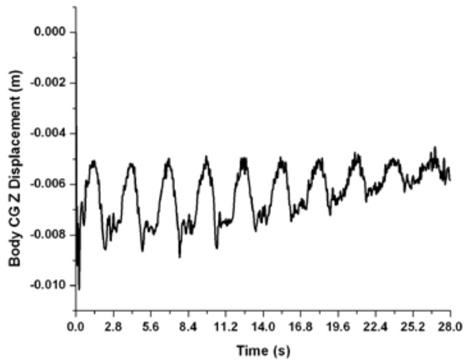


Figure 9: Body CG Z Displacement

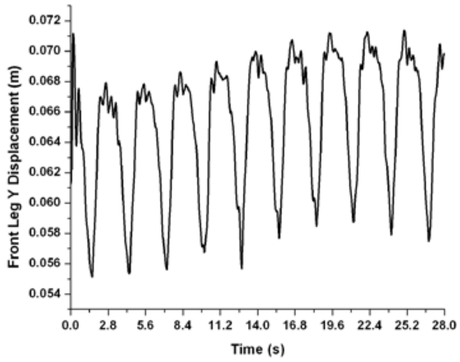


Figure 10: Front Leg Tip Y Displacement

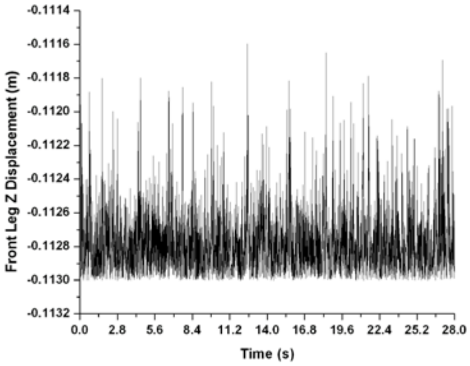


Figure 11: Front Leg Tip Z Displacement

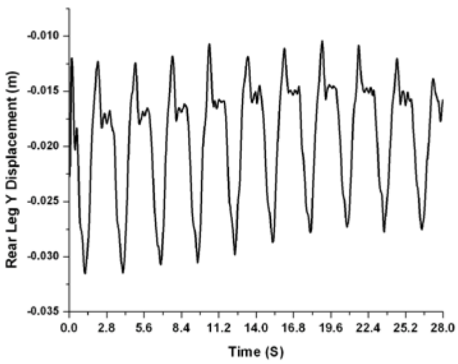


Figure 12: Rear Leg Tip Y Displacement

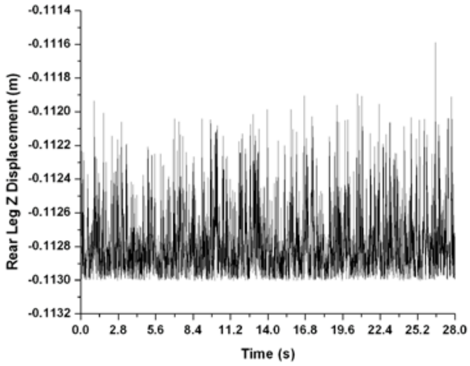


Figure 13: Rear Leg Tip Z Displacement

5. CONCLUSION

This work presents a method for trajectory control of a walking robot in workspace. The concept of using reference velocity pattern to evaluate Jacobian through PID controller is shown to be successful. The strategy can be extended for a control of walking robot in three dimensional models also.

REFERENCES

Aclan M.C., Ramos M.C., 2009, “Bipedal Robot Locomotion Using Multivariable Control”, *TENCON 2009 - 2009 IEEE Region 10 Conference*, pp. 1-6.

Angeles, J., 2007, *“Fundamentals of Robotic Mechanical Systems. Theory, Methods, and Algorithms”*, Springer.

Bernardi, R. & Da Cruz, J. J., 2007, “Kamanbaré: A tree-climbing biomimetic robotic platform for environmental research”, *International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, pp. 478-484.

Bolton W., 2007, *“Mechatronics: Electronics Control Systems in Mechanical and Electrical Engineering”*, South Asia, Pearson Education.

Ghosh A. K., Mukherjee A., Faruqi M. A, 1991, “Computation of Driving Efforts for Mechanisms and Robots Using Bond Graphs”, *Transactions of* *the ASME Journal of Dynamic, Systems, Measurement, and Control*, Vol. 113, pp. 744-748.

Kolter J. Z., Rodgers M. P., Ng A. Y., 2008, "A control architecture for quadruped locomotion over rough terrain," *in Proceedings IEEE Int. Conference on Robotics and Automation*, pp. 811-818.

Krishnan V. L, Pathak P. M., Sardana L., Jain S. C., 2010, "Simulation And Experimental Studies on Walking Robot with Flexible Legs", *4th International Conference on Integrated Modeling and Analysis in Applied Control and Automation (IMAACA 2010)*, Fes, Morocco, pp. 11-19.

Krishnan V. L., 2011, "*Modeling and Control of Quadruped Robot with Flexible Legs",* Ph.D thesis, IIT Roorkee, Roorkee, India.

Mukharjee A., Karmakar R., Samantaray A. K., 2006, "*Bond Graph in Modeling, Simulation and Fault Identification*", Indian, I. K. International.

Pathak P. M., Mukherjee A., Dasgupta A., 2005, "Impedance Control of Space Robots using Passive Degrees of Freedom in Controller Domain", *Transactions of ASME, Journal of Dynamic Systems, Measurement, and Control*., Vol. 127, pp. 564-578.

Pfeiffer F., Eltze J., Weidemann H. J., 1995, "The tum walking machine", *Intelligent Automation and Soft Computing. An International Journal* 1: pp. 307–323.

Zhao M., Liu L., Wanj J., Chen K., Zhao., Xu K., 2002, "Control system design of THBIP-I humanoid robot", *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 2253-2258.

ACKNOWLEDGEMENT

This work has been funded by Department of Science and Technology, Government of India (vide Grant No. INT/Korea/P-13) and National Research Foundation of Korea grant funded by the Korea government (MEST) (No. 2012-0006297), under Indo-Korea Joint Research in Science and Technology.

AUTHORS BIOGRAPHY

M M Gor received the B. E. -Production degree in 2000 and M. E. Mechanical-Machine Design degree in 2008, both from B. V. Mahavidyalaya, V. V. Nagar, India. He has joined the Department of Mechanical Engineering at G H Patel College of Engineering & Technology, V. V. Nagar, India in 2010.

Currently he is a Ph. D. candidate at the Department of Mechanical and Industrial Engineering, Indian Institute of Technology, Roorkee, India

P M Pathak received the B. Tech. degree in 1988 from NIT, Calicut, India and M. Tech. degree in 1998 from IIT, Kanpur, India both in Mechanical Engineering. He received Ph. D. degree in 2005 from IIT, Kharagpur, India.

Currently he is an Assistant Professor in the Department of Mechanical and Industrial Engineering, Indian Institute of Technology, Roorkee, India. His areas of interest are space robotics, walking robots, dynamics and control.

A K Samantaray received the Ph.D degrees in Mechanical Engineering, from Indian Institute of Technology, Kharagpur.

Currently he is an Associate Professor in Department of Mechanical Engineering, Indian Institute of Technology, Kharagpur, India. His areas of interest are Systems and Control, Nonlinear Mechanics, Vehicle system dynamics and Rotor dynamics.

Jung-Min Yang received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science & Technology (KAIST), Korea, in 1993, 1995, and 1999, respectively.

Currently he is an Associate Professor in Electrical Engineering Department of Catholic University of Daegu, Korea. His research interests are in gait study of walking machines and control of asynchronous sequential machines.

Seong Woo Kwak received his B.S., M.S., and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science & Technology (KASIT), Korea, in 1993, 1995 and 2000, respectively.

Currently he is an Associate Professor in Department of Electrical Engineering, Keimyung University, Korea. His research interests are in control of asynchronous sequential machines, space-borne digital electronics, and fault-tolerance in real time systems.

ARMA MODEL BASED GPC

Yuchen Zhao[a], Aníbal Zanini[b]

[a]Department of Electronics, College of Engineering, Buenos Aires University
[b] Department of Electronics, College of Engineering, Buenos Aires University

[a]gustavozhao@gmail.com, [b]azanini@fi.uba.ar

ABSTRACT

In this paper an Auto-Regressive Moving Average (ARMA) model based Generalized Predictive Control (GPC) is presented. The controller derived from this method will automatically contain an integrator. The control law is obtained by minimizing a quadratic objective function. An analytical solution can be found in the absence of constraints. The presented method can deal with stable, unstable and non-minimum phase processes. A concept of suboptimal control is introduced in order to reduce calculating burden and refine dynamic aspects of a controlled system in case of need. It is straightforward to achieve the double integral action which is required in some industrial processes. Furthermore, it is very easy to incorporate the terms *T* and *S* if it is viewed as a classical RST controller. For this purpose, a relationship between GPC and RST controllers is also presented. Some simple examples and numerical analyses of particular cases are given.

Keywords: generalized predictive control, model predictive control, auto-regressive moving average, integral action

1. INTRODUCTION

As well known the MPC presents many advantages such as the availability to control a process with long delay times, the feasibility to handle easily the multivariable case and the relative simplicity to deal with constrained control.

Predictive control algorithms are based on an assumed model of the process and on an assumed scene for the future control signals. From the end of 1970's, a lot of predictive control algorithms were presented. Model Algorithmic Control (MAC) (Richalet, Rault, Testud, and Papon 1976, 1978), Dynamic Matrix Control (DMC) (Cutler and Ramaker 1980). Predictive Functional Control (PFC) (Richalet, Abu el Ata-Doss, Arber, Kuntze, Jacubash, and Schill 1987) and Extended Horizon Adaptive Control (EHAC) (Ydstie 1984) are ones of them. The GPC method was proposed by Clarke *et al.* (Clarke, Mohtadi, and Tuffs 1987). It has been studied intensively in the industrial and academic circles and has been successfully implemented in numerous industrial applications (Clarke 1988). It is based on the CARIMA (Controlled Auto-Regressive and Integrated Moving Average) model. The expectation of a quadric function measuring the discrepancy between the predicted output and the predicted reference over prediction horizon plus another quadric function measuring control efforts is the index to be optimized. The optimization results a sequence of control signal. But only the first one is applied to the process to be controlled. A new sequence is calculated during the next sampling interval after a new measurement of output is obtained. This is called receding-horizon control. Based on the same concepts, the ARMA model based GPC is proposed in this paper as an alternative or a complement to enrich the goodness of the GPC.

2. PROCESS MODEL

Assume that the process dynamics are characterized by the local-linearized model

$$\begin{aligned} x(k) &= \frac{B(z^{-1})}{A(z^{-1})}[u(k-d)+v(k-d)] \\ y(k) &= x(k)+e(k) \end{aligned} \tag{1}$$

where $A(z^{-1})$ and $B(z^{-1})$ are polynomials in the backward shift operator without any common factors, u is the control signal, v is a disturbance acting on the input of the process and e is the measurement noise acting on the process output.

However, when taking into account neither the disturbance nor the measurement noise, the model (1) becomes

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})}u(k-d) \tag{2}$$

3. OUTPUT PREDICTION

One of the basic ideas of the predictive control is to rewrite the model of the process in order of obtaining an explicit expression for the output at a future time.

3.1. Stable Process

Consider the model (2)

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k-d) \tag{3}$$

h steps ahead,

$$y(k+h) = \frac{B(z^{-1})}{A(z^{-1})} u(k+h-d) \tag{4}$$

Subtracting Eq. (3) from Eq. (4), we obtain

$$\begin{aligned} y(k+h) - y(k) &= \frac{B(z^{-1})}{A(z^{-1})} u(k+h-d) - \frac{B(z^{-1})}{A(z^{-1})} u(k-d) \\ &= F_h(z^{-1}) u(k+h-d) + \frac{G_h(z^{-1}) - q^{1-d} B(z^{-1})}{A(z^{-1})} u(k-1) \end{aligned}$$

where d is the dead time, F_h y $z^{-(h-d+1)} G_h$, whose degree is $h-d$ y $n-1$ respectively, are the quotient and remainder of the division B/A, namely

$$\frac{B(z^{-1})}{A(z^{-1})} = F_h(z^{-1}) + z^{-(h-d+1)} \frac{G_h(z^{-1})}{A(z^{-1})} \tag{5}$$

where

$$F_h(z^{-1}) = f_0 + f_1 z^{-1} + \cdots + f_{h-d} z^{-(h-d)}$$

Define

$$H_h(z^{-1}) = G_h(z^{-1}) - z^{1-d} B(z^{-1}) \tag{6}$$

So

$$y(k+h) = F_h u(k+h-d) + \frac{H_h}{A} u(k-1) + y(k) \tag{7}$$

where the first term of the right hand depends on $u(k), \cdots, u(k+h-d)$, the second term can be treated as a filter and a constant error is assumed over the predication horizon. That is, from Eq. (1) and (7),

$$\varepsilon(k) = \frac{B}{A} v(k) + e(k) \qquad \forall h$$

For simplicity, the last two terms of the right hand side of Eq. (7) are denoted as $\tilde{y}_h(k)$, which is called free response and the first term is called forced response for depending on the future control signals if $h > d$. Thus

$$y(k+h) = F_h(z^{-1}) u(k+h-d) + \tilde{y}_h(k) \tag{8}$$

Using the last equation for $h = 1, 2, \cdots, N$, we get

$$\begin{aligned} & y(k+1) = F_1(z^{-1}) u(k+1-d) + \tilde{y}_1(k) \\ & y(k+2) = F_2(z^{-1}) u(k+2-d) + \tilde{y}_2(k) \\ & \vdots \\ & y(k+N) = F_N(z^{-1}) u(k+N-d) + \tilde{y}_N(k) \end{aligned} \tag{9}$$

Introduce vectors

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} \hat{y}(k+1 \mid k) & \hat{y}(k+2 \mid k) & \cdots & \hat{y}(k+N \mid k) \end{bmatrix}^T \\ \mathbf{u} &= \begin{bmatrix} u(k+1 \mid k) & u(k+2 \mid k) & \cdots & u(k+N \mid k) \end{bmatrix}^T \\ \tilde{\mathbf{y}} &= \begin{bmatrix} \tilde{y}_1(k) & \tilde{y}_2(k) & \cdots & \tilde{y}_N(k) \end{bmatrix}^T \end{aligned}$$

where $\hat{y}(k+h \mid k)$ denotes the h step ahead output prediction made at the instant k and the notation $u(k+h \mid k)$ denotes the control signal for instant $k+h$ calculated at instant k. Then Eq. (9) can be expressed as

$$\mathbf{y} = \mathbf{F}\mathbf{u} + \tilde{\mathbf{y}} \tag{10}$$

where

$$\mathbf{F} = \begin{bmatrix} f_0 & 0 & \cdots & 0 \\ f_1 & f_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ f_{N-1} & f_{N-2} & \cdots & f_0 \end{bmatrix} \tag{11}$$

is a lower triangular matrix for causality.

As it will be known that the controller based on the predictor (10) incorporate automatically an integrator. In case a double integrator is required in order to follow ramp type reference, the following predictor is needed to reach the purpose.

Multiplying both sides of Eq. (2) with the term $\Delta = 1 - z^{-1}$, which is a difference operator, we get

$$\Delta y(k) = \frac{B(z^{-1})}{A(z^{-1})} \Delta u(k-d)$$

By a similar procedure, we will acquire a predictor for double integral action.

$$\mathbf{y} = \mathbf{F}\Delta\mathbf{u} + \tilde{\mathbf{y}} \tag{12}$$

3.2. Integrating Process

An integrating process can be modeled as

$$y(k) = \frac{B(z^{-1})}{\Delta A(z^{-1})} u(k-d)$$

where d is dead time, $\Delta = 1 - z^{-1}$ is the difference operator as mentioned above and A is stable and well damped. The model can be rewritten as

$$\Delta y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k-d)$$

From that we can get

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k-d) - y(k-1) \quad (13)$$

$$y(k+h) = \frac{B(z^{-1})}{A(z^{-1})} u(k+h-d) - y(k+h-1) \quad (14)$$

Subtracting Eq. (13) from Eq. (14), we get

$$\begin{aligned} y(k+h) &= \frac{B}{A} u(k+h-d) - \frac{B}{A} u(k-d) \\ &\quad + y(k) - \frac{B(z^{-1})}{A(z^{-1})} u(k-d) + y(k-1) \end{aligned}$$

where it is assumed that the sum

$$y(k) - \frac{B(z^{-1})}{A(z^{-1})} u(k-d) + y(k-1)$$

may be different from zero.

Using the identity (5) and the definition (6), iterating the last equation for $h = 1,2,\cdots N$, we get

$$\begin{aligned} y(k+1) &= F_1 u(k+h-d) + \frac{H_1}{A} u(k-1) + (2 - z^{-1}) y(k) \\ y(k+2) &= F_2 u(k+2-d) + F_1 u(k+1-d) + \frac{H_1}{A} u(k-1) \\ &\quad + (3 - 2z^{-1}) y(k) \\ &\vdots \\ y(k+N) &= \sum_{i=1}^{N} F_i u(k+i-d) + \sum_{i=1}^{N} \frac{H_i}{A} + (N+1-Nz^{-1}) y(k) \end{aligned}$$

In vector form

$$\mathbf{y} = \mathbf{F}\mathbf{u} + \mathbf{H} u(k-1) + \mathbf{M} y(k) \quad (15)$$

where the vectors $\mathbf{y}$ and $\mathbf{u}$ are the same as the aforesaid and

$$\mathbf{F} = \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ F_1 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ F_1 & F_2 & \cdots & F_N \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} \frac{H_1}{A} & \frac{H_1 + H_2}{A} & \cdots & \sum_{i=1}^{N} \frac{H_i}{A} \end{bmatrix}^T$$

$$\mathbf{M} = [2 - z^{-1} \quad 3 - 2z^{-1} \quad \cdots \quad N+1-Nz^{-1}]^T$$

Define

$$\tilde{\mathbf{y}} = \mathbf{H} u(k-1) + \mathbf{M} y(k)$$

So Eq. (15) can be written as Eq. (10), and the matrix **F** can be otherwise expressed as

$$\mathbf{F} = \begin{bmatrix} f_0 & 0 & \cdots & 0 \\ f_0 + f_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{N-1} f_i & \sum_{i=0}^{N-2} f_i & \cdots & f_0 \end{bmatrix}$$

where $f_i, i = 0,1,\cdots,N-1$, are the coefficients of the polynomial $F_h(z^{-1})$ as aforementioned.

The controlled system will have double integral action for the process itself has the integrating effect.

3.3. Unstable process

Rewriting the model (2) as

$$A(z^{-1}) y(k) = B(z^{-1}) u(k-d) \quad (16)$$

with A monic and denoting $A(z^{-1})$ as

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n} = 1 + A_1 z^{-1}$$

where n is the order of the process, Eq. (16) can thus be rewritten as

$$y(k) = Bu(k-d) - A_1 y(k-1) \quad (17)$$

and h steps ahead

$$y(k+h) = Bu(k+h-d) - A_1 y(k+h-1) \quad (18)$$

Subtracting Eq. (17) from Eq. (18) and assuming the sum $y(k) - Bu(k-d) + A_1 y(k-1)$ could not be zero, we get

$$\begin{aligned} y(k+h) &= Bu(k+h-d) - Bu(k-d) + y(k) \\ &\quad + A_1 y(k-1) - A_1 y(k+h-1) \end{aligned} \quad (19)$$

Introduce the identity

$$A_1^i B = F_h^i + z^{-(h-d+1)} G_h^i \qquad h = 1,2,\cdots,N \text{ and } i = 0,1,\cdots,h-1$$

where F_h^i contains the first $h-d+1$ terms of $A_1^i B$ and the rest is the second term of the right hand side of the identity. Notice that in the terms F_h^i and G_h^i, i is no more than a notation which indicates the relation between them and $A_1^i B$. The polynomial $[A_1(z^{-1})]^i B(z^{-1})$, $i = 0,1,2,\cdots$. is of degree $(i+1)(n-1)$ and expressed in the backward shift operator.

Iterating Eq. (19) for $h = 1,2,\cdots,N$, we can get

$$\begin{aligned}
y(k+1) &= F_1^0 u(k+1-d) + (G_1^0 - z^{1-d}B)u(k-1) \\
&\quad +(1-A_1)y(k) + A_1 y(k-1) \\
y(k+2) &= -F_1^1 u(k+1-d) + F_2^0 u(k+2-d) \\
&\quad +[-G_1^1 + G_2^0 - z^{1-d}B(1-A_1)]u(k-1) \\
&\quad +(1-A_1+A_1^2)y(k) + A_1(1-A_1)y(k-1) \\
&\vdots \\
y(k+N) &= \sum_{i=1}^{N} F_i^{N-i} u(k+i-d) \\
&\quad +\left(\sum_{i=1}^{N}(-1)^{N-i} G_i^{N-i} - z^{1-d}B\sum_{i=1}^{N}(-1)^{i-1}A_1^{i-1}\right)u(k-1) \\
&\quad +\sum_{i=1}^{N}(-1)^i A_1^i y(k) + A_1\sum_{i=1}^{N-1}(-1)^i A_1^i y(k-1)
\end{aligned}$$

The vector form is

$$\mathbf{y} = \mathbf{Fu} + \mathbf{H}u(k-1) + \mathbf{M}y(k) \qquad (20)$$

where the vectors $\mathbf{y}$ and $\mathbf{u}$ are the same as the aforesaid and

$$\mathbf{F} = \begin{bmatrix} F_1^0 & 0 & \cdots & 0 \\ -F_1^1 & F_2^0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^N F_1^N & (-1)^{N-1}F_2^{N-1} & \cdots & F_N^0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} G_1^0 - z^{1-d}B \\ G_2^0 - G_1^1 - z^{1-d}B(1-A_1) \\ \vdots \\ \sum_{i=1}^{N}(-1)^{N-i}G_i^{N-i} - z^{1-d}B\sum_{i=1}^{N}(-1)^{i-1}A_1^{i-1} \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} 1 - A_1 + A_1 z^{-1} \\ 1 - A_1 + A_1^2 + A_1(1-A_1)z^{-1} \\ \vdots \\ \sum_{i=0}^{N}(-1)^i A_1^i + z^{-1}A_1\sum_{i=0}^{N-1}(-1)^i A_1^i \end{bmatrix}$$

For instance, for the first order system

$$y(k) + ay(k-1) = bu(k-1)$$

Following will be resulted

$$\mathbf{F} = b\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -a & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (-a)^{N-1} & (-a)^{N-2} & \cdots & 1 \end{bmatrix}$$

$$\mathbf{H} = -b\begin{bmatrix} 1 & 1-a & 1-a+a^2 & \cdots & \dfrac{1-(-a)^N}{1+a} \end{bmatrix}^T$$

$$\mathbf{M} = \begin{bmatrix} 1 - a + az^{-1} \\ 1 - a + a^2 + a(1-a)z^{-1} \\ \vdots \\ \dfrac{1-(-a)^{N+1} + a[1-(-a)^N]z^{-1}}{1+a} \end{bmatrix}$$

By introducing $\tilde{\mathbf{y}} = \mathbf{H}u(k-1) + \mathbf{M}y(k)$, Eq. (20) can be expressed in the same form of Eq. (10).

The predictor developed here can be called general predictor, because it works not only for unstable processes but also for stable processes and integrating ones. An integrating process controlled by a controller based on current predictor owns the double integrating action for the process itself possesses integrating effect.

Through a similar course, we can get a predictor for double integral action.

$$\mathbf{y} = \mathbf{F}\Delta\mathbf{u} + \mathbf{H}\Delta u(k-1) + \mathbf{M}y(k)$$

It can be expressed in the form of (12) if the following is defined.

$$\tilde{\mathbf{y}} = \mathbf{H}\Delta u(k-1) + \mathbf{M}y(k)$$

3.4. Unified predictor

Assume that the polynomial $A(z^{-1})$ with degree n, can be factorized as

$$A(z^{-1}) = A_2(z^{-1})A_0(z^{-1})$$

where $A_2(z^{-1})$ contains unstable and/or poor damped modes and $A_0(z^{-1})$ contains stable and well damped modes. $A_0(z^{-1})$ and $A_2(z^{-1})$ are monic because so $A(z^{-1})$ is. Therefore $A_2(z^{-1})$ can be expressed as

$$A_2(z^{-1}) = 1 + z^{-1}A_1(z^{-1})$$

Now, the transfer function (2) can be rewritten as

$$y(k) = \frac{B}{A_0}u(k-d) - A_1 y(k-1) \qquad (21)$$

and h steps ahead

$$y(k+h)=\frac{B}{A_0}u(k+h-d)-A_1y(k+h-1) \qquad (22)$$

Subtracting Eq. (21) from Eq. (22) and rearranging the terms, we obtain

$$\begin{aligned} y(k+h)&=\frac{B}{A_0}u(k+h-d)-A_1y(k+h-1)\\ &\quad+y(k)-\frac{B}{A_0}u(k-d)+A_1y(k-1) \end{aligned} \qquad (23)$$

where it is assumed once again the sum

$$y(k)-\frac{B}{A_0}u(k-d)+A_1y(k-1)$$

may be different from zero, so a constant error is incorporated over the prediction horizon which is, according to the model (1),

$$\varepsilon(k+h)=\frac{B}{A_0}v(k-d)+A_2e(k) \qquad \forall h$$

Let us introduce the identity

$$\frac{A_1^iB}{A_0}=F_h^i+z^{-(h-d+1)}\frac{G_h^i}{A_0}$$

where $h=1,2,\cdots,N$ and $i=1,2,\cdots,N-1$, besides, F_h^i and G_h^i are the quotient and remainder of the division A_1^iB/A_0, respectively. A_1^iB is a polynomial of degree $(i+1)(n-1)$. Through iterating Eq. (23) for all values of h, we get the following.

$$\begin{aligned} y(k+1)&=F_1^0u(k+1-d)+\frac{1}{A_0}(G_1^0-z^{1-d}B)u(k-1)\\ &\quad+(1-A_1)y(k)+A_1y(k-1)\\ y(k+2)&=-F_1^1u(k+1-d)+F_2^0u(k+2-d)\\ &\quad+\frac{1}{A_0}[G_2^0-G_1^1-z^{1-d}B(1-A_1)]u(k-1)\\ &\quad+(1-A_1+A_1^2)y(k)+A_1(1-A_1)y(k-1)\\ &\vdots\\ y(k+N)&=\sum_{i=1}^{N}(-1)^{N-i}F_i^{N-i}u(k+i-d)\\ &\quad+\frac{1}{A_0}\left(\sum_{i=1}^{N}(-1)^{N-i}G_i^{N-i}-z^{1-d}B\sum_{i=1}^{N}(-1)^{i-1}A_1^{i-1}\right)u(k-1)\\ &\quad+\sum_{i=0}^{N}(-1)^iA_1^iy(k)+A_1\sum_{i=0}^{N-1}(-1)^iA_1^iy(k-1) \end{aligned}$$

The predictions can be expressed in condensed form as

$$\mathbf{y}=\mathbf{F}\mathbf{u}+\mathbf{H}u(k-1)+\mathbf{M}y(k) \qquad (24)$$

where the vectors $\mathbf{y}$ and $\mathbf{u}$ are the same as the aforesaid and

$$\mathbf{F}=\begin{bmatrix} F_1^0 & 0 & \cdots & 0\\ -F_1^1 & F_2^0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ (-1)^NF_1^N & (-1)^{N-1}F_2^{N-1} & \cdots & F_N^0 \end{bmatrix} \qquad (25)$$

$$\mathbf{H}=\frac{1}{A_0}\begin{bmatrix} G_1^0-z^{1-d}B\\ G_2^0-G_1^1-z^{1-d}B(1-A_1)\\ \vdots\\ \sum_{i=1}^{N}(-1)^{N-i}G_i^{N-i}-z^{1-d}B\sum_{i=1}^{N}(-1)^{i-1}A_1^{i-1} \end{bmatrix} \qquad (26)$$

$$\mathbf{M}=\begin{bmatrix} 1-A_1+A_1z^{-1}\\ 1-A_1+A_1^2+A_1(1-A_1)z^{-1}\\ \vdots\\ \sum_{i=0}^{N}(-1)^iA_1^i+z^{-1}A_1\sum_{i=0}^{N-1}(-1)^iA_1^i \end{bmatrix} \qquad (27)$$

After introducing $\tilde{\mathbf{y}}=\mathbf{H}u(k-1)+\mathbf{M}y(k)$, Eq. (24) can be expressed in the same form of Eq. (10).

Notice that the predictor (24) is identical to that one for stable processes if $A_0=A$, consequently $A_2=1$ and $A_1=0$. The predictor for integrating processes is the case when $A_2=1-z^{-1}$, clearly $A_1=-1$. If $A_2=A$, that means $A_0=1$, the general predictor results.

Analogously, a unified predictor for obtaining increments of control signal can be derived.

$$\mathbf{y}=\mathbf{F}\Delta\mathbf{u}+\mathbf{H}\Delta u(k-1)+\mathbf{M}y(k)$$

It can, of course, be expressed in the form (12) if corresponding definition is introduced.

By introducing

$$A_I=\begin{cases} A_1 & \text{for obtaining } u(t)\\ A_1-A_2 & \text{for obtaining } \Delta u(t) \end{cases}$$

a general expression of identity can be found

$$\frac{A_I^iB}{A_0}=F_h^i+z^{-(h-d+1)}\frac{G_h^i}{A_0} \qquad \forall h \text{ and } \forall i \qquad (28)$$

For the last, if introduce the following notation:

$$\mathbf{w} = \begin{cases} \mathbf{u} & \text{for obtaining } u(t) \\ \Delta\mathbf{u} & \text{for obtaining } \Delta u(t) \end{cases} \tag{29}$$

all the above-mentioned predictors can be expressed in a unique form

$$\mathbf{y} = \mathbf{F}\mathbf{w} + \tilde{\mathbf{y}} \tag{30}$$

4. OBJECTIVE FUNCTION

Different objective functions have been proposed for different model predictive control (MPC) algorithms. But the aim is basically the same. The distance between the future output and the reference and the control efforts are penalized. The general expression for such an objective function is

$$\begin{aligned} J(N_1, N_2, N_u) = &\sum_{i=N_1}^{N_2} \lambda(i)[y(k+i) - r(k+i)]^2 \\ &+ \sum_{j=1}^{N_u} \rho(j)[\Delta^2 u(k+j-1)]^2 \end{aligned} \tag{31}$$

where N_1 and N_2 are the minimum and maximum prediction horizons and N_u is the control horizon. $\lambda(i)$ and $\rho(j)$ are coefficients whose elections depend on pretension to future behavior of the system. The signal r is the reference.

One of the advantages of predictive control is that the system can react before the change has effectively been made if the future evolution of the reference is known a priori. That, consequently, avoids the effect of delay in the process response.

It is needed to redefine the objective function in order to obtain an optimal sequence of increments of control signal.

$$\begin{aligned} J(N_1, N_2, N_u) = &\sum_{i=N_1}^{N_2} \lambda(i)[y(k+i) - r(k+i)]^2 \\ &+ \sum_{j=1}^{N_u} \rho(j)[\Delta^2 u(k+j-1)]^2 \end{aligned} \tag{32}$$

A reference trajectory is frequently used in the objective functions (31) and (32):

$$\begin{aligned} &y_m(k) = y(k) \\ &y_m(k+h) = \alpha y_m(k+h-1) + (1-\alpha)r(k+h) \qquad h = 1, \cdots, N \end{aligned}$$

where α is a parameter of design and takes the value between zero and one, the closer to unity the smoother the approximation.

In this paper $\rho(j) = \text{constant}, \forall j$ and $\lambda(i) = 1, \forall i$, are adopted for the sake of simplicity, unless otherwise indicated.

5. CONTROL LAW

The aim of the MPC is that the future output on considered horizon should follow a determined reference signal but the control effort is taken into account too. Thus the objective function should be minimized. The product of the minimization is the control law. Namely, the objective function will be minimized by using the resultant control law. For this purpose, the following are introduced.

$$\begin{aligned} &\mathbf{r} = \begin{bmatrix} r(k+1) & r(k+2) & \cdots & r(k+N) \end{bmatrix}^T \\ &\mathbf{u}(k) = [u(k+1-d \mid k) \quad \cdots \quad u(k+N-d \mid k)]^T \\ &\mathbf{u}(k-1) = [u(k+1-d \mid k-1) \quad \cdots \quad u(k+N-d \mid k-1)]^T \end{aligned}$$

hence

$$\Delta\mathbf{u} = [\mathbf{u}(k) - \mathbf{u}(k-1)]^T$$

Assume $N_1 = 1$ without any loss of generality. The cost function (31) can be written as

$$\begin{aligned} J(1, N_2, N_u) &= (\mathbf{y} - \mathbf{r})^T(\mathbf{y} - \mathbf{r}) + \rho\Delta\mathbf{u}^{\mathbf{T}}\Delta\mathbf{u} \\ &= [\mathbf{F}\mathbf{u}(k) + \tilde{\mathbf{y}} - \mathbf{r}]^T[\mathbf{F}\mathbf{u}(k) + \tilde{\mathbf{y}} - \mathbf{r}] \\ &\quad + \rho[\mathbf{u}(k) - \mathbf{u}(k-1)]^T[\mathbf{u}(k) - \mathbf{u}(k-1)] \end{aligned}$$

where Eq. (10) is used as a general case. Minimizing the cost function with respect to $\mathbf{u}(k)$, we get

$$\mathbf{u}(k) = (\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})^{-1}[\mathbf{F}^T(\mathbf{r} - \tilde{\mathbf{y}}) + \rho\mathbf{u}(k-1)] \tag{33}$$

The first element $u(k)$ of vector $\mathbf{u}(k)$ is applied to the process. The control law is calculated again when a new measurement is obtained at the next sampling instant. Thus, the receding-horizon control concept is used.

The control law (33) incorporates implicitly the integral action. This will be proved below. It is time invariant if the process to be controlled is time invariant.

A matrix of dimension $N \times N$ has to be inverted in the calculation of the control law, where N is the prediction horizon. It is possible to introduce constraints on the future control signal in order to decrease the computations. Assume, for example, that the control signals are constant after N_u steps, with $N_u < N$,

$$u(k + N_u) = u(k + N_u + 1) = \cdots = u(k + N)$$

This implies that the control increments are assumed to be zeros after N_u steps. The control law (33) should be modified to

$$\mathbf{u}(k) = (\mathbf{F}_1{}^T\mathbf{F}_1 + \rho\mathbf{I})^{-1}[\mathbf{F}_1{}^T(\mathbf{r} - \tilde{\mathbf{y}}) + \rho\mathbf{u}(k-1)]$$

where $\mathbf{F}_1$ is a $N \times N_u$ matrix. Therefore, the matrix to be inverted is now $N_u \times N_u$ dimensions. For instance, applying mentioned assumption, the matrix described in Eq. (11) will be

$$\mathbf{F}_1 = \begin{bmatrix} f_0 & 0 & \cdots & 0 & 0 \\ f_1 & f_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{N_u-1} & f_{N_u-2} & \cdots & f_1 & f_0 \\ f_{N_u} & f_{N_u-1} & \cdots & f_2 & f_1 + f_0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{N-1} & f_{N-2} & \cdots & f_{N-N_u+1} & \sum_{i=0}^{N-N_u} f_i \end{bmatrix}$$

By an analogous procedure and with the predictor (12), we can get an optimal sequence of increments of the control signal for double integral action.

$$\Delta\mathbf{u}(k) = (\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})^{-1}[\mathbf{F}^T(\mathbf{r} - \tilde{\mathbf{y}}) + \rho\Delta\mathbf{u}(k-1)] \qquad (34)$$

The control law can be expressed as

$$\begin{aligned} \Delta u(k) &= \mathbf{L}[\mathbf{F}^T(\mathbf{r} - \tilde{\mathbf{y}}) + \rho\Delta\mathbf{u}(k-1)] \\ u(k) &= u(k-1) + \Delta u(k) \end{aligned} \qquad (35)$$

where $\mathbf{L}$ is the first row of the matrix $(\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})^{-1}$.

Now let us demonstrate one of the properties of the controller (33). Firstly we introduce a definition.

DEFINITION 1 Estimation of Output

The vector given by $\hat{\mathbf{y}} = \mathbf{F}\mathbf{u}(k-1) + \tilde{\mathbf{y}}$ *is a estimation of the vector given by* $\mathbf{y} = \mathbf{F}\mathbf{u}(k) + \tilde{\mathbf{y}}$ *with the available data at instant k before calculating* $\mathbf{u}(k)$, *because* $\mathbf{u}(k-1)$ *is obtained before instant k and* $\tilde{\mathbf{y}}$ *is obtained at instant k.*

THEOREM 1 Integral Action

The control law (33) derived by means of minimizing the cost function (31) possesses the integral action.

Proof: Assume $N_1 = d = 1$ *without any loss of generality. The control law can be rewritten as*

$$(\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})\mathbf{u}(k) - \rho\mathbf{u}(k-1) = \mathbf{F}^T(\mathbf{r} - \tilde{\mathbf{y}})$$

Subtracting the term $\mathbf{F}^T\mathbf{F}\mathbf{u}(k-1)$ *on both sides of the last equation, using the definition and gathering the terms, we get*

$$(\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})[\mathbf{u}(k) - \mathbf{u}(k-1)] = \mathbf{F}^T(\mathbf{r} - \hat{\mathbf{y}})$$

Thus

$$\mathbf{u}(k) = \mathbf{u}(k-1) + (\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})^{-1}\mathbf{F}^T(\mathbf{r} - \hat{\mathbf{y}})$$

In words, the control law (33) has the integral action. □

It deserves to be mentioned that a controlled system will have offset in steady-state if u instead of Δu is penalized. But it is an alternative penalizing u when a regulator is wanted.

Now let us introduce the suboptimal concept. When replacing the vector $\mathbf{u}(k-1) = [u_h(k-1)]$, $h = 1, \cdots, N$, with $\mathbf{1}u_1(k-1)$, where $\mathbf{1} = [1 \;\; 1 \;\; \cdots \;\; 1]^T$ an $N \times 1$ vector, in the calculating of the control signal, a suboptimal controller is obtained. The suboptimal concept can be used to reduce computation burden and overshoot. For example, the process $G(s) = 1/(s+1)$ is controlled by a controller with following parameters $N = 8, N_u = 1$ and $\rho = 0.1$. Its step response is drawn in Fig.1 (blue trajectory). Assume a quicker recovery from disturbance is required. It can be reached with larger N_u. The red trajectory (left drawing of Fig.1) is resulted when $N_u = 5$. Larger N_u means larger burden of computation. But it can be reduced by using the suboptimal concept as above-mentioned. The result is plotted in the left drawing of Fig. 1 (dashed). Notice that the dashed one and the red one are very similar. Here the reference trajectory which was mentioned in Section IV is used to lower the overshoots. However, the suboptimal concept can also be used to change other dynamic aspects of a controlled system. Fig.1 (right) depicted the result of using it to cut down the overshoot of the same process given above. After applying the suboptimal concept, the dashed trajectory is gotten.

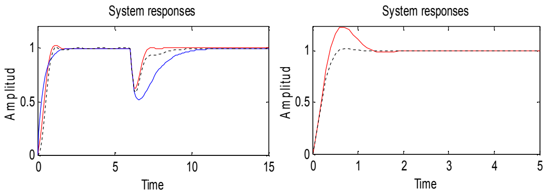


Figure1. Step Responses of the System. The Left: $N_u = 1$ (blue), $N_u = 5$ (red) and $N_u = 2$ (Dashed). A Load Disturbance Is Introduced at Time 6. The Right: Results of Applying the Suboptimal Concept to Reduce Overshoot.

In order to demonstrate the double integral action of the control law (35), the following definition is introduced.

DEFINITION 2 Estimation of Output

The vector given by $\hat{\mathbf{y}} = \mathbf{F}\Delta\mathbf{u}(k-1) + \tilde{\mathbf{y}}$ *is a estimation of the vector given by* $\mathbf{y} = \mathbf{F}\Delta\mathbf{u}(k) + \tilde{\mathbf{y}}$ *with the available data at instant k, because* $\mathbf{u}(k-1)$ *is obtained before instant k and* $\tilde{\mathbf{y}}$ *is obtained at instant k.*

THEOREM 2 Double Integral Action

The control law (35) derived by means of minimizing the cost function (32) possesses the double integral action.

Proof: Assume $N_1 = d = 1$ *without any loss of generality. Eq. (34) can be rewritten as*

$$(\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})\Delta\mathbf{u}(k) - \rho\Delta\mathbf{u}(k-1) = \mathbf{F}^T(\mathbf{r} - \tilde{\mathbf{y}})$$

Subtracting the term $\mathbf{F}^T\mathbf{F}\Delta\mathbf{u}(k-1)$ *on both sides of the last equation and manipulating algebraically, we get*

$$\Delta\mathbf{u}(k) = \Delta\mathbf{u}(k-1) + (\mathbf{F}^T\mathbf{F} + \rho\mathbf{I})^{-1}\mathbf{F}^T(\mathbf{r} - \hat{\mathbf{y}})$$

where $\hat{\mathbf{y}} = \mathbf{F}\Delta\mathbf{u}(k-1) + \tilde{\mathbf{y}}$ *as defined. Thus the sequence* $\Delta\mathbf{u}$ *is obtained by means of integrating. That signifies the increment* $\Delta u(k)$ *is given by integration. And another integral action is given by*

$$u(k) = u(k-1) + \Delta u(k)$$

Therefore the double integral action of the control law (35) is revealed. □

Previous theorems are illustrated with some simple examples. All of them are the particular case with following assumptions: constant future control, namely $u(k+i) = u(k),\quad i = 1,2,\cdots,h,$ there is no constraint on the control effort and it is desired that $y(k+h) = r(k+h),$ where h is the prediction horizon. In other words this is the case when $N_u = 1,\quad \rho = 0$ and $\lambda(i) = 0,\quad i = 1,2,\cdots,h-1$ but $\lambda(h) = 1.$

Example 1. Consider a first order process

$$y(k) + ay(k-1) = bu(k-1)$$

If it is open-loop stable and well damped, we can use the predictor (10) for stable process case. After algebraic manipulations, we obtain the control law

$$u(k) = u(k-1) + \frac{(1+a)(1+az^{-1})}{b[1-(-a)^h]}[r(k+h) - y(k)]$$

We can also use the predictor for unstable process case to obtain the control law no matter whether the process may be stable or not.

$$u(k) = u(k-1) + \frac{1+a}{b[1-(-a)^h]}\left\{r(k+h) - \left[\alpha + \beta z^{-1}\right]y(k)\right\}$$

where

$$\alpha = \frac{a[1-(-a)^{h+1}]}{1+a} \qquad \beta = \frac{a[1-(-a)^h]}{1+a}$$

Example 2. Consider the same process of Example 1 and assume it is open-loop stable and well damped. It is desired that the output of the process can follow a mixed trajectory composed of steps and ramps. So the double integral action is needed. Using the predictor (12) and manipulating algebraically, we will finally get

$$\Delta u(k) = \Delta u(k-1) + \xi[r(k+h) - (h+1-hz^{-1})y(k)]$$
$$u(k) = u(k-1) + \Delta u(k)$$

where

$$\xi = \frac{(1+a)^2(1+az^{-1})}{b[h(1+a) + a + (-a)^{h+1}]}$$

Example 3. Consider an integrating process

$$y(k) = \frac{b_1 + b_2 z^{-1}}{(1-z^{-1})(1+az^{-1})}u(k-1)$$

with the pole $p = -a$ in the unit disc and well damped. Using the corresponding predictor, we can get the following control law through same manipulations.

$$u(k) = u(k-1) + \eta[r(k+h) - (h+1-hz^{-1})y(k)]$$

where $\eta = \dfrac{(1+a)(1+az^{-1})}{h(b_1+b_2) - \dfrac{b_2 - b_1 a}{1+a}[1-(-a)^h] + h(1+a)b_2 z^{-1}}$

6. STABILITY AND ROBUSNESS

No Mathematic model is able to describe exactly a physical process. Some approximations are always made. However, it is desired that the controlled system should be insensitive to those uncertainties in the model.

In the absence of theoretical results, some numerical analyses are presented.

Let $N_1 = d$ and assume the transport delays are multiple of the sampling time for all cases. Only a few types of uncertainties are considered here.

Case 1. Consider a stable process

$$G(s) = \frac{b}{(s+a)}e^{-\tau s}$$

with nominal values $a = 2$ and $b = 0.5,$ which is sampled with sampling time 0.1 seconds, so the discrete nominal plant is

$$y(k) - 0.8187y(k-1) = 0.0453u(k-d)$$

The nominal system is stable for all $N \geq d,$ and $1 \leq N_u \leq N - d + 1.$

The simulations with $N - N_1 = 0, N_u = 1$ and $\rho = 0.1$ show the followings.

1. Uncertainty at the pole: When $d = 1,$ the system is stable for $a \geq 0.01.$ That is to say an uncertainty of about ±99.5% is permitted.

When $d = 10$, the system keeps stable for $a \geq 0.6$. That means an uncertainty of about $\pm$ 70% is permitted. Fig. 2 is step responses and Nyquist diagrams.

2. Uncertainty of the gain: When $d = 1$, the system dose not loss stability for $0.01 \leq b \leq 80$. Namely a variation about ±98% is allowed. When $d = 10$, the system is stable for $0.01 \leq b \leq 1.6$. That is a variation about ±98% is allowed. See Fig. 2.
3. Unmodelled pole: Assume that the real process has another pole -2σ but its static gain maintains the same in spite of the existence of another pole. When $d = 1$, the system is stable for $\sigma \geq 0.1$. When $d = 10$, the system is also stable for $\sigma \geq 0.1$. Fig. 3 shows step responses of the controlled system and Nyquist diagrams when the real process has a less dominant pole -10.
4. Uncertainty at the delay: When $d = 1$, an error of 12 units is tolerated. When $d = 10$, an error of 19 units is tolerated and for all the possible negative values of delay mismatch, from -1 to -9, the controlled system does not lose stability. See Fig. 4.

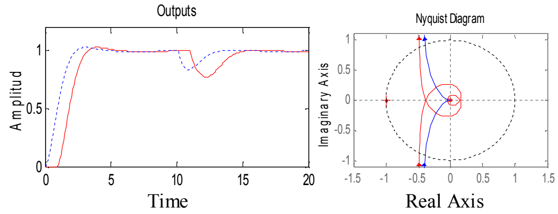


Figure 2. Responses of the Nominal System to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) when $d = 1$ (Blue) and $d = 10$ (Red).

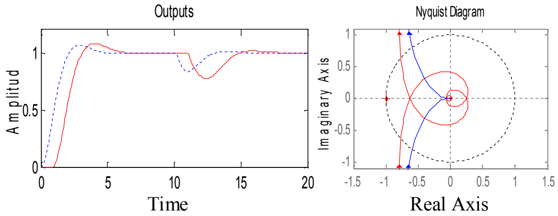


Figure 3. Responses of the System to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) with Unmodelled Mode when $d = 1$ (Blue Trajectory) and $d = 10$ (Red Trajectory).

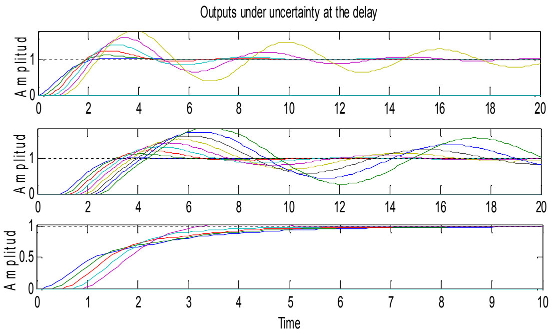


Figure 4. Step Responses of the Controlled System under the Uncertainty at the Delay, when $d = 1$ (Above) and $d = 10$ (Middle and Below). Only Those with Even Numbers of Delay Units Are Depicted for Clarity.

Case 2. Consider a stable second order process

$$G(s) = \frac{\omega^2}{s^2 + 2\zeta\omega s + \omega^2} e^{-\tau s}$$

with nominal values $\zeta = 0.5$ and $\omega = 1$ rad/s, which means the process has poles at $-0.5 \pm j0.8660$. It is sampled with sampling time 0.1 seconds.

$$y(k) = \frac{0.0047 + 0.0044z^{-1}}{1 - 1.8097z^{1} + 0.8187z^{-2}} u(k-d)$$

The nominal system is stable for $N - N_1 = 0$ and $N - N_1 \geq 4$. The simulations with $N - N_1 = 7, N_u = 1$ and $\rho = 0.1$ obtain the following results.

1. Uncertainty of the damping: When $d = 1$, the system is stable for $0.01 \leq \zeta \leq 10$. That is to say an uncertainty of about ±98% is permitted. When $d = 10$, the system keeps stable for $0.2 \leq \zeta \leq 10$. That means an uncertainty of about $\pm$ 30% is permitted. Fig. 5 is the step responses of the controlled nominal process and the correspondent Nyquist diagrams.
2. Uncertainty of the nature frequency: When $d = 1$, the system does not lose stability for $0.45 \leq \omega \leq 2.5$. Namely a variation about ±55% is allowed. When $d = 10$, the system is stable for $0.5 \leq \omega \leq 1.35$. That is a variation about ±35% is allowed. See Fig. 5.
3. Unmodelled pole: Assume that the real process has another pole -0.5σ but its static gain maintains the same in spite of the existence of another pole. When $d = 1$, the system is stable for $\sigma \geq 0.3$. When $d = 10$, the system is also stable for $\sigma \geq 0.3$. Fig. 6 shows the step responses and the Nyquist diagrams when the

real process has a less dominant pole -10. If the real process has a couple of conjugated complex poles $-0.5\sigma \pm j0.8660\varpi$ is assumed, the controlled system is stable, for example, for $\sigma \geq 3, \varpi = 1$ or for $\sigma = 1, \varpi \geq 5$ when $d = 1$, and for $\sigma \geq 1.5, \varpi = 1$ or $\sigma = 1, \varpi \geq 2$ when $d = 10$. A drawing similar to Fig. 6 can be obtained under the assumption that the real process has another less dominant couple of poles $-10 \pm j8.660$.

4. Uncertainty at the delay: When $d = 1$, an error of 6 units is tolerated. When $d = 10$, an error of 14 units is tolerated and for all the possible negative values of delay mismatch, from -1 to -9, the controlled system remains stable. See Fig. 7.

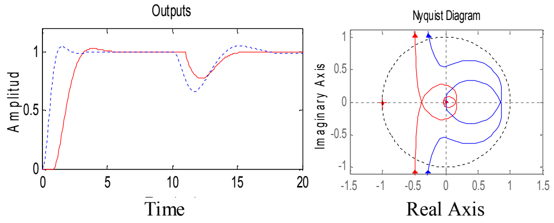


Figure 5. Responses of Nominal System to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) when $d = 1$ (Blue) and $d = 10$ (Red).

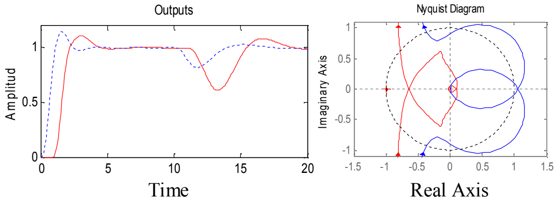


Figure 6. System's Responses to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) of the System with Unmodelled Mode when $d = 1$ (Blue) and $d = 10$ (Red).

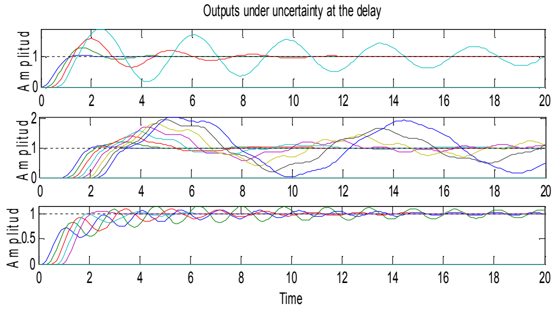


Figure 7. Responses of the Controlled System to Step Type Reference under the Uncertainty at the Delay, when $d = 1$ (Above) and $d = 10$ (Middle and Below). Only Those with Even Numbers of Delay Units Are Depicted for Clarity.

Case 3. Consider a second order integrating process

$$G(s) = \frac{a}{s(s+a)} e^{-\tau s}$$

with nominal values $a = 2$. It is sampled with sampling time 0.1 seconds.

$$y(k) = \frac{0.0094 + 0.0088z^{-1}}{(1-z^{1})(1-0.8187)z^{-2}} u(k-d)$$

The nominal system is stable for $N \geq 4$ and $N_u = 1$ when $d = 1$. It is stable for $N \geq 10$ and $N_u = 1$ when $d = 10$. For this particular case, the controlled system is not always stable for $N_u \leq N - d + 1$. The simulations with $N - N_1 = 7, N_u = 1$ and $\rho = 0.1$ when $d = 1$ and with $N - N_1 = 32, N_u = 1$ and $\rho = 0.1$ when $d = 10$ show the followings.

1. Uncertainty of the parameter a: When $d = 1$, the system is stable for $0.01 \leq a \leq 20$. That is to say an uncertainty of about $\pm 98\%$ is permitted. When $d = 10$, the system maintains stability for $0.9 \leq a \leq 20$. That means an uncertainty of about $\pm$ 55% is permitted. Fig. 8 is the step responses of the controlled nominal process and the correspondent Nyquist diagrams.
2. Unmodelled pole: Assume that the real process has another pole -2σ but its static gain maintains the same in spite of the existence of another pole. When $d = 1$, the system is stable for $\sigma \geq 2.4$. When $d = 10$, the system is stable for $\sigma \geq 3.5$. Fig. 9 shows the step responses and the Nyquist diagrams when the real process has a less dominant pole -10 or -20.
4. Delay uncertainty: When $d = 1$, an error of 2 units is tolerated and up to 4 is tolerated when $N = 18$. When $d = 10$, an error of 6 units is tolerated and up to 10 units is tolerated when $N = 50$. And for all the possible negative values of delay mismatch, from -1 to -9, the controlled system preserves stability. See Fig. 10. It is noticed that the tolerance is dependent on prediction horizon.

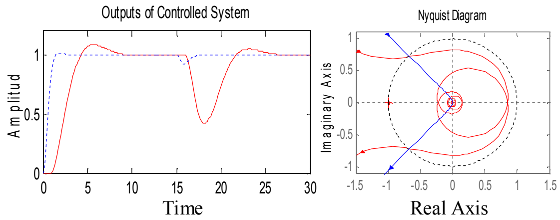


Figure 8. Responses of the Nominal System to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) when $d=1$ (Blue Trajectory) and $d=10$ (Red Trajectory).

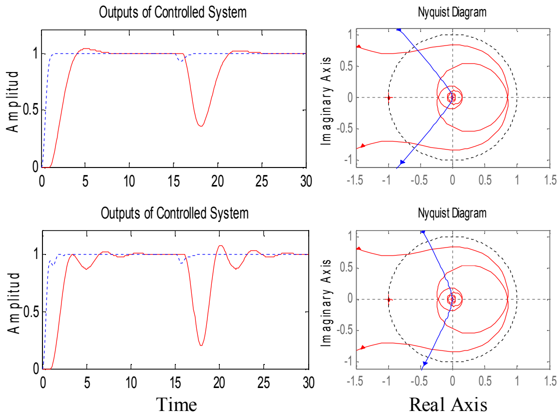


Figure 9. Responses of the Controlled System to Step Type Reference and Load Disturbance (Left), and Nyquist Diagrams (Right) with Unmodelled Mode when $d=1$ (Blue Trajectory) and $d=10$ (Red Trajectory) and when $\sigma=10$ (Superior) and $\sigma=5$ (Inferior).

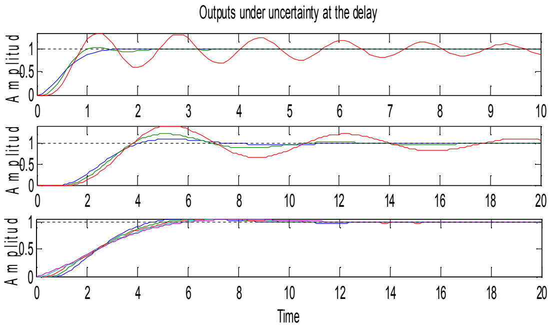


Figure 10. Step Responses of the System under the Uncertainty at the Delay, when $d=1$ (Above) and $d=10$ (Middle and Below, Only Those with Odd Numbers of Delay Units Are Depicted for Clarity).

Three numerical analyses of controlled systems have been given. Those controllers were based on the predictors for stable processes or for integrating processes. In the following the controller based on the predictor for unstable processes will be used in order to compare the proposed controller with that based on the CARIMA model.

The CARIMA model is

$$A(z^{-1})y(k)=z^{-d}B(z^{-1})u(k)+C(z^{-1})\frac{e(k)}{\Delta}$$

where e is a white noise. The comparisons will de made under the conditions $C(z^{-1})=1$ for the CARIMA model, and $S_1(z^{-1})=1, T_1(z^{-1})=1$ (see Section VII for the definitions) for the ARMA model.

The process to be controlled is a stable one.

$$y(k)-ay(k-1)=bu(1-d)$$

For the range of variation of the parameters: $0.1\le a\le 0.98$ and the possible values of gain with which the controlled system keeps stable. The numerical results show that no difference, in respect of tracking a reference and rejecting disturbances, can be observed between the controller based on the predictor for unstable processes and that based on CARIMA model when $d=1$. However, when $d=10$, the pole is big $a=0.98$, and the process gain is 5, the system controlled by controller based on predictor for unstable processes can tolerate a delay estimation error of 7 units. See Fig. 10. Nevertheless, for a delay of 10 units a delay mismatch of one unit is permitted by the controller based on CARIMA model (Camacho, Bordons, 2007).

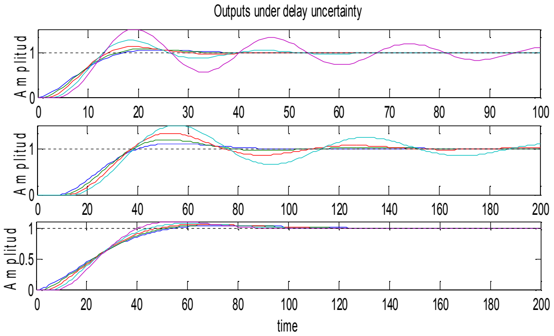


Figure11. Step Responses of the System under Delay Uncertainties, when $d=1$ (Above) and $d=10$ (Middle and Below, Only Those with Even Numbers of Delay Units Are Depicted for Clarity).

7. A CLASSICAL POINT OF VIEW

The terms T and R can be incorporated when the controller (33) or (35) is viewed as a classical controller,

$$R(z^{-1})u(k)=T(z^{-1})r(k)-S(z^{-1})y(k) \tag{36}$$

For doing so, introduce

$$y_f(k)=\frac{S_1(z^{-1})}{T_1(z^{-1})}y(k) \tag{37}$$

with $S_1(1)/T_1(1)=1$ for avoiding static error, and factorize A as $A=A_2A_0$ and use the identity (28) in order to get a general sense result. Thereby the model (2) can be rewritten as

$$A_2(z^{-1})y_f(k)=\frac{S_1(z^{-1})}{A_0(z^{-1})T_1(z^{-1})}w(k-d)$$

where w is defined as (29).

By a similar way to that in Section II, we can get a predictor with the form of Eq. (30). It can be also expressed in the form of Eq. (19),

$$y_f=\mathbf{F}\mathbf{u}+\mathbf{H}u(k-1)+\mathbf{M}y_f(k) \tag{38}$$

where $\mathbf{F}$ has the same form of (25), replacing the polynomials T_1, B and A_1 of the Eq. (26) and (27) with A_0T_1, BS_1 and A_I, respectively, we can get the new matrixes $\mathbf{H}$ and $\mathbf{M}$ for Eq. (38).

Let $\Gamma=[\gamma_i], i=1,2,\cdots,N$ be the first row of the matrix $(\mathbf{F}^T\mathbf{F}+\rho\mathbf{I})^{-1}\mathbf{F}^T$ and $\mathbf{1}=[1\ \ 1\ \ \cdots\ \ 1]^T$ an $N\times 1$ vector, where N is the prediction horizon.

Eq. (33) and (34) can be expressed as

$$\mathbf{w}(k)=(\mathbf{F}^T\mathbf{F}+\rho\mathbf{I})^{-1}[\mathbf{F}^T(\mathbf{r}-\tilde{\mathbf{y}})+\rho\mathbf{w}(k-1)]$$

It is equivalent to

$$(\mathbf{F}^T\mathbf{F}+\rho\mathbf{I})\mathbf{w}(k)-\rho\mathbf{1}w(k-1)=\mathbf{F}^T(\mathbf{r}-\tilde{\mathbf{y}})$$

because the receding-horizon concept is used. Subtracting $\mathbf{F}^T\mathbf{F}\mathbf{1}w(k-1)$ from both sides of the last equation and manipulating algebraically, we get

$$\mathbf{w}(k)-\mathbf{1}w(k-1)=(\mathbf{F}^T\mathbf{F}+\rho\mathbf{I})^{-1}\mathbf{F}^T[\mathbf{r}-\tilde{\mathbf{y}}-\mathbf{F}\mathbf{1}w(k-1)]$$

However, only the first element of the vector is used, therefore the following results.

$$(1-z^{-1})w(k)=\Gamma[\mathbf{r}-\tilde{\mathbf{y}}-\mathbf{F}\mathbf{1}w(k-1)]$$

Assume $N_1=1$ without any loss of generality and consider that the future reference keeps constant along the horizon or its evolution is unknown. That is $w(k+h)=w(k), \forall h.$ By means of algebraic manipulations, we can get

$$\begin{aligned}(1-z^{-1})R_1w(k)&=A_0T_1\sum_{i=1}^{N}\gamma_i r(k)\\&-A_0S_1\sum_{i=1}^{N}\gamma_i\sum_{j=1}^{i}[1+(-1)^j A_I^j(1-z^{-1})]y(k)\end{aligned}$$

where

$$\begin{aligned}(1-z^{-1})R_1&=(1-z^{-1})A_0T_1\\&+z^{-1}\sum_{i=1}^{N}\gamma_i\sum_{j=1}^{i}(-1)^{i-j}\left[(1-z^{-i})G_j^{i-j}+(1-z^{1-d})A_I^{i-j}BS_1\right]\\&=(1-z^{-1})A_0T_1\\&+(1-z^{-1})z^{-1}\sum_{i=1}^{N}\gamma_i\sum_{j=1}^{i}(-1)^{i-j}\left[P_1^{1-i}G_j^{i-j}+P_2^{2-d}A_I^{i-j}BS_1\right]\end{aligned}$$

where P^l denotes a polynomial of degree l. The second equality is obtained for the fact that a polynomial $1-z^{-m}$, m integers, has always a factor $1-z^{-1}$ is used (De Moivre's theorem). Thus compare the third equation from the bottom with Eq. (36), we get

$$\begin{aligned}R&=(1-z^{-1})R_1\\S&=A_0S_1\sum_{i=1}^{N}\gamma_i\sum_{j=1}^{i}[1+(-1)^j A_I^j(1-z^{-1})]\\T&=A_0T_1\sum_{i=1}^{N}\gamma_i\end{aligned}$$

The controlled system is, according to the model (1),

$$\begin{aligned}A_2x(k)&=\frac{z^{-d}B}{A_0}[u(k)+v(k)]\\y(k)&=x(k)+e(k)\\Ru(k)&=Tr(k)-Sy(k)\end{aligned} \tag{39}$$

where $A_0A_2=A$ as foresaid. By resolving Eq. (39) for x, y and u, we will get the characteristic equation.

$$AR+z^{-d}BS=0$$

A proper selection of the term or filter T_1 can improve the robustness. Signals with certain frequencies can be suppressed by requiring that the polynomial $S(z^{-1})$ vanishes at corresponding values of z. There is no steady-state error when the load disturbance is a step for the existence of an integrator. That is $R(1)=0$. Due to the latter, the present section can be regarded as another proof of Theorem 1.

Notice that the obtained controller is dependent on prediction horizon and on delay as well if $d>1$. It is not obvious whether there is virtue for this distinguishing feature. Simulations show that a small N is enough to stabilize a system and that the sensitivity of a controlled system is closely related to prediction horizon.

8. CONCLUSION

According to simulations, the ARMA model based GPC presented here can deal with a great variety of processes, stable and unstable ones as well as those of nonminimum phase. The predictor formulated for unstable process can be also used for stable ones as well as for integrating ones. However, there is difference with respect to the dynamic response of the controlled system to disturbances and to uncertainties when different predictor is used. To say roughly, a system controlled by a controller based on the predictor for unstable processes is less robust but more rapid to recover from a disturbance. The derived relationship between GPC and RST controllers indicates how to improve the controller if it is needed. There are potential applications of proposed controllers in the case of poor estimation of process delays. All the matrixes involved $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{M}$, can be calculated recursively. Present method presents difficulties when it is used to control unstable and simultaneously nonminimum phase processes. Some of them can be controlled with large N and $N_u \geq 2$.

REFERENCES

Camacho, E.F., Bordons, C., 2007. *Model Predictive Control.* London: Springer-Verlag

Clarke, D.W., 1988. Applications of Generalized Predictive Control to Industrial Processes. *IEEE Control System Magazine,* 122:49-55.

Clarke, D.W.,Mohtadi, C.,and Tuffs, P.S., 1987. Generalized Predictive Control. Part I. The Basic Algorithm. *Automatica,* 23(2):137-148.

Clarke, D.W.,Mohtadi, C.,and Tuffs, P.S., 1987. Generalized Predictive Control. Part II. Extensions and Interpretations. *Automatica,* 23(2):149-160.

Cutler, C. R., and Ramaker, B.C., 1980. Dynamic Matrix Control- A Computer control Algorithm. In *Automatic Control Conference, San Francisco.*

Richalet, J. Rault, A., Testud, J. L., and Papon, 1976. Algorithmic Control of Industrial Processes. In 4th IFAC *Symposium on Identification and System Parameter Estimation. Tbilisi* USSR.

Richalet, J., Rault, A., Testud, J.L., and Papon., J., 1978 Model Predictive Heuristic Control: Application to Industrial Processes. *Automatica,* 14(2):413-428.

Richalet, J., Abu el Ata-Doss, S., Arber, C., Kuntze, A. Jacubash, H.B., and Schill, W., 1987. Predictive Functional Control. Aplication to Fast and Accurate Robots. In *Proc. 10th IFAC Congress, Munich.*

Ydstie, B.E., 1984. Extended Horizon Adaptive Control. In *Proc.9th IFAC World Congress, Budapest, Hungary.*

DIAGNOSIS OF PEMFC BY USING STATISTICAL ANALYSIS

Zhongliang Li[a][c], **Rachid Outbib**[a], **Daniel Hissel**[b][c], **Stefan Giurgea**[b][c]

[a] Laboratorie des Sciences de l'Information et des Systemes (LSIS), University of Aix-Marseille, France
[b] FEMTO-ST (UMR CNRS 6174), ENISYS Department, University of Franche-Comte, France
[c] FCLAB (Fuel Cell Lab) Research Federation, FR CNRS 3539, rue Thierry Mieg, 90010 Belfort Cedex, France

zhongliang.li@lsis.org, rachid.outbib@lsis.org, daniel.hissel@univ-fcomte.fr, Stefan.Giurgea@utbm.fr

ABSTRACT

Fault diagnosis, especially on-line fault diagnosis is an essential issue for practical application of Polymer Electrolyte Membrane Fuel Cell (PEMFC) system. This paper proposes a diagnosis approach for PEMFC to handle the flooding fault which is considered to be a common fault. In this procedure, both fault detection and fault isolation are considered. For fault detection, the statistical characters of cell voltages distribution of a 20-cell PEMFC stack are analyzed. Parameters for describing voltage distribution characters are extracted. After that, a subset of parameters is selected in the orientation that the definition of fault is as correct as possible. A popular clustering methodology named K-means clustering (KMC) is adopted to make definition of flooding fault zone. For fault isolation, Support vector machine (SVM) classifier is trained to handle the cell voltages constructed vectors. Two different causes of flooding: increasing air humidity and decreasing of stack temperature can be discriminated by the classifier with a high correctness.

Keywords: PEMFC, diagnosis, flooding, voltage distribution, fault isolation

1. INTRODUCTION

Environmental issues have increased the demand for less polluting energy generation technologies. Developing fuel cell generator seems to be a possible solution, because they can offer substantially lower emissions, particularly of CO_2. Among different kinds of fuel cell, the PEMFC has been drawing more attention because of its high efficiency and its high power density (Wahdame et al. 2008).

However, to be widely adapted to transportation and stationary applications, the reliability and durability have to be improved. Fault diagnosis is an efficient solution to achieve this: fault diagnosis can be combined with control regulation to avoid the serious problem of system (Samy et al. 2011), it is also helpful to understand the physical and chemical mechanism (Wasterlain et al. 2011), so as to speed up the development cycle of new technologies. In addition, fault diagnosis can improve user support and acceptance by reducing down time (Tian et al. 2008).

In PEMFC diagnosis, several points must be satisfied: For transportation application of PEMFC, which aims at minimizing the embedded instrumentation to improve the reliability and decrease the cost, easy-to-monitor parameters are highly desired. For on-line diagnosis, computing time must be restricted so as to achieve diagnosis calculation within a sample time cycle. For control oriented diagnosis, fault isolation, which is to point the type of fault and its location, is an important component other than fault detection (Samy et al. 2011).

Some literatures have provided several fuel cell stack and system diagnostic methods. Physical modeling is an intuitional way to realize the aim of diagnosis (Escobet et al. 2009), however, parameters estimation is a barrier in PEMFC system. Some other works have been done to overcome the shortage of physical modeling. G. Tian et al, see (Tian et al. 2010), developed a methodology based on the analysis of the Open Circuit Voltage (OCV) in order to detect leakage fault, but it is not for on-line diagnosis. In (Yousfi Steiner et al. 2011), N. Yousfi Steiner et al. proposed neural networks model based diagnosis procedure to handle water management issues, but a series of variables, including pressures, are needed to be monitored.

In (Hissel et al. 2004), D. Hissel et al. proposed a fuzzy diagnostic model tuned by genetic algorithm, which is used to diagnose drying of membrane and accumulation of N_2/H_2O in the anode compartment. L. Alberto et al, in paper (Alberto et al. 2008), proposed an approach based on Bayesian networks, which can handle four types of faults in PEMFC system. In (Steiner et al. 2011), N. Yousfi Steiner et al. proposed a wavelet package translating methodology, flooding fault can be detected by analyzing the behavior of fuel cell stack voltage. J. Hua et al., see (Hua et al. 2011), proposed a multivariate statistical method, in which faults can be detected by analyzing principal components. Although above approaches can efficiently achieve the tasks of diagnosis, most of the literatures consider the fuel cell stack as integration. However, fuel cell stack in practice are always composed by series fuel cells, the behaviors of cells are different actually, even in normal operating state. This can be observed from the distribution of individual cell voltages (Hernandez et al.

2006). For this reason, assuming fuel stack as one block constituted by homogeneous cells seems to be arbitrary, additionally, the analysis of the statistical features of individual cells can support us more information for diagnosis.

The goal of this work is to extend the previous studies (Hernandez 2006), in which some statistical qualities of cell voltages distribution and their relations to flooding fault have been found. In this paper, methodologies are used to improve the diagnosis performance and make the previous work more completed. Summarily, the proposed approach is suitable for real-time diagnosis, and is based on data. Specifically, in this paper, four statistical parameters are extracted to analyze the distribution of cell voltages. Fault definition methodology is proposed, in which feature selection and KMC are adopted. Fault isolation is also accomplished with a high correctness by training a SVM classifier, such that the two causes of flooding fault can be discriminated.

This paper is organized as follows: In section 2, the experiments of PEMFC are introduced. In section 3, the diagnosis approach is expounded, including detailed presentations of methodologies and their using in fault detection and isolation. The results of diagnosis are given in the next section. Finally, we conclude the work in section 5.

2. DESCRIPTION OF EXPERIMENTS

The data used are collected from a 20-cells PEMFC stack, whose nominal output power is 500 W. Flooding experiments with a certain load (40A) were carried out on the test bench. Some fault controls can result in the flooding inside the fuel. Here, the flooding was induced by two ways: First, increasing the inlet air humidity in order to favor water condensation. Second, the water vapor condensation in the fuel cell is caused by decreasing the temperature of fuel cell stack. The cell voltages were sampled in the flooding evolutions. Figure 1 and figure 2 show cell voltages in flooding evolutions caused by inlet air humidity variation and stack temperature variation. The goal of this paper is to realize fault diagnosis by statistical analysis of cell voltages distribution in these two evolutions.

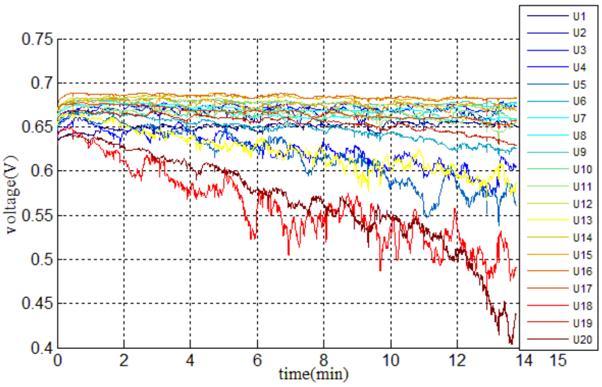


Figure 1: Cell voltages in flooding evolution caused by increasing the inlet air humidity

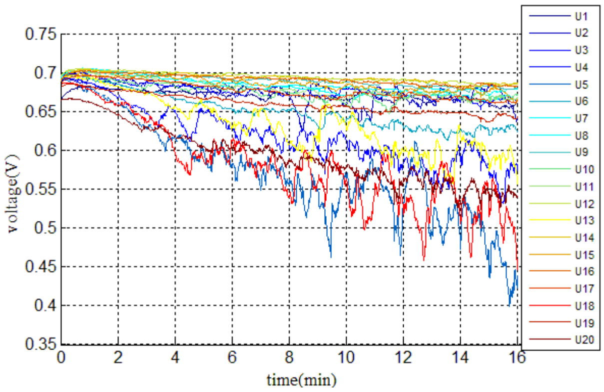


Figure 2: Cell voltages in flooding evolution caused by decreasing the stack temperature

3. THE PROPOSED APPROACH

3.1. Principle of the approach

The objective of this paper is to realize online diagnosis of PEMFC. The flow chart of diagnosis, showed in figure 3, is as follows: after the real-time samples (just individual cell voltages in this paper) are imported, some diagnosis oriented features are computed. The fault (flooding fault) can be detected if the features of the sample are in the fault zone. After fault is detected, fault isolation can be achieved by a SVM classifier. By using the fault isolation, two fault causes of flooding were been considered: the variation of inlet air humidity (cause1) and the variation of the stack temperature (cause2).

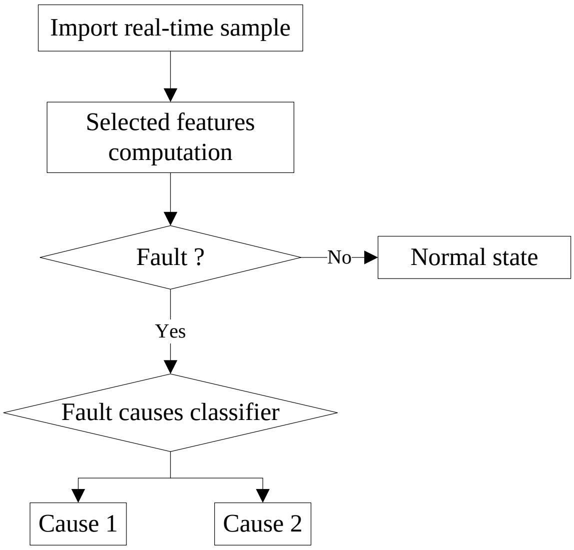


Figure 3: On-line diagnosis flow chart

In the online diagnosis flow chart, some training procedures need to be accomplished off-line. As the figure 4 shows, for the fault detection (the left figure), several statistical parameters are extracted firstly, then, a diagnosis oriented subset of parameters is selected. After that, KMC is adopted for fault definition. For fault isolation (the right figure), vectors composed by cell voltages are constructed firstly, then, SVM classifier is trained based on the vectors.

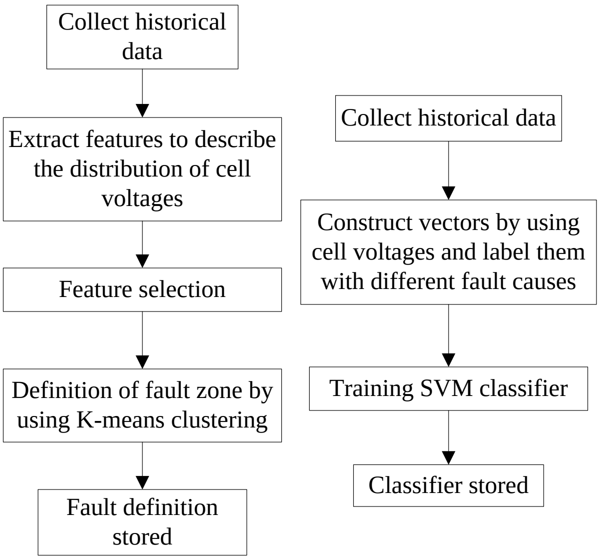


Figure 4: Flow chart of off-line training procedures

3.2. Description of cell voltages distribution

In order to describe the statistical characters of cell voltages distribution, some description variables are extracted.

3.2.1. The hypothesis test of Normal distribution

Normal distribution is a most widely used distribution model. A specific normal distribution can be described by two parameters: μ, the mean or the expectation, and σ^2, the variance. In order to extract the features to describe the distribution of the cell voltages, the hypothesis test of normal distribution is carried out, if the cell voltages follow the normal distribution, μ and σ^2 would be extracted; otherwise, some other parameters would be seeked other than these two parameters.

The *Lilliefors test* is used to test the null hypothesis that data come from a normally distributed population, when the null hypothesis does not specify which normal distribution (Lilliefors 1967).

After hypothesis test, the null hypotheses of 84.5% data samples in the flooding evolution are rejected with significance level $\alpha = 0.05$. Hence, it is arbitrary to consider the cell voltages are from normal distribution.

3.2.2. Cell voltages distribution features

The mean μ and the variance σ^2 are significant descriptive measures that locate the center and describe the dispersion of probability density function. However, they do not provide a unique characterization of the distribution. To better approximate the probability distribution of a random variable, *skewness* and *kurtosis*, two parameters that are usually used for describing a distribution statistical characters (Ramachandran and Tsokos 2009), are considered other than μ and σ^2.

Let X be a random variable, *Skewness*, which is denoted by *Sk*, is defined as

$$Sk = \frac{E[(X-\mu)^3]}{\sigma^3} \qquad (1)$$

where E denotes expected value, μ and σ^2 are mean value and variance of X . *Skewness* is used as a measure of the symmetry of a density function about its mean. Recall that a distribution is symmetric if it looks the same to the left and right of the center point. If $Sk = 0$, then the distribution is symmetric about the mean, if $Sk > 0$, the distribution has a longer right tail, and if $Sk < 0$, the distribution has a longer left tail. The *Skewness* of a normal distribution is zero.

Kurtosis, which is denoted by *Ku*, is defined as

$$Ku = \frac{E[(X-\mu)^4]}{\sigma^4} \qquad (2)$$

Kurtosis is a measure of whether the distribution is peaked or flat relative to a normal distribution. *Kurtosis* is based on the size of a distribution's tails. A high *kurtosis* distribution has longer, fatter tails. A low *kurtosis* distribution has shorter, thinner tails. A distribution which has the same *kurtosis* as a normal distribution is known as *mesokurtic*. It is known that the kurtosis for a standard normal distribution is $Ku = 3$. A distribution with positive excess of 3 is called *leptokurtic*, while a distribution with negative excess of 3 is called *platykurtic* (Ramachandran and Tsokos 2009).

Four parameters: μ, σ^2, *Sk* and *Ku* of cell voltages in the flooding evolutions caused by increasing the humidity of inlet air and decreasing the temperature of stack are as figure 5 and figure 6.

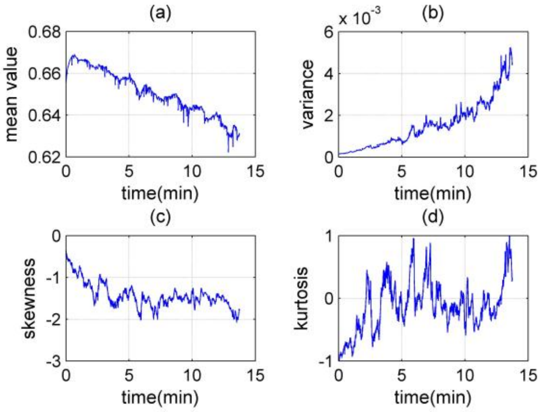


Figure 5: Four statistical variables of cell voltages in flooding evolution caused by humidity variation

From the figures, some intuitional characteristics can be observed:

1. Mean value decreases and variance increases in the flooding evolution.
2. *Skewness* is always negative, which means the probability density distribution function of cell voltages has a longer left tail.
3. *Kurtosis* is always with a negative excess of 3 which means that the probability density function is *leptokurtic*.
4. *Skewness* and *kurtosis* show no monotonous variation in the flooding evolution.

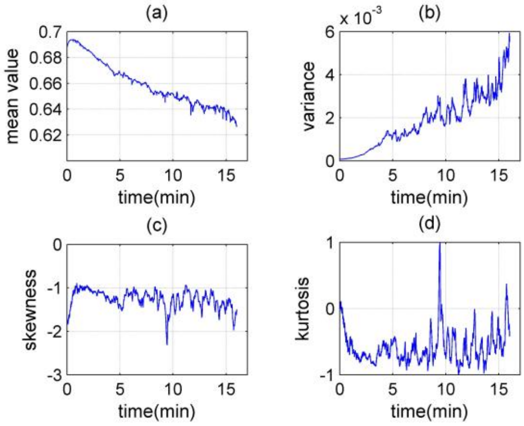


Figure 6: Four distribution parameters in flooding evolution caused by temperature variation

Until now, four typical parameters to describe the properties of cell voltages distribution have been extracted. Fault diagnosis will be preceded based on the four parameters.

3.3. Feature selection

Although all these four parameters: μ, σ^2, *Sk* and *Ku* are useful for cell voltages distribution description, for the fault diagnosis, just a subset of the parameters has a significant role. The goal of feature selection procedure is to find and to select the subset which improves the performance of fault diagnosis. In order to evaluate the performance of feature selection quantificationally, a criterion denoted as *Cr* is defined. The number of non-void subsets of dataset $\{\mu, \sigma^2, Sk, Ku\}$ is $2^4-1 = 15$. The criterions of each subset are calculated and compared, the parameters of the subset which has the most optimal criterion will be chosen as the final features for diagnosis.

The computation of the criterion is connected to the definition of the fault. In the flooding evolutions, it is considered that the vapor condenses as time, which results in the intensive to flooding. So if we allocate the samples in flooding evolution into two clusters (normal cluster and fault cluster), the normal cluster must be the one of which the major part is in the front of the time axis, while the major part of fault cluster is in the tail of the time axis. Consequently, the performance of the clustering can be evaluated by the proportion of overlaps of the two clusters in the time domain. More specifically, the evaluation criterion of fault definition is defined as

$$Cr = N_{error}\left(\frac{1}{N_1}+\frac{1}{N_2}\right) \tag{3}$$

where N_1 and N_2 are the numbers of samples in normal cluster and fault cluster, N_{error} is the number of samples in normal cluster whose indexes exceed N_1.

3.4. Fault definition

3.4.1. KMC

Clustering is an unsupervised learning algorithm which allocates data points to a certain number of clusters. It is used widely to collect similar data. So clustering is suited for fault definition. KMC is one of the simplest and most popular to solve clustering problem (Macqueen 1967).

The KMC problem aims at allocating a dataset $\{x_1, x_2, ..., x_N\}$ into *C* clusters, which are denoted as $\omega_1,\ldots,\omega_C$, so as to minimize the within-cluster sum of squares which is defined as

$$J = \sum_{i=1}^{C}\sum_{x_n\in\omega_i} \| x_n - \bar{x}_i \|^2 \tag{4}$$

where $\bar{x}_i$ is the center of cluster ω_i,

$$\bar{x}_i = \frac{1}{N_i}\sum_{x_n\in\omega_i} x_n \tag{5}$$

N_i is number of samples in cluster ω_i.

The implementation of KMC is as follows (Güneş et al. 2010):

Stage1: Choose *C* initial cluster centers $\bar{x}_1,...,\bar{x}_C$ randomly from the *N* points $\{x_1, x_2,\ldots, x_N\}$.

Stage2: Assign point x_n to the cluster ω_i, $i = 1,\ldots,C$, if

$$\| x_n - \bar{x}_i \| < \| x_n - \bar{x}_j \|, \qquad j = 1,...,C, j \neq i \tag{6}$$

Stage3: Compute new cluster centers

$$\bar{x}_i^{new} = \frac{1}{N_i}\sum_{x_n\in\omega_i} x_n \tag{7}$$

Stage4: Repeat stage2 and 3 until the centroids no longer move. If

$$\| \bar{x}_i^{new} - \bar{x}_i \| < \varepsilon, \qquad i = 1,...,C \tag{8}$$

the computation process is terminated, otherwise back to stage2.

Although there is no guarantee of achieving a global minimum, the convergence of the algorithm is ensured (Yiakopoulos et al. 2011).

3.4.2. Application of KMC

KMC is used for fault definition, the goal of which is to divide "normal state" zone and "fault state" zone in the selected feature space. In this artical, fault definition is accomplished by two stages: Firstly, data in flooding evolution are labeled as "normal state" or "fault state". This is achieved by KMC procedure. After that, different zones are gotten based on the centroids of clusters. The points of the boundary between two zones are equidistant from their centroids.

3.5. Fault isolation

The tasks of fault diagnosis contain not only the fault detection, but also fault isolation which is to fix the position and the causes of faults. Fault isolation is beneficial to modify the control in order to slack or eliminate the fault, also it play a positive part in the design and manufacture of products. In this article, the flooding fault is formed by two causes: increasing of the humidity of the inlet air and decreasing of stack temperature. The objective of this part is to isolate these two causes. The fault data can be obtained after fault definition, and data of these two situations have been labeled, hence, the problem is a classification problem.

3.5.1. SVM

SVM is a classification method developed by V. Vapnik (Vapnik 1999) and has been widely applied last two decades. As figure 7 shows, the general description of a 2-class SVM is: SVM establishes the optimal separating hyperplane that allocates the majority of points of the same class in the same side, whilst make the smallest distances between the two classes and the hyperplane, which is called margin, to be maximized. The points of the two classes which create the hyperplane are called support vectors.

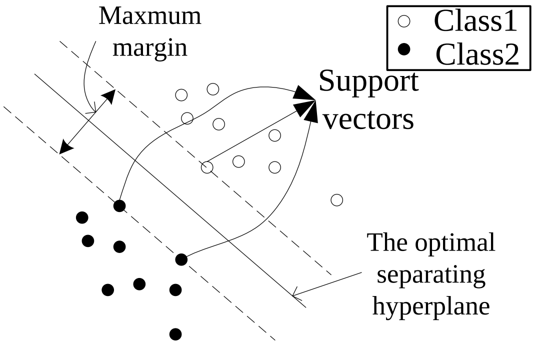


Figure 7: Classification of two classes using SVM

Given the training data $\{\mathbf{x}_n, y_n\}$, n = 1,..., N. where the $\mathbf{x}_n \in \mathbf{R}^M$ and $y_n \in \{1;-1\}$. $\mathbf{x}_n$ is an input vector and y_n indicates the class of $\mathbf{x}_n$. It is possible to determine the hyperplane separating data as

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0 \tag{9}$$

where $\mathbf{w} \in \mathbf{R}^M$ is normal to the hyperplane, b is a bias. A distinct separating hyperplane should satisfy the constraints

$$f(\mathbf{x}_n) \geq 1 \quad \text{if } y_n = 1 \tag{10}$$

$$f(\mathbf{x}_n) \leq -1 \quad \text{if } y_n = -1 \tag{11}$$

Support vectors lie on the planes $\mathbf{w}^T\mathbf{x}_n + b = \pm 1$. The margin is $2/\|\mathbf{w}\|$, consequently, the problem can be converted to find the minimum of $\|\mathbf{w}\|^2/2$. Taking into account the misclassification errors with slack variables ξ_n and the error penalty D, the optimal hyperplane separating the data can be obtained as a solution to the following optimization problem:

$$\min \quad D\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 \tag{12}$$

subject to

$$\begin{cases} y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1 - \xi_n \\ \xi_n > 0 \end{cases} \tag{13}$$

After introducing kernel functions to extend the SVM to nonlinear classification domain, the optimization problem (12-13) can be converted to a dual quadratic optimization problem as

$$\max L(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m k(\mathbf{x}_n, \mathbf{x}_m) \tag{14}$$

subject to

$$\begin{cases} 0 \leq a_n \leq D \\ \sum_{n=1}^{N} a_n y_n = 0 \end{cases} \tag{15}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ satisfies $\mathbf{w} = \sum_{n=1}^{N} a_n y_n \mathbf{x}_n$, and $k(\mathbf{x}_n, \mathbf{x}_m)$ denotes kernel function on $\mathbf{x}_n$ and $\mathbf{x}_m$.

The associated class label y of a new vector $\mathbf{x}$ is calculated as follow

$$y = sign\{\sum_{n=1}^{N} a_n y_n k(\mathbf{x}_n, \mathbf{x}) + b\} \tag{16}$$

Three typical kernel functions are presented in table1.

Table 1: Three typical kernel functions

Kernel	$k(\mathbf{x}_i,\mathbf{x}_j)$
Linear	$\mathbf{x}_i^T\mathbf{x}_j$
Polynomial	$(\mathbf{x}_i^T\mathbf{x}_j + 1)^d$
Gaussian	$\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/c)$

Direct solution of the quadratic programming problem (14-15) using traditional techniques is often infeasible due to the demanding computation and memory requirements, a more practical approach sequential minimal optimization (SMO) is used (Platt 1998).

The basic SVM is a binary classifier. There are a couple of approaches to expand this methodology to multi-classification situation such as "one-against-all", "one-against-one", and directed acyclic graph SVM (DAGSVM). In (Hsu and Lin 2002), C. Hsu et al. made a conclusion and a comparison among different approaches.

3.5.2. Application of SVM

In order to find the differences between humidity variation caused flooding and temperature variation caused flooding, cell voltage evolutions in the two situations are presented by 3-D figures as figure 8 and figure 9.

The index number of cell is considered as representative of the distance from air entrance to the cell. From the figures, it can be seen that the cells near to the exit are more likely to be fault in the situation of humidity variation; while the cells near to air entrance are more likely to be fault in the temperature variation situation. Hence, the spatial distributions of cell voltages in the two situations are various. If we denote the cell voltages from air entrance to air exit as a vector $\mathbf{v} = [v_1, v_2, \ldots, v_{20}]^T$, then, the spatial distribution information can be kept by the indexes of elements in the vector. Consequently the cell voltages constructed vectors are used as training data for SVM.

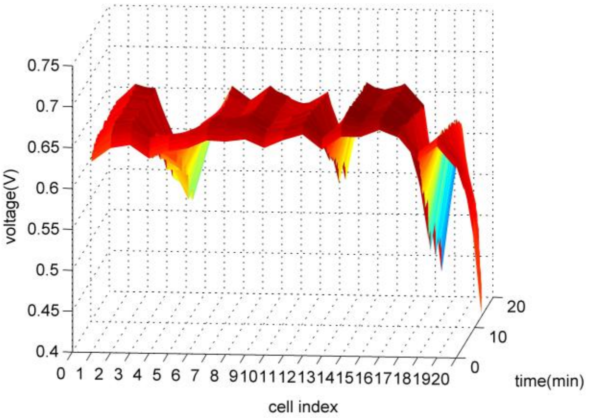


Figure 8: Cell voltages spatial distribution in the flooding evolution caused by humidity variation

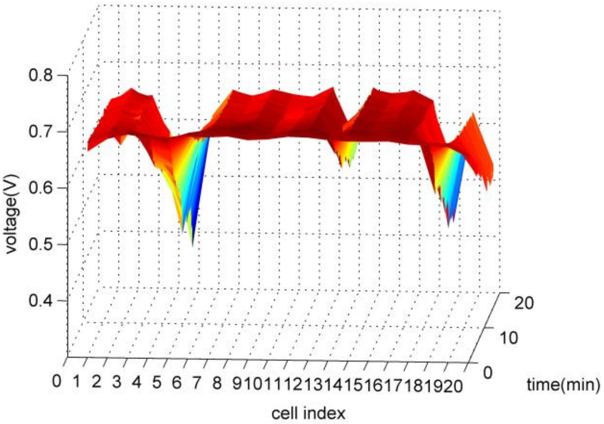


Figure 9: Cell voltages spatial distribution in the flooding evolution caused by temperature variation

3.5.3. Evaluation of the classifier

To evaluate the performance and of the SVM classifier, and also to estimate how accurately it will perform in practice, a popular cross-validation methodology which named K-fold cross-validation is used (Kohavi 1995).

In K-fold cross-validation, the total data is randomly divided into K subsets. Of the K subsets, $K-1$ are chosen as training data and the rest one subset is used for test the classifier. The training and test process is then repeated K times, so that each of the K subsets is used once as the test data. The averaged test result is then obtained to evaluate the classifier. The advantage of K-fold cross-validation ensures that all data are used for both training and test, each observation is used for test exactly once.

4. RESULTS

4.1. Feature selection result

Considering the fault detection, the data in two experiments are analyzed together. The four statistical parameters presented in 3.2 are normalized to $[-1,1]$ firstly. After computing and comparing the criterions of 15 different subsets, the criterion of the subset composed by μ and σ^2 is the smallest of all. In other word, the combination of these two parameters is most suitable for fault diagnosis.

4.2. Fault definition result

KMC is carried out in the 2-dimension space composed by two features: μ and σ^2. The clustering result is as figure 10 shows, the criterion variable Cr of clustering is 0.048, which indicates a good performance of clustering.

After K-means clustering procedure, two centroids are obtained. The “fault zone” and “normal zone” can be divided by drawing a line in the middle of these two centroids as figure 10 and 11 show. For a new sample, it is fault or not can be told by its location is above or below the line.

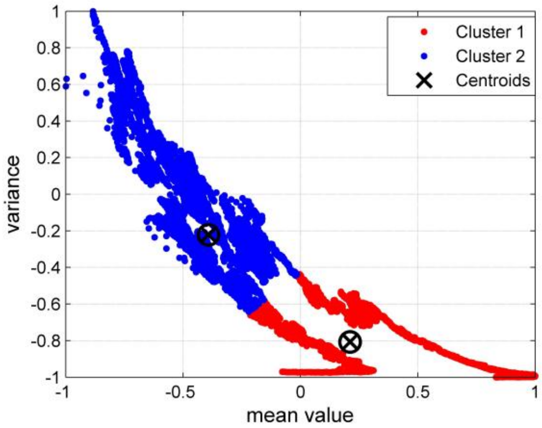


Figure 10: Cluster result on 2-dimenstion space of μ and σ^2

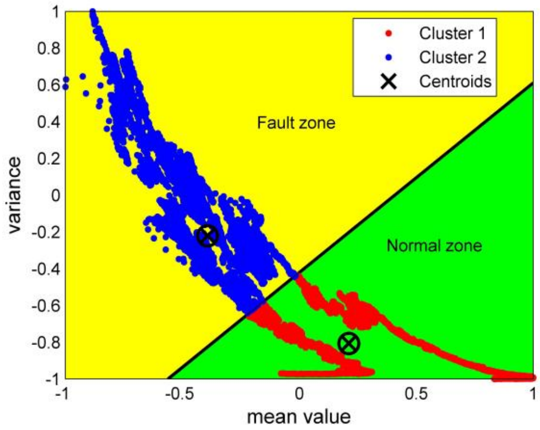


Figure 11: Fault zone and normal zone partition

4.3. Fault isolation result

The SVM is carried out on cell voltages composed vectors, Gaussian kernel (as in table 1) is selected with parameter $c = 0.5$. Error penalty parameter D is chosen to be 10^4 after several attempts.

In the validation procedure, K-fold cross-validation is used, the number of folds K is set to be 10. On account of that the different folds are chosen randomly, 5 independent tests are preceded. The average error rates of all the tests are kept to be 0. Hence, the performance of SVM for discriminating the causes of flooding is excellent.

After training the total 628 fault vectors, only 9 support vectors which are used for classify new vectors are obtained. Consequently, the computation time will be short enough for online diagnosis.

5. CONCLUSION

This paper presents a fault diagnosis approach for PEMFC stack based on statistical analysis. For fault detection, four statistical parameters of cell voltages: μ, σ^2, *Sk*, *Ku*, are extracted to describe the distribution. Feature selection and KMC are used to define fault zone. For fault isolation, SVM is adopted to classify the cell voltages composed vectors into different fault cause classes. The proposed diagnosis approach can detect flooding fault in fuel cell stack and efficiently classify the causes of flooding fault into two classes: increasing the humidity of inlet air and decreasing the temperature of fuel cell stack. The proposed approach takes account of otherness of cells in the fuel cells stack, and is suited for online diagnosis.

REFERENCES

Alberto, L. et al., 2008. On-line fault diagnostic system for proton exchange membrane fuel cells. *Journal of Power Sources*, 175, pp.419-429.

Escobet, T. et al., 2009. Model-based fault diagnosis in PEM fuel cell systems. *Journal of Power Sources*, 192(1), pp.216-223.

Güneş, S., Polat, K. & Yosunkaya, Ş., 2010. Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Systems with Applications*, 37(12), pp.7922-7928.

Hernandez, A., 2006. *Diagnostic d'une pile a combustible de type PEFC*. Thesis (PhD). UNIVERSITE DE TECHNOLOGIE DE BELFORT MONTBELIARD.

Hernandez, A. et al., 2006. Fuel cell fault diagnosis : A stochastic approach . In *Industrial Electronics, 2006 IEEE International Symposium on*. pp. 1984-1989.

Hissel, D., Péra, M.C. & Kauffmann, J.M., 2004. Diagnosis of automotive fuel cell power generators. *Journal of Power Sources*, 128, pp.239-246.

Hsu, C.-wei & Lin, C.-jen, 2002. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transaction on Neural Networks*, 13(2), pp.415-425.

Hua, J. et al., 2011. Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method. *International Journal of Hydrogen Energy*, pp.1-10.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. pp. 1137-1143.

Lilliefors, H.W., 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, pp.399-402.

Macqueen, J., 1967. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. In *Proceedings of 5th Berkeley symposium on mathematical statistics and probability*. pp. 281-297.

Platt, J.C., 1998. Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines. *Technical Report MSR-TR-98-14, Microsoft Research*, pp.1-21.

Ramachandran, K.M. & Tsokos, C.P., 2009. *Mathematical Statistics with Applications*, Academic Press.

Samy, I., Postlethwaite, I. & Gu, D.-W., 2011. Survey and application of sensor fault detection and isolation schemes. *Control Engineering Practice*, 19(7), pp.658-674.

Steiner, N.Y. et al., 2011. Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform. *International Journal of Hydrogen Energy*, 36(1), pp.740-746.

Tian, G. et al., 2008. Diagnosis methods dedicated to the localisation of failed cells within PEMFC stacks. *Journal of Power Sources*, 182, pp.449-461.

Tian, G. et al., 2010. Identification of failed cells inside PEMFC stacks in two cases: Anode/cathode crossover and anode/cooling compartment leak. *International Journal of Hydrogen Energy*, 35(7), pp.2772-2776.

Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 10(5), pp.988-99.

Wahdame, B. et al., 2008. Analysis of a Fuel Cell Durability Test Based on Design of Experiment Approach. *IEEE Transaction on Energy Conversion*, 23(4), pp.1093-1104.

Wasterlain, S. et al., 2011. Development of new test instruments and protocols for the diagnostic of fuel cell stacks. *Journal of Power Sources*, 196, pp.5325-5333.

Yiakopoulos, C.T., Gryllias, K.C. & Antoniadis, I. a., 2011. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 38(3), pp.2888-2911.

Yousfi Steiner, N. et al., 2011. Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling. *International Journal of Hydrogen Energy*, 36(4), pp.3067-3075.

GENERALIZED CONTROLLED SWITCHED BOND GRAPH STRUCTURES WITH APPLICATIONS TO ABRUPT FAULT MODELING

Nacusse, Matías A.$^{(a,b)}$ **and Junco, Sergio J.**$^{(a)}$

$^{(a)}$ LAC, Laboratorio de Automatización y Control, Departamento de Control, Facultad de Ciencias Exactas, Ingeniería y
Agrimensura, Universidad Nacional de Rosario, Ríobamba 245 Bis – S2000EKE Rosario – Argentina.
$^{(b)}$ CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas.

$^{(a,b)}$nacusse@fceia.unr.edu.ar, $^{(a)}$sjunco@fceia.unr.edu.ar

ABSTRACT

This paper presents two controlled switched Bond Graph structures with fixed causality. The first is the Switchable Structured Bond, an interconnection structure extending the idea of switchable bonds that can represent all commutation modes between two subsystems. The second is called Generalized Switched Junction Structure and can represent all the interconnections enforced by commutations involving bond graph elements around standard 0- and 1-junctions. Both structures, defined with fixed causality for modeling and simulation purposes, can be internally represented with standard bond graph elements. To keep fixed the causality assignment even under switching, some algebraic constraints are added to the equation set of the switched structures, which in the Bond Graph domain can be represented with residual sinks. Both structures preserve causality under ideal (zero transition time) switching. Adding parasitic components as an alternative, non-ideal, approximate approach to switching can also be accomplished with the second structure just performing a minor modification on its internal implementation with basic bond graph components.

Keywords: Bond Graphs, Switched Systems, Switched Structures, Residual Sinks, Abrupt Faults.

1. INTRODUCTION

Frequently in engineering abrupt changes in physical systems are considered to occur instantaneously. This is mainly due to the fact that the behavior the engineer is interested in has a time scale much bigger than that of the abrupt change, and that the details inside the time window of this change are not relevant to the behavior under study. Thus, ignoring them results in saving time and effort. As this practice departs from the assumptions of continuity and smoothness underlying classical physics, it requires special modeling and simulation (M&S) and analysis tools to handle the systems it yields, see (Mosterman and Biswas, 1998) for a sound discussion of M&S issues related to this problem.

Bond Graphs (BG) constitutes a graphical energy-based modeling tool originally conceived to represent the continuous dynamics of physical systems (Karnopp et al. 2000, Borutzky 2010). Many tools have been proposed in the BG domain to extend its basic component set in order to also model ideal switching processes: *MTFs modulated with* gain taking values over the set $\{0, 1\}$ (Asher 1993, Dauphin-Tanguy and Rombaut 1997); an *ideal switch* as a new bond graph element (Strömberg, Top, and Söderman 1993); a switch as an ideal current source and a voltage source (Demir and Poyraz 1997); *switchable bonds* (Broenink and Wijbrans 1993); *con-trolled junctions* (Mosterman and Biswas 1995, 1998); *Petri nets* to represent discrete modes and transition between them (Allard, Helali, Lin, and Morel 1995); and the *SPJ* or *Switched Power Junction* formalism. See (Umarikar and Umanand 2005) for an introduction to the latter modeling technique and a brief description and discussion of the pros and cons of all the others.

The results in this paper are twofold. First, after a critical review of the *switchable bond* concept (Broenink and Wijbrans 1993), a modification of it, called *S-Bonds*, is proposed. Second, it is considered how switching affects and modifies structures originally represented with elementary **0-** and **1-**junctions. The consequence of this is the introduction of two new BG components, called Generalized Switched Junction Structure (GSJ), allowing to represent all the structural changes induced by switched interconnections among the elements around the original **0-** and **1**-junctions.

Switching in a physical system can be considered under different perspectives. The research presented here was conducted in the BG-domain from a system dynamics point of view. In order to fix ideas, consider the standard state-space description $dx/dt = f(x, u; p)$, where x and u are the state- and input-vectors, and p a vector of system parameters. This model can be modified in different ways by switching, the mildest one being just a change in the values of the parameters (p-Before Switching changes into p-After Switching: $p_{BS} \rightarrow p_{AS}$) without further consequences. But more substantial modifications can occur, like changes in the vector field f defining the dynamics ($f_{BS} \rightarrow f_{AS}$) or, even more dramatic, changes in the set of state- (and/or input-) variables ($x_{BS} \rightarrow x_{AS}, u_{BS} \rightarrow u_{AS}$), with or without changes in the system order. Moreover, it could happen that the explicit standard form be not longer attainable and substituted by a differential-algebraic system, or more generally, a differential-implicit form. All these effects are related to changes in the causality assignment if using elements of the standard BG-set, so that any tool devised to modeling switching in the BG-domain must be able to somehow address this issue.

Some M&S software do not allow changes in the causality of the model during the simulation. A possible approach to solve this is duplicating the elements with changing causality. As each of these causality-alternating, duplicated elements represents in fact a unique physical phenomenon, this modeling approach is not Object Oriented Modeling (OOM) compliant.

Also, the causal constraints at the origin of the causality switching of some elements can be broken adding some parasitic BG components, and models can be obtained with fixed causality. However these parasitic components increase the order of the model and make it stiff, which, practically, is not convenient for simulation purposes and, conceptually, enters in conflict with the ideal switch approach chosen to model the commutations. Besides this, the parasitic components are usually not related to the physical system from a macroscopic point of view, which complicates the task of parameterizing them.

Aiming at simulation with fixed causality, the causality changes are avoided in this paper following an approach already presented in (Nacusse and Junco 2010): residual sinks (Borutzky 2010) are introduced in the model to break the causality constraints produced by switching. The residual sink component injects the necessary effort or flow in order to make vanish the power conjugated variable into the sink. This bond graph component adds an algebraic constraint which implies that a DAE system describes the system dynamics. The constraint must be numerically solved at each integration step, through an explicit calculation if the constraint can be solved analytically off-line prior to the simulation, otherwise implicitly, with the consequent increment of the computational cost.

Another problem associated to the change of causality between modes is the possible appearance of discontinuities or jumps in the state trajectories, which is solved with the re-initialization of the storage elements after a switching occurrence (Nacusse and Junco 2010).

The paper is organized as follows. Section 2 presents some background results on switchable bonds, switched power junctions and residual sinks, employed in what follows. Section 3 present the Structure Switchable Bond or S_Bonds and the GSJ or Generalized Switched Junction Structures as the main results of the paper, and illustrate them with switching problems in an electric circuit. Section 4 applies the new results to two classic switched power electronic converters. Section 5 addresses the application of the GSJ structure to a fault modeling problem in a two tank hydraulic system. It is stressed that only fault modeling is addressed and not FDI. Finally, Section 6 presents some conclusions.

2. BACKGROUND RESULTS

This section recalls the basics on SPJs, residuals sinks and switchable bonds, tools which are going to be used further in this paper.

2.1. SPJ: Switched Power Junctions

The SPJs are generalizations of the standard 0- and 1-junctions (Umarikar and Umanand 2005). They are represented as receiving the effort (0_s) or flow (1_s) information from *more than one bond*. To prevent from the causal conflicts this would otherwise imply, control signals, taking values over the set {1, 0}, are added to the new elements. Only one of these signals is allowed to have the value 1 at a given time instant, the remaining being zero. In this way, only one of the bonds imposing effort (0_s) or flow (1_s) is selected (i.e., becomes operative) and the value zero is imposed to the power co-variables of the remaining bonds, which results in their disconnection.

Figure 1 shows the SPJs with causality assignment and eqs. 1 and 2 express the mathematical relationships -for the 0_s and the 1_s, respectively- among the power variables and the control signals U_i injected to select the appropriate bond. In (Junco et al. 2007) the SPJs have been interpreted in terms of the classical 0- and 1-junctions and MTFs modulated by a gain taking the values 0 or 1. In (Nacusse et al. 2008) the implementation of the 0_S and the 1_S as new standard elements of the 20sim basic library has been presented (available at http://www.fceia.unr.edu.ar/dsf/I&D/BG.html).

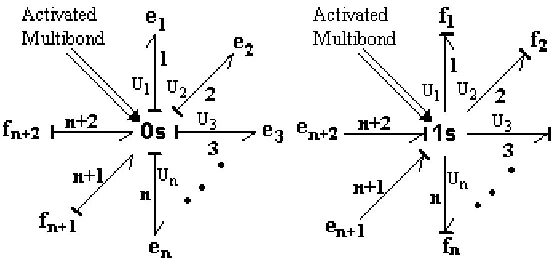


Figure 1. Switched Power Junctions with causality assignment.

$$\begin{aligned} &\mathrm{Effort} = U_1 e_1 + U_2 e_2 + \cdots + U_n e_n \\ &f_i = U_i(f_{n+1} + f_{n+2}) \quad ; \quad i = 1, \ldots, n \end{aligned} \tag{1}$$

$$\begin{aligned} &\mathrm{Flow} = U_1 f_1 + U_2 f_2 + \cdots + U_n f_n \\ &e_i = U_i(e_{n+1} + e_{n+2}) \quad ; \quad i = 1, \ldots, n \end{aligned} \tag{2}$$

The simple electrical circuit in Figure 2 illustrates how to use the SPJ technique.

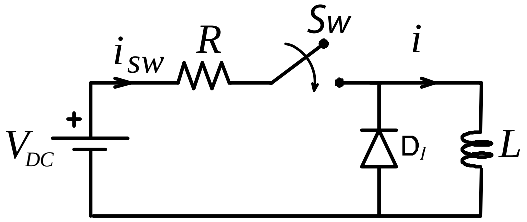


Figure 2. Switched electrical circuit

The circuit contains two switching elements, an ideal switch (an on-off commanded transistor, for instance) and a free-wheel diode, which have the complementary logic states {switch open, diode closed} and {switch closed, diode open}, so that only one control variable is necessary in the Switched BG (SwBG).

In the SwBG of Figure 3, the current commutation of the ideal switch is modeled with the 1_s and the source $\mathbf{S_f}$ (f≡0), whereas the voltage commutation at the diode is modeled with the 0_s plus the resistor **R** labelled *D1* (it models the diode's conduction state). Qualitatively it works as follows (consider eqs. (1) and (2) to get a more complete and precise quantitative description of this BG's behavior): the 1_s selects either the $\mathbf{S_f}$-bond below it (*m*=1, switch OFF) or the bond on its right (*m*=0, switch ON) to impose, respectively, zero current or the inductance current to the submodel to its left. The 0_s chooses the **R**(*D1*)-bond below it (*m*=1, switch OFF) or the bond on its left (*m*=0, switch ON) to impose, respectively, the voltage of the source-resistor series or the diode voltage-drop to the inductance. Summarizing, in this example, each SPJ chooses the bond below it (*m*=1, switch OFF) or, alternatively, both SPJs select the bond joining them (*m*=0, switch ON).

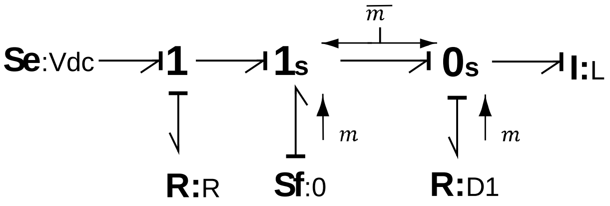


Figure 3. SwBG using SPJ of the switched circuit.

2.2. Switchable bonds

The *switchable bonds* presented in (Broenink and Wijbrans 1993) are controlled bonds commanded by a control signal m that can take the values 1 or 0 and indicates the presence or absence of the bonds in the BG model. The dashed power line indicates that this bond is only conditionally present.

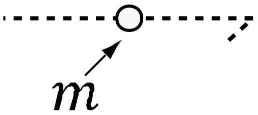


Figure 4. Switchable bond representation

This approach has some problems, caused by the fact that the boundary conditions on the adjacency of the switchable bonds are not always explicitly defined in all the switching modes (Strömberg 1994). This fact, known as the problem of the dangling junctions, is illustrated with the help of the SwBG in Figure 5, where the switches in the circuit of Figure 2 are modeled with switchable bonds. Again, the resistor **R**, labelled $D1$, models the diode's conduction state (note that this BG is not fully OOM-compliant, in the sense that the ideal switch, a single circuit element, has to be modeled with two switchable bonds). The problem arises when the switch is OFF and the switchable bonds commanded by m are disconnected (m = 0): the source Se and the resistor R receive each an undefined flow information, each from an otherwise disconnected 1-junction. There is no problem with the switchable bond commanded by $\overline{m}$, which connects the **I** and the **R**($D1$). Also the other circuit configuration (m = 1) is properly defined.

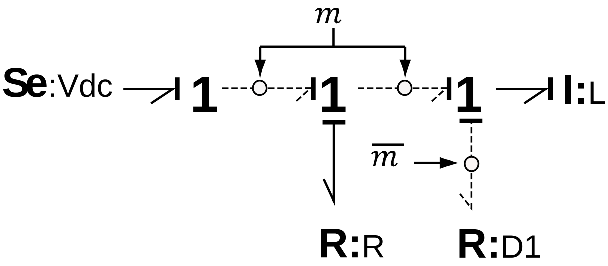


Figure 5. SwBG model of the switched electrical circuit with switchable bonds.

2.3. Residual sinks

Residual sinks are traditionally used to break causal conflicts in BG models yielding the same results as adding Lagrange multipliers (Borutzky 2010). This element injects its output variable, effort or flow, into the rest of the system, computed as to make vanish the power conjugated variable, the input into the sink.

A residual sink can be interpreted as an energy store where its parameter tends to zero. For example, an effort residual sink can be interpreted as a C element in integral causality. If the parameter C tends to zero, then $\dot{e}$ is determined by the algebraic equation $\Delta f = 0$.

$$C\dot{e} = \Delta f \tag{3}$$

Figure 6 shows the graphical representation of the effort and flow residual sink as in (Borutzky 2009).

Figure 6. Flow and effort residual sink

3. MAIN RESULTS

The Structured Switchable Bonds, S-bonds for short, are introduced in this section as an improvement of the switchable bonds, as well as the Generalized Switched Bond Graph Structures, or GSJ, as the main contributions of the paper.

3.1. S-bonds: Structured Switchable Bonds.

The S-bonds, which can be viewed as an extension of the plain switchable bonds presented by (Broenink and Wijbrans 1993), are introduced with the aim of remedying the problem of the dangling junctions previously discussed. To do this three control signals are necessary instead of just one. Indeed, with the help of the three control variables it is possible not only to determine the presence or absence of the switchable bond, but also to explicitly and univocally define the boundary conditions of the BG-elements adjacent to the switchable bond in each switching mode. The symbology adopted is shown in Figure 7.

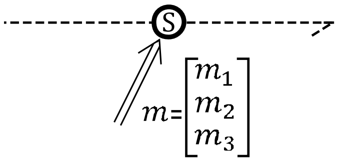


Figure 7. S-bond representation.

Without loss of generality, the behavior of the S-bond is explained with the help of Figure 8, where causality indicates that the effort e_b is imposed by Σ_a and the flow f_a is calculated by Σ_b. Besides the ground connection mode the S-bond enforces $e_b = e_a$ and $f_a = f_b$, also the switched modes must be specified where each subsystem can independently reach two modes, the zero flow (ZF) and the zero effort (ZE) mode. There are also five different operation or switching modes, which calls for three boolean-like control variables (where only 5 combinations out of the 2^3=8 will be employed).

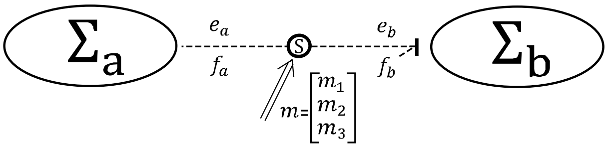


Figure 8. S-bond interconnecting two subsystems.

When Σ_b is in the ZE mode the S-bond imposes zero effort to Σ_b. On the contrary when Σ_b is in the ZF mode ($f_b = 0$) the S-bond forces the value of e_b necessary to satisfy the algebraic restriction $f_b = 0$.

When Σ_a is in the ZF mode the S-bond imposes zero flow to Σ_a. On the contrary when Σ_a is in the ZE mode ($e_a = 0$) the S-bond forces the value of f_a necessary to satisfy the algebraic restriction $e_a = 0$.

$$e_b = \begin{cases} e_a \ when\ \Sigma_a \ and\ \Sigma_b \ conected \\ e_r \ when\ \Sigma_b \ is\ in\ ZF\ mode \\ 0\ when\ \Sigma_b \ is\ in\ ZE\ mode \end{cases} \tag{4}$$

$$f_a = \begin{cases} f_b when\ \Sigma_a \ and\ \Sigma_b \ conected \\ f_r \ when\ \Sigma_a \ is\ in\ ZE\ mode \\ 0\ when\ \Sigma_a \ is\ in\ ZF\ mode \end{cases} \tag{5}$$

In (4) and (5) e_r and f_r are calculated through the corresponding algebraic constraints $f_b = 0$ and $e_a = 0$.

Figure 9 shows the SwBG model of the switched electric circuit of Figure 2 modeled with S-bonds. Here, each switch is represented by only one S-bond.

The S-bond commanded by m represents the electrical switch, while the S-bond commanded by $\overline{m}$ represents the switching behaviour of the diode.

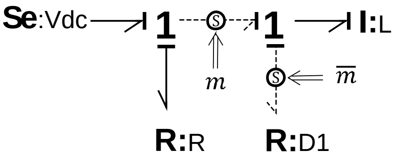


Figure 9. SwBG model of the switched electric circuit with S-bonds, $m = [m_1, m_2, m_3]$

There are no dangling junctions now: when the electric switch is OFF, the S-bond commanded by $m = [m_1, m_2, m_3]$ imposes zero flow to the 1-junction on the left and zero effort to the 1-junction on the right. At the same time, the $\mathbf{R}(D1)$ element calculates the effort imposed to the 1-junction on the right through the bond commanded by $\overline{m}$ (the diode ON). When the switch is ON (and the diode OFF) the S-bond commanded by $\overline{m}$ imposes zero flow to the $\mathbf{R}(D1)$ and zero effort to the 1-junction, while the other S-bond imposes the flow calculated by the I-element to the 1-junction on the left. The mathematical details of the control vector m are given in Table 1 in the next subsection.

3.1.1. Implementation of S-bonds with elementary BG components.

Figure 10 shows the internal representation of S-bonds using SPJs to model the mode switching and residual sinks to solve the algebraic constraints of each mode. The I/O relationships of this structure are given in (6) where e_r and f_r are imposed by the residual sinks. The behavior specified by Eqs. (4) and (5) is achieved with the combinations of the control variables $m_{1,2,3}$ given in Table 1.

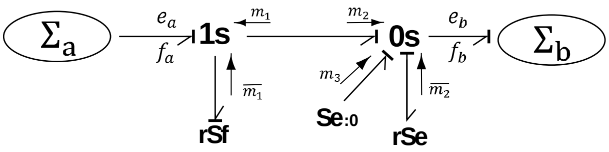


Figure 10. Internal S-bond representation

$$\begin{cases} e_b = (1 - m_3)(m_1 m_2 e_a + (1 - m_2) e_r) \\ f_a = (1 - m_3)(m_1 m_2 f_b + (1 - m_1) f_r) \end{cases} \tag{6}$$

For the sake of clarity, the model in Figure 10 uses the compact representation of SPJs and, thus, is not elementary. However, the version with BG components from the basic set can be achieved replacing the SPJs with their elementary realization as introduced in (Junco et al. 2007).

Table 1: S-bond modes and control variables.

m_3	m_2	m_1	mode
0	0	0	Σ_a in ZE and Σ_b in ZF
1	0	1	Σ_a in ZF and Σ_b in ZE
0	0	1	Σ_a and $\Sigma_b$in ZF
0	1	0	Σ_a and $\Sigma_b$in ZE
0	1	1	Σ_a and Σ_b connected

With the purpose of illustration consider the series RLC circuit of Figure 11a, where different kind of faults are expected to occur at the connection point of the resistor and the inductor, as depicted in Figs. 11b-11e. Each circuit configuration can be seen as a commutation mode between subsystems Σ_a and Σ_b. The transition among these modes and the behavior in each of them is modeled, employing S-bonds, by the SwBG of Fig. 12.

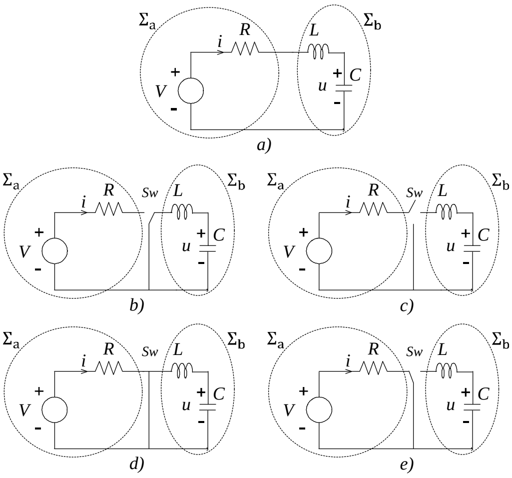


Figure 11. Series RLC circuit a) normal mode. b) Σ_a in ZF mode and Σ_b in ZE mode. c) Σ_a and Σ_b in ZF mode. d) Σ_a and Σ_b in ZE mode. e) Σ_a in ZE mode and Σ_b in ZF mode.

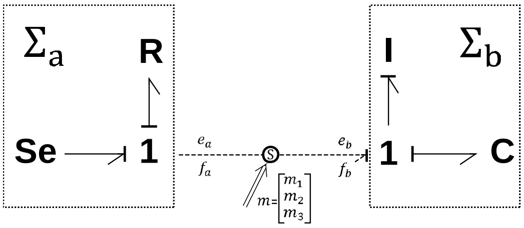


Figure 12. SwBG model of a faulty RLC circuit.

As already said, keeping fixed causality calls for the residual sinks to solve algebraic constraints in some modes. The following are the calculations of e_r and f_r for the different modes of this example. For the cases of Figs. 11c and 11e the effort on Σ_b is $e_r = e_c + \delta(t - t_C)$ where e_c is the effort of the capacitor and $\delta(t - t_C)$ is the necessary Dirac impulse of effort necessary to bring the inductance current to zero because of the switching (circuit opening) at time $t = t_C$. As in a numerical simulation the Dirac impulse cannot be implemented, to force to zero the flow in the 1-junction, the integrator of the I element must be reset to zero.

For the operation mode represented in Figure 11d, the calculus of f_r is trivial and is equal to $f_r = V/R$.

3.2. GSJ: Generalized Switched Junction Structures

The generalized switched junction structures 1-GSJ and 0-GSJ are introduced here as controlled junctions that can represent all the interconnections modes enforced by commutations involving BG-elements around the standard 0- and 1-junctions. They will be graphically represented as 0_G and 1_G.

To better understand their behavior consider that GSJ have a ground configuration where they behave like standard BG-junctions. This ground configuration is just one of their possible switching modes. In any of the other switching modes, the junction behaves as in the ground configuration but only for a subset of all the adjacent bonds, while the remaining bonds get disconnected from the junction. Thus, in a 1-GSJ (0-GSJ) these bonds do not contribute any effort (flow) to the junction, while their flows (efforts) are determined by the structural condition which their own efforts (flows) must satisfy. The configuration of a set of control variables decides which is the subset of bonds sticking to the ground junction configuration (selected bonds) and which is the subset disconnected (not selected bonds).

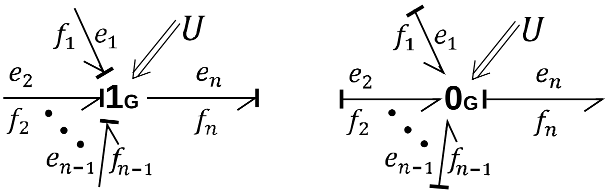


Figure 13. 1-GSJ and 0-GSJ representation

Figure 13 shows the BG iconic representation of the GSJ, where $U = [u_1, u_2, \ldots, u_n]$ is the vector of control signals. In the ground configuration the bonds numbered from 1 to $(n-1)$ impose the effort (flow) to the 1-GSJ (0-GSJ) while the n^{th} bond imposes the flow (effort) to it. Each control signal u_i $(i = 1, 2 \ldots n)$ can only take the value 1 or 0 and commands the i^{th} bond. In the case of the 1-GSJ, for $i = 1, 2 \ldots, (n-1)$, when u_i takes the value 0, the i^{th} bond does not contribute any effort to the junction (this does not necessarily means that its effort is zero!). When u_n takes the value 0, then the n^{th} bond imposes zero flow (which is transmitted by the junction to the selected bonds only) and its effort is obtained from the algebraic restriction $f_n = 0$. Equations (7) and (8) specify precisely the relationships among all the variables in the 1-GSJ and the 0-GSJ, respectively.

$$\begin{cases} e_n = u_n \sum_{i=1}^{n-1} u_i\, e_i + (1-u_n) e_r \\ f_i = u_i u_n f_n + (1-u_i) f_r \quad \forall\, i = 1 \ \ to\, n \end{cases} \tag{7}$$

$$\begin{cases} f_n = u_n \sum_{i=1}^{n-1} u_i\, f_i + (1-u_n) f_r \\ e_i = u_i u_n e_n + (1-u_i) e_r \quad \forall\, i = 1 \ \ to\, n \end{cases} \tag{8}$$

In (7) the value of e_r is calculated through the algebraic restriction $f_n = 0$ when the 1-GSJ is in the ZF mode and the value of f_r is calculated through the algebraic restriction $\sum_{j=1}^{j=m} \left(1 - u_{\mu j}\right) e_{\mu j} = 0$, where $m \le n - 1$ is the number of bonds in ZE mode and $\mu_j \in \{1, 2, \ldots, n-1\}$ (i.e., μ_j is the index of the not selected bonds). An analogue algebraic restriction is used to obtain f_r for the ZF mode of (8).

As an example of the GSJ behavior, the series circuit of Figure 11a is considered again, but in this case assuming the possible occurrence of the more ample spectrum of configurations depicted in Figures 15 and 16. All of them can be captured by the BG of Figure 14, with the control vector U defined in Table 2. There are 16 configurations in Figs. 15 and 16, the ground configuration of Fig. 16h and 15 faulty modes, so that a control vector with 4 variables is needed: $U = [u_1, u_2, u_3, u_4]$.

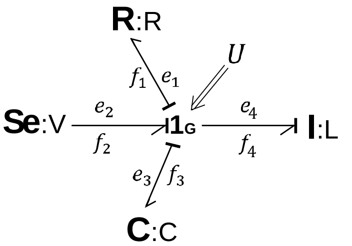


Figure 14. 1-GSJ model of switching series circuit.

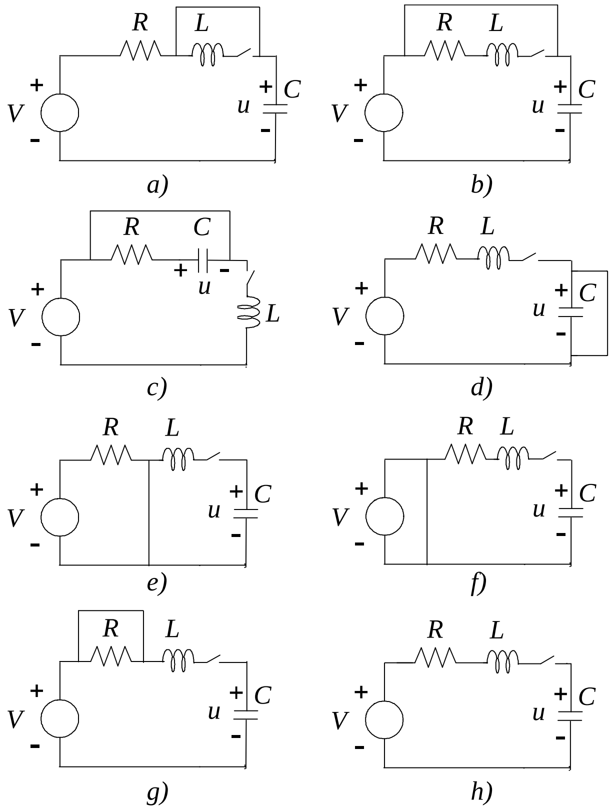


Figure 15. ZF modes of the series electrical circuit.

Table 2: Switching modes of a 1-GSJ

Fig	u_4	u_3	u_2	u_1	Modes
15a	0	0	0	0	S_eRC in ZE and I in ZF
15b	0	0	0	1	S_eC in ZE and RI in ZF
15c	0	0	1	0	RC in ZE and S_eI in ZF
15d	0	0	1	1	C in ZE and S_eRI in ZF
15e	0	1	0	0	S_eR in ZE and CI in ZF
15f	0	1	0	1	S_e in ZE and RIC in ZF
15g	0	1	1	0	R in ZE and S_eCI in ZF
15h	0	1	1	1	S_eRIC in ZF
16a	1	0	0	0	$S_eRIC\ in\ ZE$
16b	1	0	0	1	$S_eC\ in\ ZE\ and\ RI\ in\ ZE$
16c	1	0	1	0	$RC\ in\ ZE\ and\ S_eI\ in\ ZE$
16d	1	0	1	1	$C\ in\ ZE\ and\ CRI\ in\ ZE$
16e	1	1	0	0	$S_eR\ in\ ZE\ and\ CI\ in\ ZE$
16f	1	1	0	1	$S_e\ in\ ZE\ and\ RIC\ in\ ZE$
16g	1	1	1	0	$R\ in\ ZE\ and\ S_eCI\ in\ ZE$
16h	1	1	1	1	$Standard\ 1-junction$

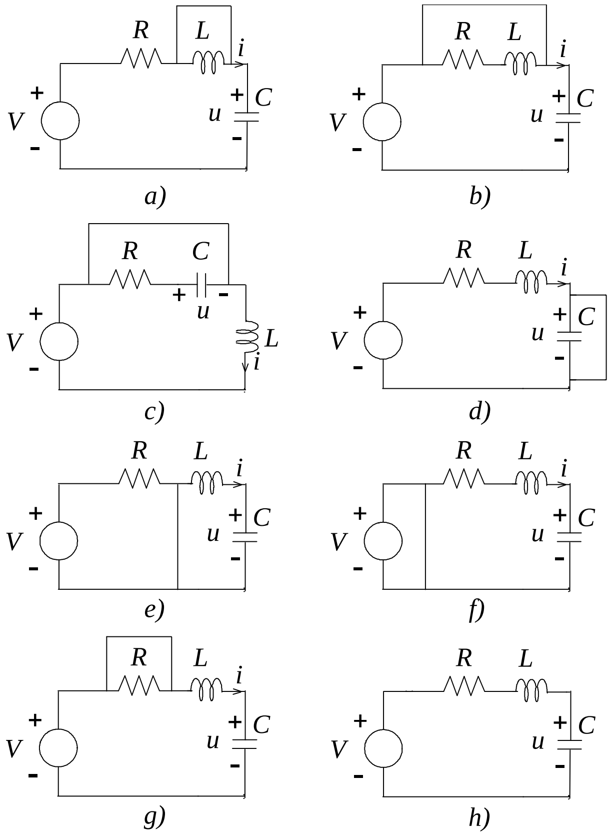


Figure 16. ZE modes of the series electrical circuit.

3.2.1. Representation of GSJ with atomic BG elements.

Following the reasoning proposed in (Junco et al. 2007) for the SPJ, also the GSJs can be represented by standards BG components as in Figures 17 and 18. The control signal enters in the BG multiplying the power variables through *MTFs*; the algebraic operations between power variables are carried out by the standard junctions of the BG formalism and the algebraic constraints are added using residual sinks.

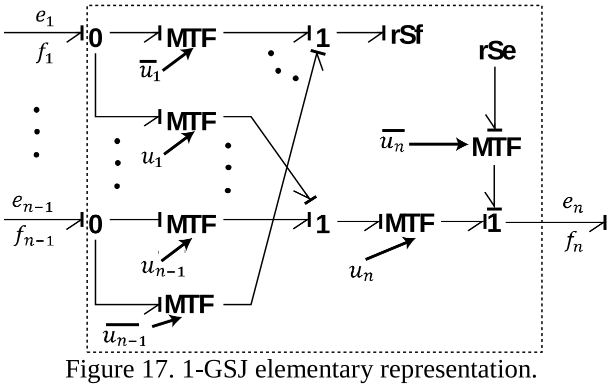


Figure 17. 1-GSJ elementary representation.

Remark: The elementary representations of Figs. 17 and 18 are also useful if the modeling approach with instantaneous commutations is resigned in favor of an approximation using parasitic components: it suffices to replace the residual sinks with the parasitic components, or with MTFs plus resistors, as done in (Borutzky 2010 and Dauphin-Tanguy and Rombaut 1997).

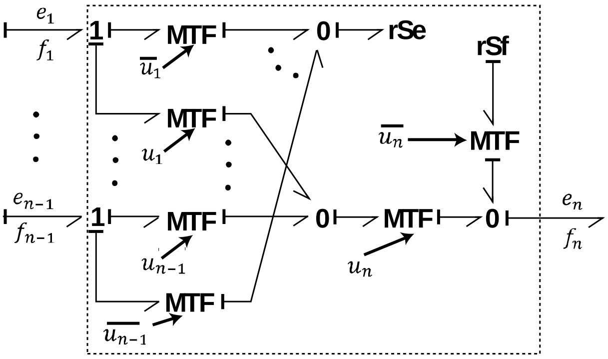


Figure 18. 0-GSJ atomic representation.

4. APPLICATION EXAMPLES: SWITCHED ELECTRONIC CIRCUITS

Modeling some switched circuits with GSJ and S-bonds, this section suggests a modeling technique.

4.1. Buck converter

The Buck converter of Figure 19 (a reducing DC-DC voltage converter: the output voltage *u* is less or equal than the input voltage *V*) contains an ideal switch (in practice, a switched transistor) and a free-wheel diode. In normal operation the diode (modeled as a resistor **R,** labelled *D1* in conduction state) and the switch have complementary logic states; in some cases, a third operation mode called discontinuous mode can take place, when the current through the diode becomes zero and both, switch and diode, are in the off state.

The basic modeling idea is to use a 0-GSJ (1-GSJ) when/where the switch commutates the application of an effort (flow) variable. In this example, the first case applies when the calculation of the potential $\mathscr{P}$ changes according to the switch state, so that a 0-GSJ must be used to represent it. The system is modeled considering the switch closed (corresponds to the 0-GSJ in its ground state), which yields the SwBG of Figure 20, endowed with an appropriate causality assignment and the control vector $U = [u_1, u_2, u_3]$.

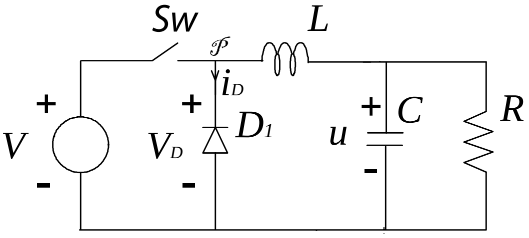


Figure 19. Schematic circuit of a Buck converter.

When the ideal switch is ON the diode is OFF $e_2 = e_3$ and e_1 is calculated through the algebraic restriction $f_1 = 0$. When the ideal switch is OFF *and* the diode is ON *and* its current i_D is less than zero (cf. the positive sense of the current i_D in Figure 19: $i_D = -f_1$), *then* e_2 is calculated through the algebraic restriction $f_1 - f_2 = 0$ and $f_3 = 0$. While when the diode is OFF (discontinuous operation mode of the circuit), e_2 is calculated through the algebraic restriction $f_2 = 0$. All the Buck converter operation modes are presented in Table 3.

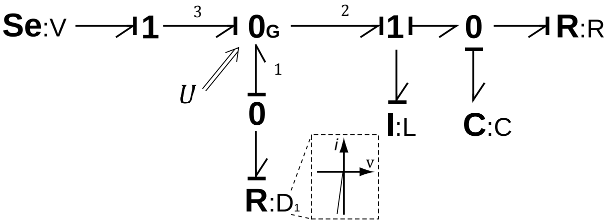


Figure 20. Buck converter SwBG model using GSJ.

Table 3: Buck converter modes

u_3	u_2	u_1	Modes
1	1	0	Switch ON, Diode OFF
0	0	0	Diode ON Switch OFF
0	1	0	Diode OFF Switch OFF

4.2. Boost converter

The Boost converter, depicted in Figure 21, is an amplifying DC-DC voltage converter, where the output voltage *u* is greater or equal than the input voltage *V*. This circuit has two operation modes, switch ON and diode OFF (mode M_1), and the opposite mode (M_2) switch OFF and diode ON. As in the previous example the diode is modeled, in conduction state, as a resistor **R** labelled *D1*.

Figure 22 shows the SwBG obtained for the Boost converter following the modeling technique suggested at the beginning of this section. It uses 1-GSJ considering that the current path is swtiched at node $\mathscr{N}$. As the diode switches its current between zero and a positive value, while the current through the inductance is always positive, the flow is imposed to the 1-GSJ by the resistor **R(***D1***)**. This causality assignment forces derivative causality in the inductance which is not desirable. The different operation modes of the Boost converter according to Figure 22 are reached with the control signals presented in Table 4.

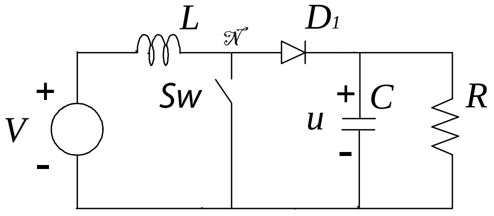


Figure 21. Schematic circuit of a Boost converter.

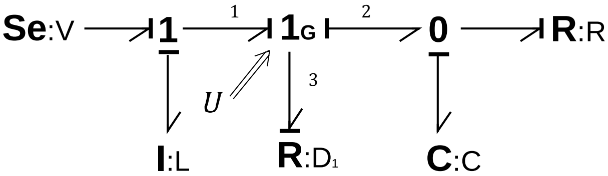


Figure 22. SwBG of the Boost converter with 1-GSJ.

Table 4: Boost converter modes of Figure 22.

u_3	u_2	u_1	modes
1	1	1	Diode ON, Switch OFF
1	0	0	Diode OFF, Switch ON

To enforce integral causality in the inductance, the 1-GSJ can be replaced by a 0-GSJ and an effort source can be placed to break the causality conflict. The resulting SwBG model is depicted in Figure 23 and Table 5 shows the different combination of the control signals to reach the operation modes.

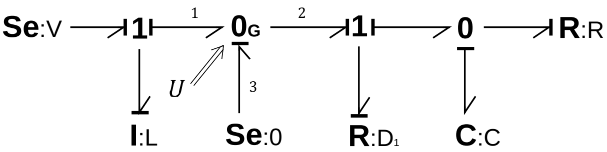


Figure 23. SwBG of the Boost converter with 0-GSJ.

Table 5: Boost converter modes of Figure 23.

u_3	u_2	u_1	modes
1	0	1	Switch ON, Diode OFF
1	0	0	Switch OFF, Diode ON

Instead of GSJs, Figure 24 uses a S-bond to model the switching in the Boost converter; its interpretation is straightforward: when the switch is ON (diode OFF) the S-bond imposes ZE to the series $S_e - I$ and ZF to the rest of the circuit. Whereas when the switch is OFF (diode ON) the S-bond works as a standard bond connecting both sub-circuits. Table 6 shows the combination of the control signals for the different configurations (cf. Eqs. 6).

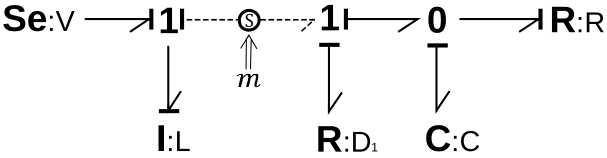


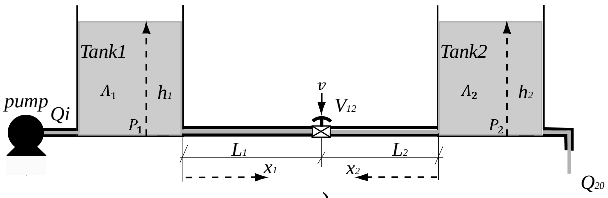
Figure 24. SwBG of the Boost converter with S-Bonds.

Table 6: Boost converter modes with S-bonds

m_3	m_2	m_1	modes
0	0	0	Switch ON, Diode OFF
0	1	1	Switch OFF, Diode ON

5. APPLICATION TO FAULT MODELING: FAULTY TWO TANK SYSTEM.

The application example consists in two tanks separated by a distance $L = L_1 + L_2$ and connected by two pipes and a valve (V_{12}) which controls the flow passage as shown in Figure 25. The pipe 1 connects the Tank1 with the valve V_{12} and has a length L_1, while pipe 2 connects the valve V_{12} with the Tank2 and has a length L_2. Figure 26 shows the associated BG model.



a)

Figure 25. Two tanks physical system.

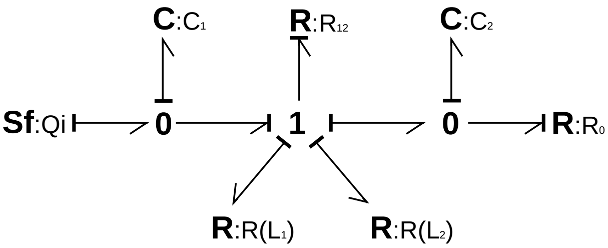


Figure 26. BG model of healthy two tank system.

The following constitutive relationships of the BG elements of Figure 26 are assumed: $R(x_i)\colon Q = \frac{a_i}{x_i + D} \sqrt[2]{\Delta P}\ sign(\Delta P)$ where a_i and x_i (with $i = 1{,}2$) represents the cross section and length of the pipes while D match the value of the restriction when $x_i = 0$; $R_{12}\colon Q = a_{12} v \sqrt[2]{\Delta P}\, sign(\Delta P)$ where a_{12} is the discharge coefficient of the valve and v is the opening control of the valve; $R_0\colon Q = a_0 \sqrt[2]{\Delta P}\, sign(\Delta P)$ where a_0 represents the cross section of outlet hole from Tank2; $C_i = \frac{A_i}{\rho g}$ (with *i=1,2*) are the tanks hydraulic capacities where A_1 and A_2 are the cross section areas of the tanks ρ is the constant density of the liquid, g is the gravitational acceleration.

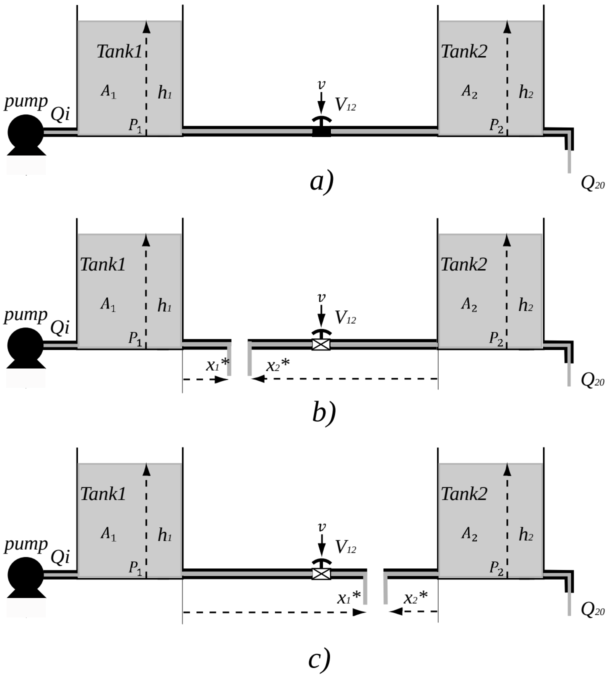


Figure 27. Fault modes of the two tank system. a) Valve V_{12} blocked, b) broken pipe at x_1^*, c) broken pipe at x_2^*.

In this example three different abrupt faults are considered for modeling purposes, as shown in Figures 27a,b,c. The first one is a blockage in the valve V_{12}, the second and the third one corresponds to the pipe broken at position x_1^* and x_2^* respectively. All these faults break the shared flow constraint of the pipes and the valve. So, to represent the structural changes produced by the faults a 1-GSJ can be placed instead of the standard 1-junction, which yields the SwBG model of Figure 28.

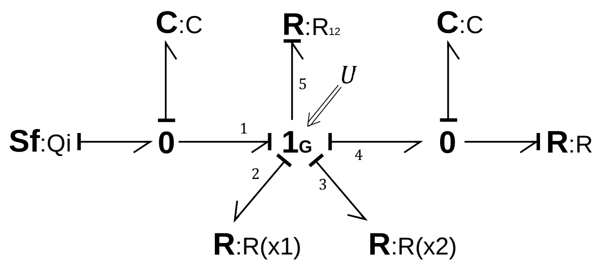


Figure 28. SwBG model of the faulty two tank system.

Table 7 shows the combinations of the control signals that generate the healthy and the faulty modes. Figures 29, 30, 31 explicitly show, in the form of BG models, the different calculations implemented by the 1-GSJ in the faulty modes as determined by the signals of the control vector $U = [u_1, u_2, u_3, u_4, u_5]$.

Table 7: two tank modes

u_5	u_4	u_3	u_2	u_1	Two tank process modes
1	1	1	1	1	Healthy
1	0	0	1	1	Pipe 2 broken at x_2^*.
1	1	1	0	0	Pipe 1 broken at x_1^*.
0	1	1	1	1	Valve V_{12} blocked (ZF mode)

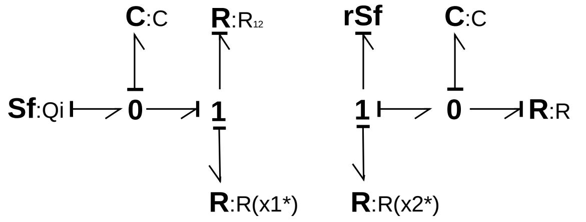


Figure 29. Pipe 2 broken at x_2^*. , $U = [1,1,0,0,1]$

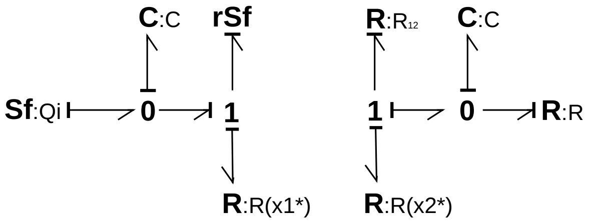


Figure 30. Pipe 1 broken at x_1^* , $U = [0,0,1,1,1]$.

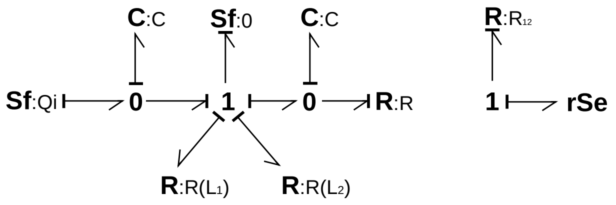


Figure 31. Valve V_{12} blocked, $U = [1,1,1,1,0]$

5.1. Simulation results.

In this subsection some simulation results are presented to show the correct behavior of the GSJs.

The following parameters are used in the simulations (Samantaray and Ould Bouamama 2008): $A_i = 1.45\ 10^{-2}\ m^2$ $a_{12} = 1.593\ 10^{-2} kg^{1/2} m^{1/2}$, $Q_i = 1\ m^3/s$, $L_1 = 1m$, $a_0 = 1.596\ 10^{-2}\ kg^{1/2} m^{1/2}$, $L_2 = 1\ m$ and $a_i = 0.03\ kg^{1/2} m^{1/2}$, $D = 0m$, $v = 1$.

In all simulation responses from top to bottom, P_1 is the pressure at the bottom of Tank1 in N/m^2, P_2 is the pressure at the bottom of Tank2 in N/m^2, Q_{1out} is the output mass flow of Tank1 in m^3, Q_{2in} is the input flow mass of Tank2 in m^3 and Q_{2out} is the output flow mass of Tank2 in m^3.

Figure 32 shows the simulation response of a fault in the pipe that connects Tank1 with the valve V_{12}. The fault occurs at time $T = 310\ s$ and at a distance $x_1^* = 0.5\ m$, which implies that $x_2^* = 1.5\ m$.

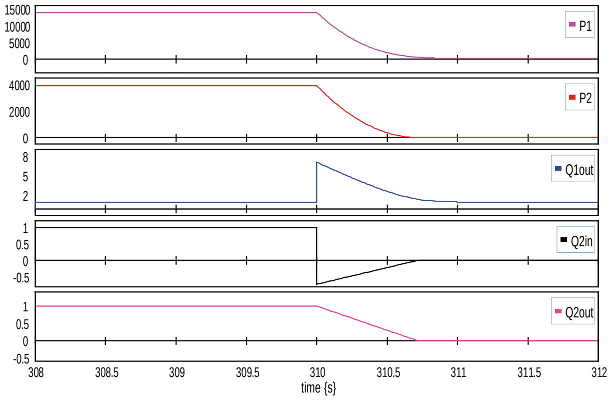


Figure 32.Simulation response with pipe 1 broken at x_1^*.

Figure 33 shows the simulation response of a fault in the pipe that connects Tank2 with the valve V_{12}. The fault occurs at time $T = 310\ s$ and at a distance $x_2^* = 0.5\ m$ which implies that $x_1^* = 1.5\ m$.

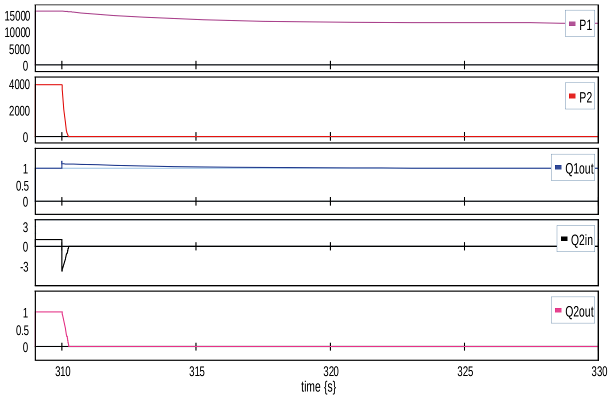


Figure 33.Simulation response with pipe 2 broken at x_2^*.

To perform a simulation of sequential structural faults, the control signal U starts with $U = [1,1,1,1,1]$ (system in healthy mode), then changes to $U = [0,1,1,1,1]$ (commutation to "valve blocked") and, finally, switches to $U = [0,1,1,0,0]$ ("valve blocked and pipe 1 broken at x_1^*"). Notice that the latter faulty mode is not in Table 7. Figure 34 shows the simulation response of this sequence of structural faults. At $T = 300\ s$ the valve V_{12} gets blocked; then, pipe 1 breaks at $T = 302\ s$ at $x_1^* = 0.5\ m$.

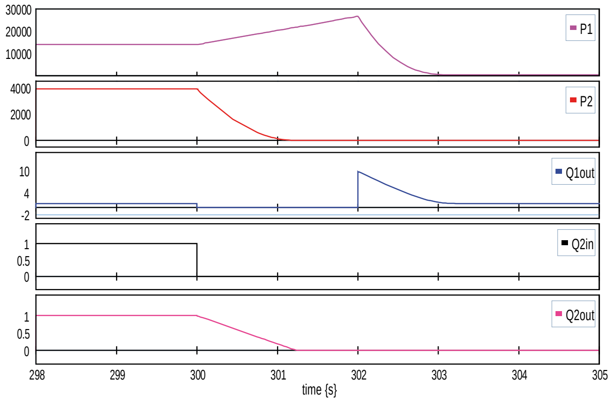


Figure 34. Simulation response with sequential faults.

6. CONCLUSIONS

This paper introduced two new fixed-causality formalisms to handle ideal switching processes -i.e., commutations happening within a null time span- in the Bond Graph domain. The first one, called Switchable Structured Bond, S-Bond for short, allows to model the power connection/ disconnection (presence/absence of a bond) between two subsystems and, at the same time, solves the “dangling junction” problem known to happen in the classical switchable bonds. The second one, called GSJ for Generalized Switched Junction Structure, allows to represent the classical structure of a standard BG-junction (called the ground configuration of the GSJ) *plus* all possible commutations involving the elements joined by the structure in its ground configuration. Both, a macro definition or representation and an internal implementation with elementary BG-components are provided for each of the new structures. A minor modification of the GSJ internal representation allows to alternatively adopt an approximate approach to switching modeling with the use of parasitic components. A procedure to construct the switched bond graphs models using these new techniques has been suggested. Also, application examples of controlled and fault-induced switching have been provided, together with some simulation results in the latter case.

ACKNOWLEDGMENTS

The authors wish to thank CONICET (the Argentine National Council for Scientific and Technological Research), and SeCyT-UNR (the Secretary for Science and Technology of the National University of Rosario) and PICT 0650-2008 for their financial support.

REFERENCES

Allard, B., Helali, H., Lin, Ch., Morel, H., 1995. Power converter average model computation using the bond graph and Petri Net Techniques. *IEEE Power Electronic Specialist Conference*, pp. 830–836.

Asher, G., 1993. The robust modeling of variable topology circuits using bond graphs. *Proc. ICBGM’93, the Int. Conf. on Bond Graph Modeling and Simulation*. pp. 126–131. San Diego, CA.

Borutzky, W., 2010. *Bond Graphs. A Methodology for Modelling Multidisciplinary Dynamic Systems*. Erlangen, San Diego: SCS Publishing House.

Borutzky, W., 2009. Bond graph model-based fault detection using residual sinks. *Proc. IMechE, Part I: Journal of Systems and Control Engineering*, Vol. 223, London: PEP. pp. 337-352.

Broenink, J. and Wijbrans, K., 1993. Describing discontinuities in bond graphs. *Proceedings ICBGM’93*. pp. 120–125 San Diego, CA.

Dauphin-Tanguy, G. and Rombaut, C., 1997. A Bond Graph Approach For Modeling Switching Losses of Power Semiconductor Devices. *Proc. ICBGM’97*. pp. 207-212. Phoenix, Arizona, USA.

Demir, Y. and Poyraz, M., 1997. Derivation of state and output equations for systems containing switches and a novel definition of switch using the bond graph model. *J. Franklin Institute*. 334 (2) 191–197.

Lin, H. and Antsaklis, P., 2007. Switching Stabilizability for Continuous-Time Uncertain Switched Linear Systems. *IEEE TAC*. Vol. 52, N°4. pp 633-646.

Junco, S., Diéguez, G., Ramírez, F., 2007. On commutation modeling in Bond Graphs, *Proc. Int. Conf. on Bond Graph Modeling and Simulation*, pp. 12-19. San Diego, CA.

Karnopp, D., Margolis, D., Rosenberg, R., 2000. *System Dynamics. Modeling and simulation of mechatronic systems*, 3rd edition, Wiley Inter-science, New York.

Mosterman, P. and Biswas, G., 1995. Behaviour generation using model switching. A hybrid bond graph modeling technique. *Proc. ICBGM’95*. pp. 177–182. Las Vegas, USA.

Mosterman, P. and Biswas, G., 1998. A theory of discontinuities in physical systems models. *Journal of the Franklin Institute*, 335B (3), 401–439.

Nacusse, M., Junco, S., Donaire, A., 2008. Automatizando el Modelado y Simulación con Bond Graph de sistemas Físicos Conmutados. *Mecánica Computacional*. XXVII, pp. 3479-3494.

Nacusse, M., Junco, S. 2010. Automatic handling of commutations in Bond Graph based Simulation of Switched Systems. *Proc. 4th Int. Conf. on Integrated Modeling and Analysis in Applied Control and Automation (IMAACA2010)*. pp. 27-35. Fes, Morocco.

Strömberg, J., Top, J., Söderman, U., 1993. Variable causality in bond graph caused by discrete effects. *Proc. Int. Conf. on Bond Graph Modeling and Simulation*, pp. 115–119. San Diego, CA,.

Strömberg, J. E. 1994. *A mode switching modelling philosophy*. Phd Thesis, Department of Electrical Engineering. Linköping University, Sweden.

Samantaray, A.K., Ould Bouamama ,B. 2008. *Model-based Process Supervision*. Springer Verlag London Ltd.

Umarikar, A. and Umanand, L., 2005. Modelling of switching systems in bond graphs using the concept of switched power junctions. *Journal of the Franklin Institute*, Vol. 342, pp. 131-147.

UNKNOWN INPUT OBSERVER FOR DETECTING CURRENT SENSOR FAULTS FOR A DFIG IN WIND TURBINE

H. OUYESSAAD [(a)], **H. CHAFOUK** [(a)]

[(a)] IRSEEM : Technopole du Madrillet, Avenue Galilée, 76801 St Etienne du Rouvray France
[(a)] hakim.ouyessaad@esigelec.fr, [(a)] houcine.chafouk@esigelec.fr.

ABSTRACT

The objective of this paper is to detect and isolate the presence of sensor faults in dynamical systems. Unknown input observers are used which is then used to generate residuals based on the DOS observer architecture (Dedicated Observer Scheme). This diagnosis strategy is applied on the double-feed induction generator (DFIG) in wind turbines. The structure of a DOS is used for detection and isolation of multiple sensor faults. The approach is validated using signals obtained from a simulated DFIG system. The main contribution of this paper is the modelling of induction generator for wind turbines and the use unknown input observers to detect multiples and simultaneous faults in current sensors. The simulation model of DFIG is developed using MATLAB.

Keywords: wind turbine, DFIG, UIO, observer DOS, current sensor

1. INTRODUCTION

In recent years, environmental issues play an increasingly important role in our daily lives. This is particularly due to an awareness of people about the consequences of some pollution on the environment and climate conditions. This work aims to improve safety, reliability and performance of wind turbine and to predict the evolution of degraded mode, in order to improve the availability of the system (Odgaard, Thogersen and Stroustrup 2009; Blanke, Kinnaert, Lunze, Staroswiecki and Schroder 2010), The main goal of this studies-to detect, isolate faults and to determining the origin of abnormalities (i.e. failure of sensors or actuators, system malfunctions).

With the increasing size of wind turbine (Rothenhagen, Thomsen and Fuchs 2007), there is a need for improving diagnostic techniques in order to detect the favourable arguments for the occurrence of fault. Several works concern the detection of fault and their isolation in the wind system. In (Rothenhagen, Thomsen and Fuchs 2007) an observer was setup to detect sensor fault in the turbine hub. Reconfiguration has been proposed by (Rothenhagen and Fuchs 2009). An observer with unknown inputs has been proposed by (Nielsen 2009; Gálvez-Carrillo and Kinnaert 2011). In this paper, we have shown interest in the detection and isolation of currents sensors faults, for example internal faults in the DFIG are caused by the component of the generator (rotor and stator magnetic circuits, stator windings, mechanical air gap). The generator is often exposed to perturbation as the origin of the noise from the environment of the generator, the uncertainties of measurements, sensors faults or actuators. The construction of the observers is knowledge-based for the model of the generator to observe.

In this paper an unknown input observers are used as well (Chen and Patton 1999). The faults are considered as unknown inputs which should be estimated this can be done by introducing internal fault models. Other examples of this usage of the unknown input observers can be seen in (Odgaard and Mataji 2008) where the former reports similar scheme applied on fault detection of power plant coal mills and the latter estimate power coefficients for wind turbine, some examples can be found of fault detection and accommodation of wind turbine. An observer based scheme for detection of sensor faults for blade root torque sensor is presented in (Wei, Verhaegen and Van den engelen 2008). An unknown input observer based scheme was in (Rothenhagen, Thomsen and Fuchs 2007) proposed to detect such faults in a wind turbine.

In present paper, the problem of sensors fault detection and isolation the multiple and simultaneous currents sensors, in DFIG driven by a wind turbine application has been addressed. The application of DOS is possible if the system is observable.

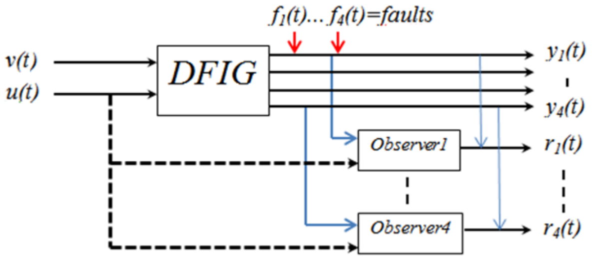


Figure 1: Structure of Observer DOS

2. SYSTEM DESCRIPTION

The DFIG is one of the most used wind generator (Khojet EL Khail 2006.), “Figure 2”. Several recent papers, confirmed by industrial realizations, demonstrate the viability of this device in a wind system.

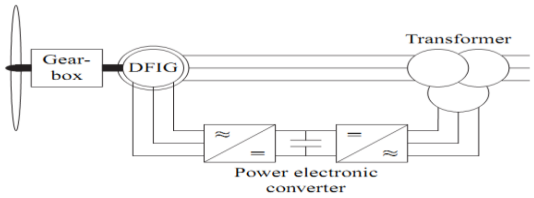


Figure 2: Wind turbine with a DFIG

The presence of a converter may result in large variations of the rotor voltages with high frequencies, rectifiers and inverters are used to solve this problem. Wind energy captured by the turbine is converted by the DFIG and is transmitted to the network by the rotor and the stator windings. The guidance system of the blades was used in order to adjust lift of the blade to the wind speed. So, it limits the power generated by the generator. With such a system, the blade is controlled by a system called "pitch control", this guidance system is not study in this paper.

3. MODEL OF THE DFIG

3.1. PARK TRANSFORMATION

The Park transformation is defined by the rotation matrix of the rotating field. It consists in the projection of the three phase coordinates (*a, b, c*), in frame (*d, q*) (El Aimani 2004). In this frame, the *d*-axis is chosen to coincide with stator axis at *t = 0* and *q* axis is lag by 90 degree with the *d* axis of the direction of rotation. To investigate the DFIG the d-q model is required. The used Park transformation is given by equation (1).

$$[P(\theta)] = \sqrt{\frac{2}{3}}\begin{bmatrix} \cos(\theta) & \cos(\theta-\frac{2\pi}{3}) & \cos(\theta-\frac{4\pi}{3}) \\ -\sin(\theta) & -sim(\theta-\frac{2\pi}{3}) & -\sin(\theta-\frac{4\pi}{3}) \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (1)$$

The transformation is valid for both stator and rotor, equation (2) describes the global modeling of the generator, use the voltage and fluxes become respectively.

$$\left.\begin{aligned} V_{ds} &= R_s i_{ds} + \frac{d\emptyset_{ds}}{dt} - w_s \emptyset_{qs} \\ V_{qs} &= R_s i_{qs} + \frac{d\emptyset_{qs}}{dt} + w_s \emptyset_{ds} \\ V_{dr} &= R_r i_{dr} + \frac{d\emptyset_{dr}}{dt} - w_r \emptyset_{qr} \\ V_{qr} &= R_r i_{qr} + \frac{d\emptyset_{ds}}{dt} + w_r \emptyset_{dr} \end{aligned}\right\} \quad (2)$$

where *V* stands for voltages (V), *I* stands for currents (A), *R* stands for resistors *(Ω)*, ∅ stands for flux linkages (V·s). Indices *d* and *q* indicate direct and quadrature axis components respectively while *s* and *r* indicates stator and rotor quantities respectively. w_s and w_r is the stator and the (mechanical) rotor speed of generator.

3.2. Observer Design

It is common when modeling a system, to involve inputs that are not measurable but which nevertheless affect the state. Use the term to refer unknown inputs, and reconstruction of the state of these systems can only be done under certain condition, the observer used are called unknown input observers (Odgaard and Stoustrup 2012 a, 2009b).

Doubly fed induction machines comprise of a wound stator and a wound three phase rotor, where the rotor windings can be accessed by brushes. Usually, the stator is connected to the grid, and the rotor is fed by an inverter. In this work, an Input Observer based on an unknown input observer (UIO) with input reconstruction is used to observe the stator voltages of the DFIG. UIO observe the states of this system, where an input, the unknown input, is not used. For the reconstruction of the unknown inputs the UIO needs the measured outputs of the system and the known inputs. Using the estimated states and the known inputs, the unknown input may be reconstructed. This is thoroughly described in (Rothenhagen and Fuchs 2009). The state space model used for the DFIG is given in equations (3) to (8). The bilinear character of the DFIG is represented by the matrix A_0. The reference frame is described by matrix A_1. The input is split up into known and unknown inputs (4), with their respective input matrices B and F. It is assumed that all currents are measurable, therefore C is unity matrix.The system matrices are explicitly given in (6), and (7), where ω_r is the mechanical rotor frequency, ω_a is the rotational frequency of the reference frame.

Using this system description, it is possible to easily convert the system from stator fixed into a synchronous reference frame or any other, since the influence of the rotation is described by ω_a. Explicitly, a stator fixed system is using ω_a=0, while a system oriented uses the stator angular frequency ω_a= ω_s=2π50 s $^{-1}$. A mor detailed desctiption may be found in (Rothenhagen and Fuchs 2009).

$$\dot{x}(t) = Ax(t) + Bu(t) + Fv(t) \quad (3)$$
$$y(t) = Cx(t)$$
$$u(t) = [Vdr\ Vqr]^T,\ v(t) = [Vds\ Vqs]^T \quad (4)$$
$$A = A_0 + A_1\omega_a \quad (5)$$

$$A_0 = \begin{bmatrix} -\frac{Rs}{\sigma Ls} & \frac{p\omega_r L_h^2}{\sigma LsLr} & \frac{L_h Rr}{\sigma LsLr} & \frac{p\omega_r L_h}{\sigma Ls} \\ -\frac{p\omega_r L_h^2}{\sigma LsLr} & \frac{-Rs}{\sigma Ls} & -\frac{p\omega_r L_h}{\sigma Ls} & \frac{L_h Rr}{\sigma LsLr} \\ \frac{L_h Rs}{\sigma LrLs} & \frac{-p\omega_r L_h}{\sigma Ls} & -\frac{Rr}{\sigma Lr} & -\frac{p\omega_r}{\sigma} \\ \frac{p\omega_r L_h}{\sigma Lr} & \frac{RsL_h}{\sigma LsLr} & \frac{p\omega_r}{\sigma} & \frac{-Rr}{\sigma Lr} \end{bmatrix} \quad (6)$$

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad (7)$$

$$B=\begin{bmatrix} \frac{1}{\sigma.Ls} & 0 \\ 0 & \frac{1}{\sigma.Ls} \\ -\frac{L_h}{\sigma.LsLr} & 0 \\ 0 & -\frac{L_h}{\sigma.LsLr} \end{bmatrix} F=\begin{bmatrix} -\frac{L_h}{\sigma.LsLr} & 0 \\ 0 & -\frac{L_h}{\sigma.LsLr} \\ \frac{1}{\sigma.Lr} & 0 \\ 0 & \frac{1}{\sigma.Lr} \end{bmatrix} \quad (8)$$

$(\sigma) = 1 - \frac{L_h^2}{L_r\, L_s}$: is the coefficient of Blondel

where $x(t) = [i_{ds}\ i_{qs}\ i_{dr}\ i_{qr}]^T$ is the state system, $u(t)$ is the control vector input, $y(t)$ are measured and the output, $v(t)$ is the vector of unknown input. The matrices A, B, F and C are matrices known the parameters of matrix are defined in “Table 6”, constant and consistent with the dimension signals. It is possible to reconstruct the system state observed despite the presence of unknown inputs.

The input of the system is divided in to input known and unknown (4), with their respective input matrices B and F. It is assumed that all currents are measurable output. The structure of the UIO can be defined as:

$$\dot{z}(t) = Nz(t) + Mu(t) + Ly(t) \quad (9)$$
$$\hat{x}\,(t) = z(t) + Ey(t) \quad (10)$$

wher N, M, L and E are matrices designed to achieve decoupling from the unknown input ans as well obtain an optimal observer and have to be designed in such a way $\hat{x}$ converge asymptotically to x. The matrices in the unknown input observer are found using the following equation (11)-(21), since system matrices are assumed constant.

$$P = I_n - EC \quad (11)$$

The estimation error of the stat $e(t)$, is given by:

$$e(t) = Px(t) - z(t) \quad (12)$$

It is necessary to satisfy the following conditions, if the matrices FC has full rank line, the equation (15) determine completely the matrix E of the observer. As a consequence, the observer error has to converge to zero.

$$0 = (PA - LC - NP) \quad (13)$$
$$0 = (PB - M) \quad (14)$$
$$0 = PF \quad (15)$$
$$E = F(CF)^+ + \Theta(I_m - CF(CF)^+) \quad (16)$$

The matrix N must be stable, Θ is matrix selected dy the placement of poles of N in order to reenter stability, and $(CF)^+$ is the generalized matrix of (CF):

$$(CF)^+ = ((CF)^T\, CF)^{-1}(CF)^T \quad (17)$$
$$P = I_n - F(CF)^+C + \Theta(I_m - CF(CF)^+)C \quad (18)$$
$$K = L - NE \quad (19)$$
$$N = PA - KC \quad (20)$$

At last the matrices N and K have to be designed. Therefore matrix M (14) is introduced for clarity. Solving (14) for N yields to (21). The eigenvalues of have to be in the left hand, the gain K guarantee the stability of the matrix N

$$N = (I_n - F(CF)^+C)A - [\ \Theta\ K] \begin{bmatrix} (I_m - CF(CF)^+)CA \\ C \end{bmatrix} \quad (21)$$

3.3. Estimation of unknown inputs

As seen earlier, the unknown input observer can estimate the state variables of the system independently unknown input $v(t)$. It is then possible to use the reconstructed state obtained also seek to estimate the unknown input. The choice of stator voltage as unknown inputs in that practice, the stator voltage are not measurable and they are not available, this is why we use the unknown input observer for the estimated

$$\dot{y}(t) = C\dot{x}(t) = CAx(t) + CBu(t) + CFv(t) \quad (22)$$
$$\hat{v}\,(t) = (CF)^+(\dot{y}(t) - CBu(t) - CA\hat{x}(t)) \quad (23)$$

The error estimation of the state tends asymptotically to zero, so $\hat{v}\,(t)$ asymptotically approaches $v(t)$. In other words, $\hat{v}\,(t)$ is the estimate of $v(t)$.

4. FAULTS DETECTION SCHEME DESIGN

The presence of faults sensors, denoted $f_c(t)=[\ f_1\ f_2\ f_3\ f_4]$ influence on the estimation error outputs. D_c is the matrix distribution of the additive faults.

$$\dot{x}(t) = Ax(t) + Bu(t) + Fv(t) \quad (24)$$
$$y(t) = CAx(t) + D_cf_c(t) \quad (25)$$

The estimation error output by unknown input observer can be seen as a dynamic system output follows:

$$\dot{\varepsilon}(t) = N\hat{x}(t) + (NE - L)D_cf_c(t) - ED_c\dot{f}c(t) \quad (26)$$
$$\hat{y}(t) = C\hat{x}(t) + D_cf_c(t) \quad (27)$$
$$\hat{x}(0) = Px_0 - z_0 \quad (28)$$

$$\dot{\varepsilon}(t) = (A - NP - LC - ECA)x(t) + (B - M - ECB)u(t) + (F - ECF)v(t) - D_cf_c(t) - ED_c\dot{f}c(t) - N\varepsilon(t) \quad (29)$$

It is therefore clear that, absent fault the signal $\hat{y}(t)$ converges asymptotically to zero, to the extent the matrix N is stable, $\varepsilon(t)$ the observer error.

4.1. Filter Bank for DFIG

The state observer for fault detection and isolation is a well-known problem. To overcome this problem, one can use a filter bank to estimate the dynamical behaviors of the system, in order to detect then and to isolate the fault. The first kind of filter bank is the Dedicated Observer Scheme (DOS) proposed by R. N. Clark in 1978 (Clark 1978; Trinh and Chafouk 2011; Ichalal, Marx, Ragot and Maquin 2006). And the second one, Generalized Observer Scheme (GOS) proposed by P. M. Frank in 1987 (Frank 1987) respectively. Each filter bank is composed by a number of observers, which are supplied with all inputs and different subsets of outputs. The observer which receives a faulty measurement provides a bad estimate of the variables, while the estimation of other observers converges to the measurements of corresponding outputs, except the output error.

5. SIMULATION RESULTS

The generation of residuals, is a fundamental step in designing a diagnosis based on models. In theory, a residual should be zero in the absence of fault and significantly different from zero in the contrary case. It is then necessary to introduce detection thresholds to avoid false alarms (Orjuela, Marx, Ragot and Maquin 2010). These thresholds can be set by the user from a statistical analysis of residues and taking. A structuring of residuals can accomplish the detecting and isolation of the faults (Gálvez-Carrillo and Kinnaert 2011; Ouyessaad, Chafouk and Lefebvre 2011). In this case, the i^{th} observer is controlled by the i^{th} output of the system and all inputs. The "Figure 1" shows the architecture of the bank of observers DOS. The faults of sensors affecting measurement $y(t)$ are denoted $f1(t)$, $f2(t)$, $f3(t)$ and $f4(t)$. Note later $r_{i,j}(t)$ the fault indicator signal (residual) calculated from the difference between the j^{th} system output and the i^{th} output estimated by the i^{th} observer DOS. If the output has a defect then there is a bad estimate of the state and residual $r_{1,i}$ may be affected.

In this section, the detection of current sensor faults by the observer Luenberger will be focused. Then, the isolation of the fault will be addressed. For this purpose, the following fault scenario will be used:

The first scenario is to introduce fault only the first step $y_1(t)$. The first fault was injected at $t = 0.3$ s and disappears $t = 0.4$ s, the fault consists of constant amplitude equal to 15% of the maximum amplitude of the output rotor current Idr, "Table 1".

Table 1: Faults Scenario I

Number	Fault number	Starting time (s)
1	$f_1(t)$	$0.3 \leq t \leq 0.4$

The observer 1 reconstructs the model output using only the output y_1 and all inputs of the system, in this step the measurement y_1 is affected by a fault f_1.

This output has a fault, then there is a poor estimate of the state and residuals $r_{1,1}$, $r_{1,2}$ $r_{1,3}$ and $r_{1,4}$ away from zero "Figure 4". To show the fault f_1 during its presence in y_1 (red ellipse) (Idr presence f_1) "Figure 3".

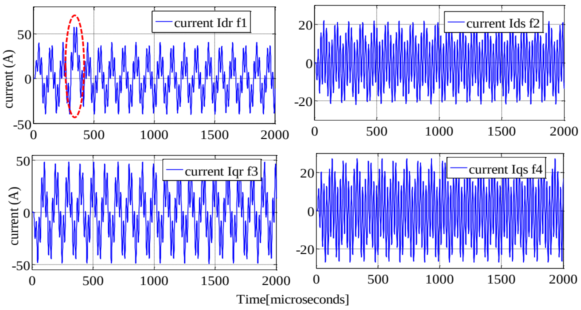


Figure 3 : Rotor and stator currents, can be seen the fault f1 in the rotor current idr

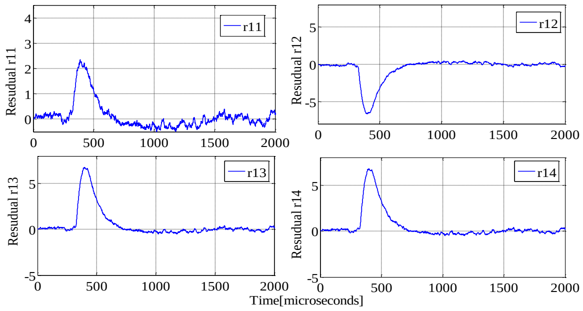


Figure 4: Fault residual for *idr*, the fault f_1 is detected during its presence in the observer 1

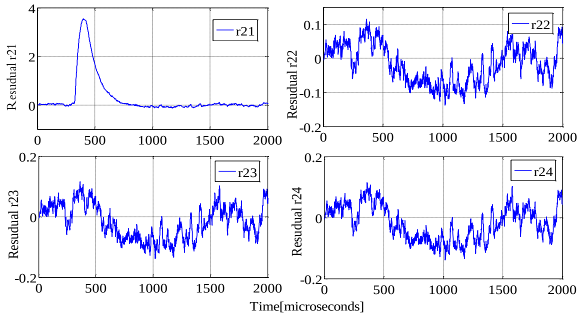


Figure 5 Fault residual for *idr*, the fault f_1 is detected during its presence in the observer 2

A sensor fault (f_1) is present in $r_{2,1}$ from time $0.3 \leq (t) \leq 0.4$. Using the observer 2 see "Figure 5", this sensor fault is detected without any false positive detection, other observers 3 and 4 confirmed this result, r_{31} and r_{41} with observers DOS. The signatures of the different faults are given in "Table 2".

Table 2: Signature of Faults

	Obs1				Obs2			
	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$	$r_{2,4}$
$f1$	1	1	1	1	1	0	0	0
	Obs3				Obs4			
	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$	$r_{3,4}$	$r_{4,1}$	$r_{4,2}$	$r_{4,3}$	$r_{4,4}$
$f1$	1	0	0	0	1	0	0	0

According to this table, the signatures $r_{1,j}= [1\ 1\ 1\ 1]$, $r_{2,j}= [1\ 0\ 0\ 0]$, $r_{3,j}= [1\ 0\ 0\ 0]$ and $r_{4,j}= [1\ 0\ 0\ 0]$ correspond the fault $f1$. Presented in "Figure 4", "Figure 6"and "Table 2". It is possible to conclude the fault f_1 appears in y_1 .

The second scenario consists to introduce multiple faults in the outputs y_1 and y_3, the following fault scenario will be used "Table.3".

Table 3: Faults Scenario II

Number	Fault number	Starting time (s)
1	$f_1(t)$	$0.9 \leq t \leq 1$
2	$f_3(t)$	$0.9 \leq t \leq 1$

It should be noted that in the simulation a measurement noise is added to the output of the DFIG (here, a random signal with zero mean and variance equal 1).

This information is confirmed by other observers. It is possible to conclude that the fault f_1 appear in the output $y1$ in interval 0.9(s) ≤ t ≤ 1(s), and that the fault f_3 appear in the output y_3 in interval 0.9(s) ≤ t ≤ 1(s). The fault consists of a window of constant amplitude equal to 15% of the maximum amplitude of the output “Figure 6”, present the fault detection the four residuals $r_{3,j}$, in observer 3. The observers 2 and 4 confirmed this results, ($r_{2,1}$, $r_{2,3}$) for observer 2, and ($r_{4,1}$, $r_{4,3}$) with the observers 4 show “Figure 7”.

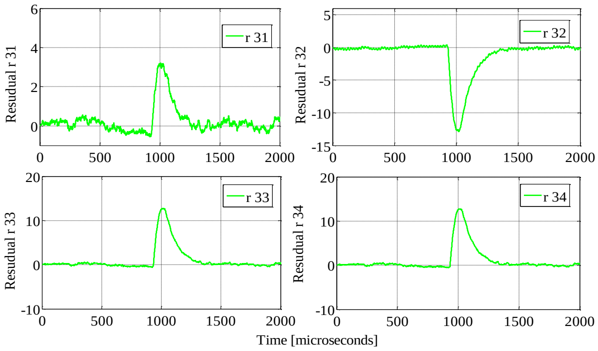


Figure 6: Fault residual in the observer 3

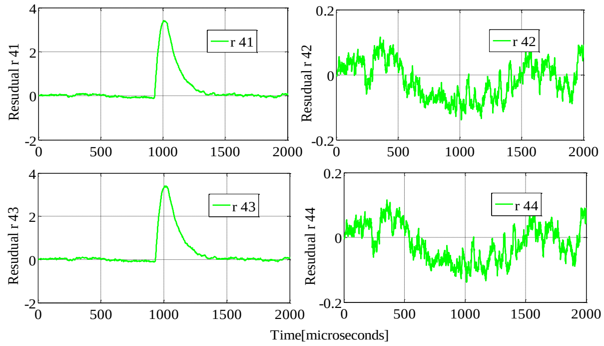


Figure 7: Fault residual in the observer 4

The third scenario consists to introduce the fault multiple and simultaneous in the outputs y_1 , y_2 and y_3, the following faults scenario will be used, see “Table.4”.

Table 4: Faults scenario VI

Number	Fault number	Starting time (s)
1	$f_1(t)$	0.9 ≤ t ≤1
2	$f_3(t)$	0.9 ≤ t ≤1
3	$f_2(t)$	1.3 ≤ t ≤1.4

The fault f_1 consists of constant amplitude equal to 20% of the maximum amplitude of the output y_1, f_3 = 5% of the output y_3 and f_2 = 10% of the output y_2, show “Figure 9”.

In this section introduction the multiples and simultaneous faults, with different amplitudes intended to show the sensitivity and robustness of the observers to detected the faults “Figure 9”.

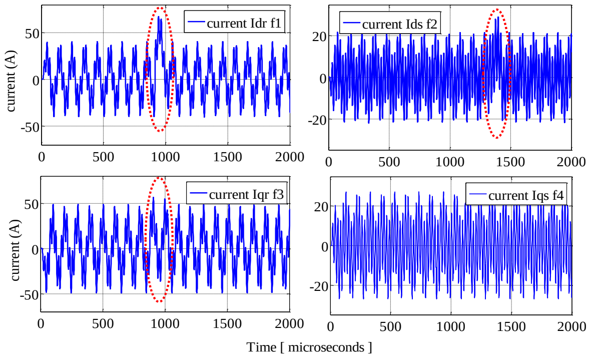


Figure 8: Rotor and stator currents, can be seen the fault f_1, f_2 and f_3 in the current sensor

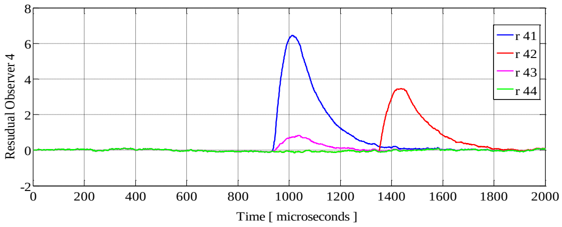


Figure 9: The four residuals signals $r_{4,j}$ in the observer four

The logical rules of the decision unit allow to detect and to locate the fault. The threshold is choose equal to 3σ (σ is the standard deviation of the signal). “Figure 10”, witch present the four faults detection residuals $r_{1,4}$, $r_{4,2}$, $r_{4,3}$ and r_{44}. “Table 5” see the different event of the occurrence of multiple and simultaneous fault, detected and located by the observer DOS.

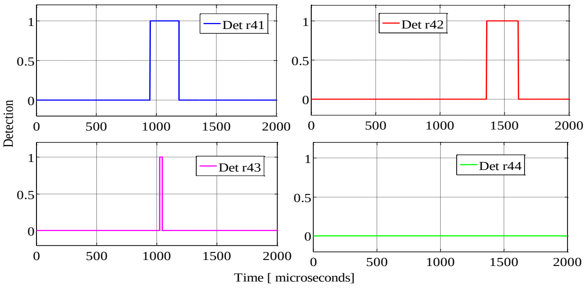


Figure 10: Single fault detection

Table 5: Generalized Table of Signature Faults

	Obs1				Obs2			
	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$	$r_{2,4}$
$f1$	1	1	1	1	1	0	0	0
$f2$	0	1	0	0	1	1	1	1
$f3$	0	0	1	0	0	0	1	0
$f4$	0	0	0	1	0	0	0	1
	Obs3				Obs4			
	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$	$r_{3,4}$	$r_{4,1}$	$r_{4,2}$	$r_{4,3}$	$r_{4,4}$
$f1$	1	0	0	0	1	0	0	0
$f2$	0	1	0	0	0	1	0	0
$f3$	1	1	1	1	0	0	1	0
$f4$	0	0	1	0	1	1	1	1

APPENDIX

V_{ds}; V_{qs} stator voltages in *d-q* reference frame
V_{dr}; V_{qr} rotor voltages in *d-q* reference frame
i_{ds}; i_{qs} stator currents in *d-q* reference frame
i_{dr}; i_{qr} rotor currents in *d-q* reference frame

Table 6: Parameters of the DFIG, 22kW (Rothenhagen K., Thomsen S., and Fuchs F. W., 2007)

Parameters	Values	Meaning
L_h	45.8 mH	Mutual inductance
L_s	46.8 mH	Stator inductance
L_r	46.8 mH	Rotor inductance
R_s	0.1315 Ω	Stator resistor
R_r	0.1070 Ω	Rotor resistor
P	2	Pairs of poles

6. CONCLUSION

In this paper, the problem of current sensor fault detection and isolation, for double-fed induction generator driven by a wind turbine application has been addressed. An unknown input observer scheme and a statistical detection algorithm have been used as residual generation and decision system, respectively. The approach has been validated using simulated signals of a double-fed induction generator for wind turbine. Through simulations, it has been demonstrated that multiple current sensor faults for rotor and stator have been correctly detected and isolated with a DOS observer scheme. The future extension of this work is to improve the performance of observers and insensitivity on measurement noise, in a test bank in real time. We are interesting in the FDI problem for other sensors of wind turbine (voltage, wind speed). The FDI problem for time varying rotational speed of the rotor will also be studied.

ACKNOWLEDGMENTS

This work was conducted in part of a concerted research financially supported by the "Couneil Regional Haute Normandie Service Recherche, Innovation".

REFERENCES

Blanke, M., Kinnaert, M., Lunze, J., Staroswieck, M. i and Schroder, J., 2010. Diagnosis and fault tolerant control. *2nd ed. Springer.*

Chen, J., and Patton, R. J.,1999 . Robust model-based fault diagnosis for dynamic systems. *Kluwer academic publishers*, first edition.

Clark, R. N.,1978. Instrument fault detection. *IEEE trans. aerospace electron systems*, vol. AES14, pp. 456–465.

El Aimani, S:, 2004. Modeling of different technologies of wind turbines integrated into a network of medium voltage. *Thesis (PhD).* Ecole Centrale in Lille (France).

Frank, P. M., 1987. System fault diagnosis reliability and related knowledge-based approaches. *D. Reidel Publishing Company.*, vol. 1 – Fault. diagnostics and Realibility, ch. Fault diagnosis in dynamic systems viastate estimation - A Survey, pp. 35–98.

Gálvez-Carrillo, M., Kinnaert, M., 2011. Sensor fault detection and isolation in doubly-fed induction generators accounting. for parameter variations. *Renewable energy*, 36 1447e1457.

Ichalal, D., Marx, B., Ragot J., and Maquin, D.,2006. Diagnostic of the nonlinear systems by multi-model approach. *JIME* 2006, Poitiers (France).

Khojet EL Khail, Sejir., 2006. Commande vectorielle d'une machine asynchrone doublement alimentée (MADA). *PhD Thesis*. Polytechnique of Toulouse.

Ouyessaad, H., Chafouk, H.,and Lefebvre D., 2011. Détection de défaut des capteurs de courant dans une GADA basée sur le banc de filtres de Kalman. *3SGS Conference* university of Valenciennes

Odgaard, P.F., and Stoustrup, J., 2012. Fault tolerant control of wind turbines using unknown input observers. *In Proceedings of the 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Mexico City, Mexico.

Odgaard, P. F., Thogersen, P. B., and Stoustrup, J., 2009. Fault isolation in parallel coupled wind turbine converters. *In Proceedings of European Wind Energy Conference 2009.* Marseille, France,

Odgaard, P.F., and Stoustrup, J., 2009. Unknown input observer based scheme for detecting faults in a wind turbine converter. *In proceedings of the 7th IFAC symposium on fault detection, supervision and safety of technical trocesses*, page 161–166, Barcelona, Spain.

Odgaard, P., and Mataji, B., 2008. Observer based fault detection and moisture estimating in coal mill. *Control engineering practice*. pages:909–921.

Orjuela, R., Marx, B., Ragot, J., and Maquin, D., 2010. Diagnostic of the nonlinear systems by multi-model approach. *Sixth International Conference of automatic, CIFA 2010*, Nancy (France)

Rothenhagen, K. and F. Fuchs, W., 2009. Doubly fed induction generator model-based sensor fault detection and control loop reconfiguration. *IEEE transactions on industrial electronics*, vol. 56, no. 10, pp. 4229–4238.

Rothenhagen, K., Thomsen, S., and Fuchs, F. W., 2007.Voltage sensor fault detection and reconfiguration for a doubly fed induction generator. *In IEEE international symposium on diagnostics for Electricmachines, power electronics and drives*. Amiens, France, pp.377–382.

Trinh, D., H., and Chafouk, H., 2011. Fault diagnosis and isolation using kalman filter bank for a wind turbine generator. *IEEE/MED Conference CORFU*, Greece.

Wei, X., Verhaegen, M., and Van den engelen T., 2008. Sensor fault diagnosis of wind turbines for fault tolerant. *In Proceedings of the 17th world congress the international federation of automatic control*, page 3222–3227, Seoul, South Korea,

Unknown input observer: a physical approach

Dapeng Yang, Christophe Sueur
Ecole Centrale de Lille, LAGIS UMR CNRS 8219, BP 48, 59651, Villeneuve d'Ascq Cedex France
E-mail: christophe.sueur@ec-lille.fr

Abstract— **The object of this paper is the study of a new unknown input observer for linear models. This new observer has some classical restrictive conditions proposed for other ones: infinite structure condition and Hurwitz conditions for invariant zeros. The main contribution are twofold: this new observer has the same state space representation as the initial state model and it is shown to be very accurate. Simulation results are proposed for a DC motor example with analysis of two other classical methodologies.**

Keywords: Unknown input observers, Bond graph, linear models, invariant zeros

1. INTRODUCTION

The unknown input and state observability problem (UIO) is a well known problem because for control design with a state space approach, the state vector $x(t)$ cannot be entirely measured and the system is often subject to unknown inputs $d(t)$ (disturbance or failure...) which must be estimated, as proposed in the state space representation (1). In this state equation $x \in \Re^n$ is the state vector, $z \in \Re^p$ is the vector set of measured variables and $y \in \Re^p$ is the vector of output variables to be controlled. The input variables are divided into two sets $u \in \Re^m$ and $d(t) \in \Re^q$ which represent known and unknown input variables respectively.

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Fd(t) \\ z(t) = Hx(t) \\ y(t) = Cx(t) \end{cases} \tag{1}$$

Different approaches give solvability conditions and constructive solutions for the unknown input observer problem. For LTI models, constructive solutions with reduced order observers are first proposed with the geometric approach [13], [4], [2]. Constructive solutions based on generalized inverse matrices taking into account properties of invariant zeros are given in [19] and then in [20] and [15] with observability and detectability properties. Full order observers are then proposed in a similar way (based on generalized inverse matrices) in [7] and [6], but with some restriction on the infinite structure of the model. The algebraic approach is proposed in [25] and in [5] for continuous and discrete time systems, without restriction on the infinite structure of the model. The structural invariants which play a fundamental role in this problem have been extensively studied in many papers and books [2], [21], [23], [16], [12], [17]. The knowledge of zeros is often an important issue because zeros are directly related to some stability conditions of the controlled system and the infinite structure is often related to solvability conditions.

The objective of this paper is the development of a new observer for linear systems when there are two kinds of inputs: measured and unmeasured inputs. The second section deals with the problem statement in an usual way, with the recall of two classical approaches and some conditions of application. In the third section, the new approach is proposed and then an application on a DC motor is proposed. Simulations are proposed for this new observer and for two classical approaches.

2. PRELIMINARIES

2.1. UIO existence conditions

In the literature, the different proposed approaches consider first the finite structure of $\Sigma(H,A,F)$ and then its infinite structure. The finite structure gives some stability conditions on the UIO and the infinite structure some conditions on the existence of the UIO.

The concepts of strong detectability, strong* detectability and strong observability have been proposed in [14]. System (1) (with only unknown input $d(t)$) is strongly detectable if $z(t) = 0$ for $t > 0$ implies $x(t) \to 0$ with $(t \to \infty)$ and system (1) is strong* detectable if $z(t) = 0$ for $t \to \infty$ implies $x(t) \to 0$ with $(t \to \infty)$.

The strong detectability corresponds to the minimum-phase condition, directly related to the zeros of system $\Sigma(H,A,F)$ (finite structure) defined as to be the values of $s \in \mathscr{C}$ (the complex plane) for which (2) is verified.

$$\mathtt{rank}\begin{pmatrix} sI - A & -F \\ H & 0 \end{pmatrix} < n + \mathtt{rank}\begin{pmatrix} -F \\ 0 \end{pmatrix} \tag{2}$$

Proposition 1: [14] The system $\Sigma(H,A,F)$ in (1) is strongly detectable if and only if all its zeros s satisfy $Re(s) < 0$ (minimum phase condition).

Proposition 2: [14] The system $\Sigma(H,A,F)$ in (1) is strong* detectable if and only if it is strongly detectable and in addition rank$[HF]$=rank$[F]$.

Proposition 3: [14] The system $\Sigma(H,A,F)$ in (1) is strongly observable if and only if it has no zeros.

The infinite structure of multivariable linear models is characterized by different integer sets. $\{n_i'\}$ is the set of infinite zero orders of the global model $\Sigma(C,A,B)$ and $\{n_i\}$ is the set of row infinite zero orders of the row sub-systems $\Sigma(c_i,A,B)$. The infinite structure is well defined in case of LTI models [8] with a transfer matrix representation or with a graphical representation (structured approach), [9].

The row infinite zero order n_i verifies condition $n_i = \min \left\{k|c_iA^{(k-1)}B \neq 0\right\}$. n_i is equal to the number of derivations of the output variable $y_i(t)$ necessary for at least one of the input variables to appear explicitly. The global infinite zero orders [10] are equal to the minimal number of derivations of each output variable necessary so that the input variables appear explicitly and independently in the equations. The infinite structure can also be defined for others models, such as $\Sigma(H,A,F)$.

In order to solve the UIO problem for systems in (1), a necessary condition called *observer matching condition* for the existence of observers is often required (see [19]; [7]): rank$[HF]$ = rank$[F]$. For a SISO model, the infinite zero order of model $\Sigma(H,A,F)$ is equal to 1. When this condition is not satisfied [11] proposed unknown input sliding mode observers after implementing a procedure to get a canonical observable form of systems. This method can also be extended in the nonlinear case. [5] and [1] gave an intrinsic solution with an algebraic approach. Necessary and sufficient conditions are that system $\Sigma(H,A,F)$ is left invertible and minimum phase. The LTI system $\Sigma(H,A,F)$ in (1), supposed to be asymptotically observable with unknown input, is rapidly observable if, and only if, it zero dynamics is trivial.

2.2. UIO Synthesis

Two approaches are recalled in this paper. The goal is to compare the performances of these observers with the new one proposed in this paper.

Approach with pseudo-inverse: An observer proposed by [6] has the form (3).

$$\begin{cases} \dot{\xi}(t) = N\xi(t) + Jz(t) + Gu(t) \\ \hat{x}(t) = \xi(t) - Ez(t) \end{cases} \tag{3}$$

where $\hat{x}(t) \in \Re^n$ is the estimate of $x(t)$. Matrices N, J, G and E with constant entries have appropriate dimensions. [6] studied the model with unknown inputs in the state and in the measurement equations. Here the model is simplified without considering unknown inputs in the measurement equation, i.e., $z = Hx(t) + Dd(t), D = 0$.

Let $P = I + EH$, the observer reconstruction error is $e = x - \hat{x} = Px - \xi$. The dynamic of the estimation error is given by $\dot{e} = Ne + (PA - NP - JH)x + (PB - G)u + PFd$. Hence, in the error variable equation some relations must be satisfied: $PA - NP - JH = PB - G = PF = 0$. In order to solve these equations, some generalized inverse matrices must be defined because in the previous equations some matrices are not square. Moreover, model $\Sigma(H,A,F)$ has a stable observer if the model is strong* detectable.

Approach with output derivation: The unknown input observer for a SISO model [5] with control input is written in (4).

$$\begin{cases} \dot{\hat{x}} = (PA - LH)\hat{x} + Q\left(z^{(r)} - U\right) + Lz + Bu \\ \hat{d} = \left(HA^{r-1}F\right)^{-1}\left(z^{(r)} - HA^r\hat{x} - U\right) \end{cases} \tag{4}$$

r is the infinite zero order of $\Sigma(H,A,F)$. $\hat{d}$ is the estimation of d and the matrices Q and P verify: $Q = F\left(HA^{r-1}F\right)^{-1}$, $P = I_n - QHA^{r-1}$, and $U = \sum_{i=0}^{r-1} HA^iBu^{(r-1-i)}$. The main idea of the method is to implement derivations on the output variable $z(t)$ to let the unknown input variable $d(t)$ appears explicitly. Note that the control input must be derivable ($r-1$ times). For MIMO models, the extension of the procedure was proposed by [11].

The dynamic of the estimation error of state variables is $\dot{e} = \dot{x} - \dot{\hat{x}} = (PA - LH)(x - \hat{x})$. One has $\lim_{t\to\infty} e(t) = 0$ for any $x(0)$, $\hat{x}(0)$, $d(t)$ and $u(t)$. The estimation of d can be written as $\hat{d} = \left(HA^{r-1}F\right)^{-1}HA^r(x - \hat{x}) + d$. As $\lim_{t\to\infty} e(t) = 0$, then $\lim_{t\to\infty}\hat{d}(t) = d(t)$.

This observer is stable if the finite structure of $\Sigma(H,A,F)$ is stable.

3. NEW APPROACH

3.1. New UIO

If a somewhat physical approach is proposed, some assumptions are possible for the state space model deduced for example from a bond graph representation. The model $\Sigma(H,A,F)$ is also supposed to be a SISO model.

Asumption 1. It is supposed that the SISO system $\Sigma(H,A,F)$ defined in (1) is controllable/observable and that the state matrix A is invertible.

With Asumption 1, a derivative causality assignment is possible for bond graph models (physical model without null pole). The extension to models with non invertible state matrix is straight for bond graph models, because a graphical approach can be proposed in that case. It is not proposed in this paper.

The state equation (1) without output variable y is now rewritten as (5).

$$\begin{cases} x(t) = A^{-1}\dot{x}(t) - A^{-1}Bu(t) - A^{-1}Fd(t) \\ z(t) = HA^{-1}\dot{x}(t) - HA^{-1}Bu(t) - HA^{-1}Fd(t) \end{cases} \tag{5}$$

If matrix $HA^{-1}F$ is invertible (Model $\Sigma(H,A,F)$ has no null invariant zero), the disturbance variable can be written in equation (6) and then the estimation of the disturbance variable can be written in equation (7).

$$d(t) = -(HA^{-1}F)^{-1}[z(t) - HA^{-1}\dot{x}(t) + HA^{-1}Bu(t)] \tag{6}$$

$$\hat{d}(t) = -(HA^{-1}F)^{-1}[z(t) - HA^{-1}\dot{\hat{x}}(t) + HA^{-1}Bu(t)] \tag{7}$$

From the state equation (5), a new estimation is proposed for the state vector, defined in equation (8), which can also be written as (9), which is similar to a classical estimation, but with a difference in the last term. It needs the derivation of the measured variable. Matrix K is used for pole placement.

$$\hat{x}(t) = A^{-1}\dot{\hat{x}}(t) - A^{-1}Bu(t) - A^{-1}F\hat{d}(t) + K(\dot{z}(t) - \dot{\hat{z}}(t)) \tag{8}$$

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + F\hat{d}(t) - AK(\dot{z}(t) - \dot{\hat{z}}(t)) \qquad (9)$$

For these three observers, the estimate of the state vector is the solution of a first order differential state equation which is not the state equation of the model for the two first observers. In our approach, the state equation is the same (model and observer), with only an extra term for the observer. This new observer is thus much more simpler. Note that most of the works proposed in the literature do not take into account the control inputs.

3.2. *Properties of the observer*

In this section, some properties of the UIO are enounced and proved. It is proved that this new observer requires the matching condition defined in some well known approaches [14], [6] and that in that case, fixed poles of the estimation error are all the invariant zeros of system $\Sigma(H,A,F)$, which means that this system must be strong* detectable.

The convergence of the disturbance variable can be verified with equation (10), obtained from (6) and (7).

$$d(t) - \hat{d}(t) = (HA^{-1}F)^{-1}HA^{-1}(\dot{x}(t) - \dot{\hat{x}}(t)) \qquad (10)$$

The estimation of the disturbance variable converges to the disturbance variable only if $(\dot{x}(t) - \dot{\hat{x}}(t))$ converges asymptotically. Convergence of the state estimation must be proved with the study of the observer fixed poles.

In order to simplify notations, new matrices N_{BO} and N_{BF} are introduced in (11).

$$\begin{cases} N_{BO} = A^{-1} - A^{-1}F(HA^{-1}F)^{-1}HA^{-1} \\ N_{BF} = A^{-1} - A^{-1}F(HA^{-1}F)^{-1}HA^{-1} - KH \end{cases} \qquad (11)$$

From (5) and (8), with $e(t) = x(t) - \hat{x}(t)$ it comes (12).

$$e(t) = N_{BF}\dot{e}(t) \qquad (12)$$

In equation (12), conditions for pole placement are studied. If matrix N_{BF} is invertible, a classical pole placement is studied, and the error variable $e(t) = x(t) - \hat{x}(t)$ does not depend on the disturbance variable. The conditions for (8) to be an asymptotic state observer of $x(t)$ is that N_{BF} must be an Hurwitz matrix, i.e., has all its eigenvalues in the left-hand side of the complex plane. Properties of the observer are studied in the next part.

A necessary condition for the existence of the state estimator is proposed in Proposition 4.

Proposition 4: A necessary condition for matrix N_{BF} defined in (11) to be invertible is that $HF \neq 0$.

Proof In Proposition 4, matrix $N_{BF}F$ is equal to $[A^{-1} - A^{-1}F(HA^{-1}F)^{-1}HA^{-1} - KH]F$, thus it can be rewritten as $N_{BF}F = A^{-1}F - A^{-1}F(HA^{-1}F)^{-1}HA^{-1}F - KHF = KHF$. If condition $HF \neq 0$ is not satisfied, the Kernel of matrix N_{BF} is not empty, which means that matrix N_{BF} is not invertible and that this matrix contains at least one null mode, thus pole placement is not possible (all its eigenvalues are not in the left-hand side of the complex plane).□

Condition defined in proposition 4 is exactly the same condition defined for the well-known observers defined in [14] and [6]. It means that the infinite zero order between the disturbance variable $d(t)$ and the measured variable $z(t)$ is equal to 1.

It is now supposed that $HF \neq 0$ is satisfied. Two properties are proved. First, it is proved that matrix N_{BO} has one eigenvalue equal to 0 and that the other eigenvalues are the inverse of the invariant zeros of system $\Sigma(H,A,F)$. In that case in matrix N_{BF}, fixed modes are the inverse of the invariant zeros of model $\Sigma(H,A,F)$ and the only eigenvalue which can be chosen is related to the null eigenvalue of matrix N_{BO}.

Proposition 5: In matrix N_{BF} defined in (11), only 1 pole can be chosen with matrix K.

Proof Since $HN_{BO} = HA^{-1} - HA^{-1}F(HA^{-1}F)^{-1}HA^{-1}$, it comes $HN_{BO} = 0$. If a state model is considered with the state matrix N_{BO} and the output vector H, it is proved that this model is not observable since vector H^t is orthogonal to the state matrix. Moreover, the rank of the observability matrix of $\Sigma(H,N_{BO})$ is equal to 1. If a classical state estimation is proposed for this model, only 1 pole can be assigned, which is also true for matrix N_{BF}, because matrix N_{BF} can be considered as the state matrix of an equation error when estimating the state of system $\Sigma(H,N_{BO})$.□

Proposition 6: The eigenvalues of matrix N_{BO} defined in (11) are the inverse of the invariant zeros of system $\Sigma(H,A,F)$ ($n-1$ modes) plus 1 eigenvalue equal to 0.

Proof: appendix A

Proposition 7: The fixed poles of the estimation equation defined in (12) are the invariant zeros of system $\Sigma(H,A,F)$.

Proof From Proposition 6, the eigenvalues of matrix N_{BO} are the inverse of the invariant zeros of system $\Sigma(H,A,F)$ with an eigenvalue equal to 0, and since N_{BF} is invertible and only one pole can be chosen, all the fixed poles are the non null eigenvalues.□

4. EXAMPLE

The previous procedures are applied on a DC motor modeled by bond graph [18] and [22]. At the analysis step, proposed methods on bond graph models do not require the knowledge of the value of parameters, because intrinsic solvability conditions can be given and a formal calculus can be proposed at the synthesis level. First some properties of bond graph models are recalled and then the example is studied. Note that this system could be studied without the bond graph approach, and that other state variables could be chosen.

4.1. *Bond graph approach*

In a bond graph model, causality and causal paths are useful for the study of properties, such as controllability, observability and systems poles/zeros. Bond graph models with integral causality assignment (BGI) can be used to determine reachability conditions and the number of invariant zeros by studying the infinite structure. The rank of the controllability matrix is derived from bond graph models with derivative causality (BGD) [24].

An LTI bond graph model is controllable if and only if the two following conditions are satisfied: first there is a causal path between each dynamical element and one of the input sources and secondly each dynamical element can have a derivative causality assignment in the bond graph model with a preferential derivative causality assignment (with a possible duality of input sources). The observability property can be studied in a similar way, but with output detectors. Systems invariant zeros are poles of inverse systems. Inverse systems can be constructed by bond graph models with bicausality (BGB) which are thus useful for the determination of invariant zeros.

The concept of causal path is used for the study of the infinite structure of the model.The order of the infinite zero for the row sub-system $\Sigma(h_i, A, F)$ is equal to the length of the shortest causal path between the i^{th} output detector z_i and the set of disturbance input sources. The global infinite structure is defined with the concepts of different causal paths (not recalled here). The number of invariant zeros is determined by the infinite structure of the BGI model. The number of invariant zeros associated to a controllable, observable, invertible and square bond graph model is equal to $n - \sum n'_i$. For bond graph models, invariant zeros equal to zero can be directly deduced from the infinite structure of the BGD model [3].

An example of a DC motor is used to show the procedure for designing the UIO observer. The BGI model of the system with a disturbance signal is given in Fig. 1, and the state-space equations are presented in (13), with $x = (p_L, p_J)^t = (x_1, x_2)^t$ the state vector. Since the state equation is written from a bond graph model, the state vector contains energy variables, for example p_L is the magnetic flux in the inductance. z is the measured output variable, it is the rotational speed of the motor drive shaft, $z = \frac{p_J}{J}$. u is the control input variable and d the disturbance input variable (disturbance torque). The input $u(t)$ is the Heaviside unit step function, i.e. $u(t) = 100\Gamma(t)$. The disturbance d of the system is a pulse signal with start time $0.005s$, end time $0.006s$ and amplitude $10N$.

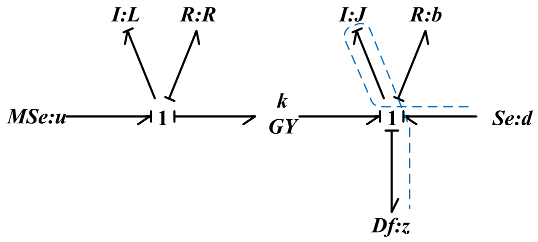


Fig. 1. BGI model of the DC moteur

$$\begin{cases} \dot{x}_1 = -\frac{R}{L}x_1 - \frac{k}{J}x_2 + u \\ \dot{x}_2 = \frac{k}{L}x_1 - \frac{b}{J}x_2 + d \\ z = \frac{1}{J}x_2 \end{cases} \tag{13}$$

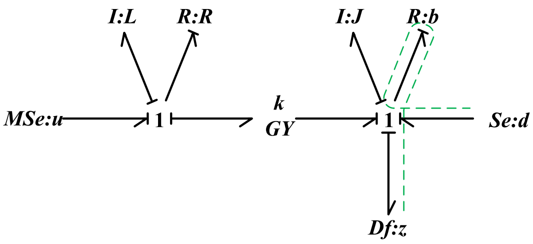


Fig. 2. BGD model of the DC moteur

The bond graph model is controllable and observable (a derivative causality can be assigned, fig. 2). The numerical values of system parameters are shown in Table I. In this part, some simulation results will be shown by the bond graph software 20-sim®.

TABLE I

NUMERICAL VALUES OF SYSTEM PARAMETERS

L	R	k	J	b
1.6×10^{-4} H	0.29 Ω	0.225	1×10^{-4} kgm^2	1×10^{-4} Nm/Wb

The rotor angular velocity ω and it's estimate $\hat{\omega}$ with an initial condition for the state variables $x_1 = 0$ and $x_2 = 0.01$ are studied in each case. Then the disturbance variable d and it's estimate $\hat{d}$ and the estimation errors for the state variables are drawn.

4.2. New observer

The design of the observer proposed in the previous section can thus be redesigned from a bond graph approach.

The causal path length between the output detector $Df : z$ and the disturbance input $Se : d$ is equal to 1, Fig. 1, path $Df : z \rightarrow I : J \rightarrow Se : d$, thus the matching condition is verified, and there is an invariant zero in the system $\Sigma(H, A, F)$. After calculations or analysis of the bond graph model with a bicausal assignment, the invariant zero is $s = -\frac{R}{L}$ which verifies the minimum phase condition. The bond graph representation of the observer is drawn in Fig. 3 in a general form without values for parameters.

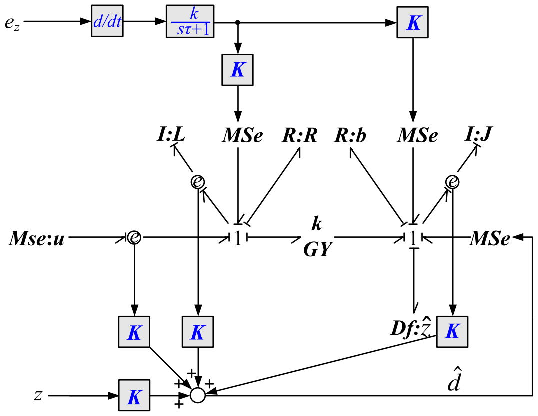


Fig. 3. Observer with the bond graph representation

In the state estimation equation defined in (8), matrix $K = [k_1, k_2]^t$ is used for pole placement. With some formal calculus, the two poles of matrix N_{BF}^{-1} defined in the state estimation error equation (12) are $s=-\frac{R}{L}$ and $s=-\frac{J}{k_2}$. The first fixed pole is equal to $-\frac{R}{L}=-1812.5$. The second one is chosen at $s=-2000$, thus $k_2=0.2$.

The two estimated variables $\hat{\omega}$ and $\hat{d}$ are very close to the real variables, Fig. 4, Fig. 5 and the estimation errors for the state variables converge rapidly to zero, Fig. 6. The different figures prove the accuracy of this new UIO.

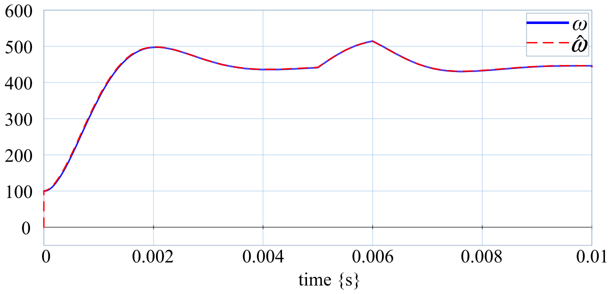


Fig. 4. The rotor angular velocity ω and it's estimate $\hat{\omega}$

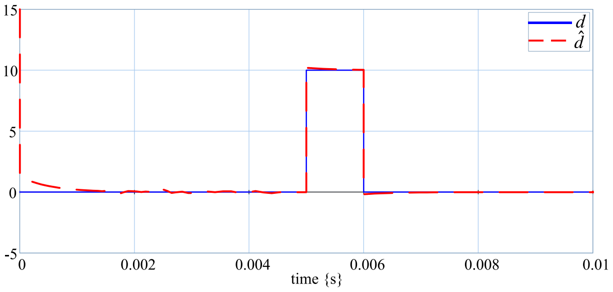


Fig. 5. The disturbance variable d and it's estimate $\hat{d}$

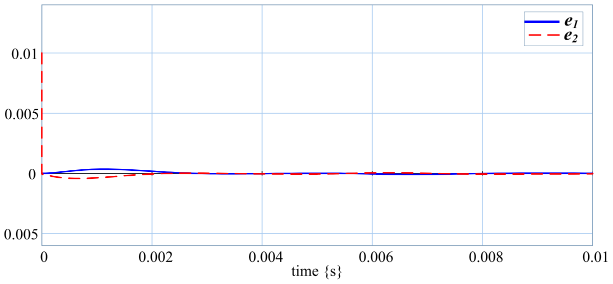


Fig. 6. Trajectories $e_i=x_i-\hat{x}_i, i=1,2$ with UIO in (9)

4.3. Other observers

In order to compare the different observers, the same model is studied with the two other observers.

Matrix Z is used to place poles of the observer for the UIO defined with generalized inverse matrices. One pole is fixed (invariant zero of system $\Sigma(H,A,F)$), another is placed at $s=-2000$. The matrices of the observer are

$$N=\begin{bmatrix}-1812.5 & -2250\\ 0 & -2000\end{bmatrix}\quad E=\begin{bmatrix}0\\ -0.0001\end{bmatrix}$$

$$J=\begin{bmatrix}-0.225\\ 0\end{bmatrix}\qquad G=\begin{bmatrix}1\\ 0\end{bmatrix}$$

Variables ω and $\hat{\omega}$, Fig. 7, are very close. The estimation errors for the two state variables are displayed in Fig. 8. [6] did not proposed the estimate of the unknown input. In that case, results are similar, but the structure of the observer is much more complex.

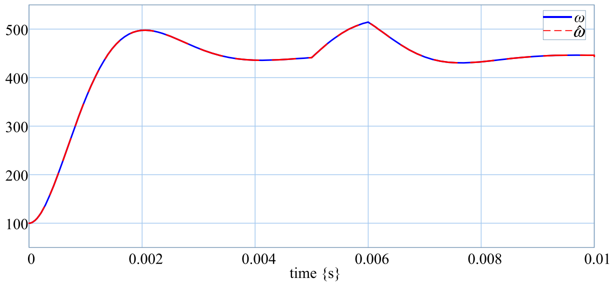


Fig. 7. The rotor angular velocity ω and it's estimate $\hat{\omega}$

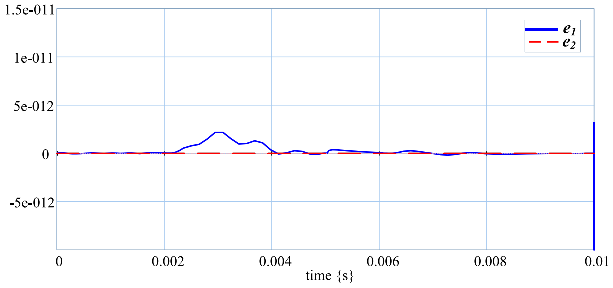


Fig. 8. Trajectories $e_i=x_i-\hat{x}_i, i=1,2$ of the system

With the algebraic approach, Matrix L is used to place poles of the observer. One pole is fixed (invariant zero of system $\Sigma(H,A,F)$), another is placed at $s=-2000$. The matrices of the observer are $Q=\begin{bmatrix}0\\ 0.0001\end{bmatrix}$ $PA-LH=\begin{bmatrix}-1812.5 & -2250\\ 0 & -2000\end{bmatrix}$

$$L=\begin{bmatrix}0\\ 0.2\end{bmatrix}\qquad U=0$$

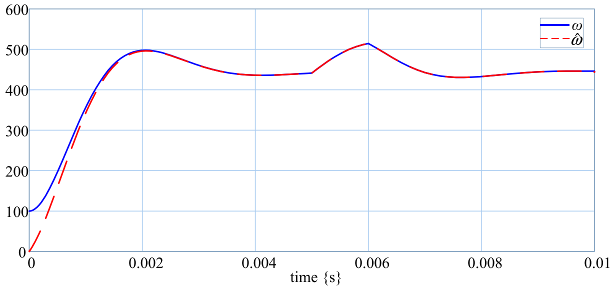


Fig. 9. The rotor angular velocity ω and it's estimate $\hat{\omega}$

Variables ω and $\hat{\omega}$ are drawn in Fig. 9. The estimation of d is $\hat{d}=0.0001(\dot{y}-\begin{bmatrix}14062500 & -10000\end{bmatrix}\hat{x})$. The comparison of the estimation of the unknown input $\hat{d}(t)$ and $d(t)$ is shown in Fig. 10. The estimation errors for the two state variables are displayed in Fig. 11. In that case, our new observer is more accurate.

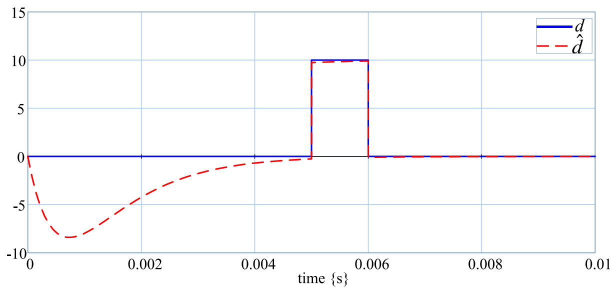


Fig. 10. Trajectories d and $\hat{d}$ of the system

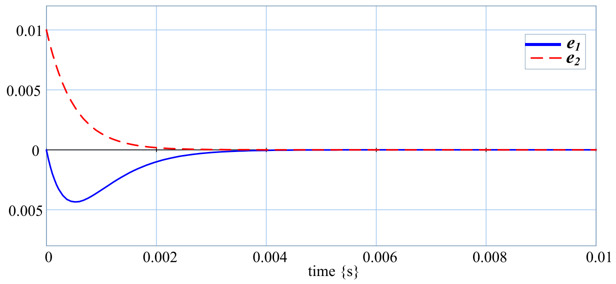


Fig. 11. Trajectories $e_i = x_i - \hat{x}_i, i = 1,2$ of the system

5. CONCLUSION

In this paper, a new input and state observer is proposed for linear systems. The classical assumption on the strong* detectability property is necessary. Two significant facts concern the simplicity of the observer synthesis and the efficiency of this observer. This observer is proposed in a SISO context, but can be easily extended to linear MIMO Systems.

REFERENCES

1. J. P. Barbot, M. Fliess, and T. Floquet. An algebraic framework for the design of nonlinear observers with unknown inputs. *Proceedings of the 46th IEEE CDC, New Orleans, USA, Dec. 12-14*, 2007.
2. G. Basile and G. Marro. A new characterization of some structural properties of linear systems: Unknown-input observability, invertibility and functional controllability. *International Journal of Control*, 17(5):931943, 1973.
3. J. M. Bertrand, C. Sueur, and G. Dauphin-Tanguy. Bond graph for modeling and control: structural analysis tools for the design of input-output decoupling state feedbacks. *Proceedings of International Conference on Bond Graph Modeling and Simulation, Phoenix*, pages 103–108, 1997.
4. S.P. Bhattacharyya. Observer design for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 23:1483–484, 1978.
5. J. Daafouz, M. Fliess, and G. Millerioux. Une approche intrinsèque des observateurs linéaires à entrèes inconnues. *CIFA 2006, Bordeaux*, 2006.
6. M. Darouach. Complements to full order observer design for linear systems with unknown inputs. *Applied Mathematics Letters*, 22:1107–1111, 2009.
7. M. Darouach, M. Zazadinski, and S.J. Xu. Full-order observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 39:606–609, 1994.
8. J. M. Dion and C. Commault. Smith-mcmillan factorizations at infinity of rational matrix functions and their control interpretation. *System & Control Letters*, 1:312–320, 1982.
9. J. M. Dion and C. Commault. Feedback decoupling of structured systems. *IEEE Transactions on Automatic Control*, 38:1132–1135, 1993.
10. P. L. Falb and W. A. Wolovich. On the decoupling of multivariable systems. *Preprints JACC, Philadelphia*, pages 791–796, 1967.
11. T. Floquet and J.P. Barbot. *Advances in variable structure and sliding mode control*, chapter A canonical form for the design of unknown input sliding mode observers, pages 271–292. Springer, 2006.
12. E. G. Gilbert. The decoupling of multivariable systems by state feedback. *SIAM Journal of Control and Optimization*, 7:50–63, 1969.
13. R. Guidorzi and G. Marro. On wonham stabilizability condition in the synthesis of observers for unknown-input systems. *IEEE Trans. Automat. Control*, 16:499–500, 1971.
14. M. L. J. Hautus. Strong detectability and observers. *Linear Algebra and its Applications*, 50:353–368, 1983.
15. M. Hou and P.C. Muller. Design of observers for linear systems with unknown inputs. *IEEE Trans. Automat. Control*, 37:871–875, 1992.
16. T. Kailath. *Linear Systems*. Prentince Hall, Englewood-Cliff, N.J., 1980.
17. R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in Mathematical System Theory*. McGraw-Hill, New York, 1969.
18. D. Karnopp, D. Margolis, and R. Rosenberg. *System dynamics: a unified approach*. John Wiley & Sons, 1975.
19. P. Kudva, N. Viswanadham, and A. Ramakrishna. Observers for linear systems with unknown inputs. *IEEE Trans. Automat. Control*, 25:113–115, 1980.
20. B.J. Miller and R. Mukunden. On designing reduced-order observers for linear time-invariant systems subject to unknown inputs. *Internat. J. Control*, 35:183–188, 1982.
21. A. S. Morse. Structural invariants of linear multivariable systems. *SIAM J. Control*, 11:446–465, 1973.
22. R. Rosenberg and D. Karnopp. *Introduction to physical system dynamics*. McGraw Hill, 1983.
23. H.H. Rosenbrock. *State space and multivariable theory*. Nelson, London, England, 1970.
24. C. Sueur and G. Dauphin-Tanguy. Bond-graph approach for structural analysis of mimo linear systems. *Journal of the Franklin Institute*, 328:55–70, 1991.
25. H. L. Trentelman, A. A. Stoorvogel, and M. Hautus. *Control theory for linear systems*. London, UK: Springer, 2001.

APPENDIX

Consider the quadruple $\Sigma(A^*, B^*, C^*, D^*)$, with $A^* = A^{-1}$, $B^* = -A^{-1}F$, $C^* = HA^{-1}$ and $D^* = -HA^{-1}F$. This model is controllable and observable, thus the invariant zeros of this model are the zeros of the Smith matrix defined in (14).

$$S(s) = \begin{pmatrix} sI - A^{-1} & A^{-1}F \\ HA^{-1} & -HA^{-1}F \end{pmatrix} \tag{14}$$

With the usual properties of matrix determinant, it comes $\det S(s)$=$\det HA^{-1}F.\det(sI - A^{-1} + A^{-1}F(HA^{-1}F)^{-1}HA^{-1})$, and thus $\det S(s) = \det(sI - N_{BO})$. The invariant zeros of the quadruple $\Sigma(A^*, B^*, C^*, D^*)$ are the poles of matrix N_{BO}.

With classical symbolic equivalent operation, it comes:

$$\begin{vmatrix} sI - A^{-1} & A^{-1}F \\ HA^{-1} & -HA^{-1}F \end{vmatrix} \sim \begin{vmatrix} sI - A^{-1} & (sI - A^{-1})F + A^{-1}F \\ HA^{-1} & HA^{-1}F - HA^{-1}F \end{vmatrix} \tag{15}$$

$$\det S(s) = \begin{vmatrix} sI - A^{-1} & sF \\ HA^{-1} & 0 \end{vmatrix} \sim \begin{vmatrix} sA - I & sF \\ H & 0 \end{vmatrix} \tag{16}$$

Thus $\det S(s) = \det(sA - I).\det(H(sA - I)^{-1}sF)$ and it comes $\det S(s) = s.\det(s^{-1}I - A).\det(H(s^{-1}I - A)^{-1}F)$. Thus the roots of $S(s)$ are all the inverse of the invariant zeros of system $\Sigma(H, A, F)$ with a root equal to 0.□

REDUCING VIBRATIONS ON FLEXIBLE ROTATING ARMS THROUGH THE MOVEMENT OF SLIDING MASSES: MODELING, OPTIMAL CONTROL AND SIMULATION

Eder Terceiro [(a)]
Agenor de Toledo Fleury [(b)]
Mechanical Engineering Department, Escola Politécnica, Universidade de São Paulo [(a),(b)]
Mechanical Engineering Department, Centro Universitário da FEI [(b)]
ederterceiro@yahoo.com.br [(a)] agfleury@fei.edu.br [(b)]

August 16, 2012

ABSTRACT

This paper brings contributions on the proposal of use of translational motions of sliding masses to minimize vibrations induced by the rotational motion of a light flexible manipulator (rotating arm). This system is inspired by rotating cranes used to transport loads. Optimal control methods have been used to generate the slider trajectories while the flexible manipulator performs a rotational maneuver from a fixed to other fixed configuration. This approach has led to good solutions even in case of quite quick maneuvers, as, for example, a 90^o beam rotation in just 1 second, using 1 or 2 sliders (Terceiro, 2002). In the present paper, the complete motion equations for any number of masses are firstly presented, in order to emphasize the complexity of the coupled elastic-rotational-translational motions. Simplifying assumptions are pointed out and the corresponding optimal control problems (OCP) are obtained. Optimal trajectories, generated according to different Indexes of Performance and different problem parameters, are analysed and compared in order to get feasible movements for the set.

Keywords: Optimal Control, Vibration Control, Flexible Robotics, Lightweight Structures, Composed Motion

1 INTRODUCTION

For many applications, structural flexibilities must be considered in early design stages in order to assure good vibration attenuation in modern machine design. The problem of flexible structures has worried many authors, today's literature on the subject is extensive. We can cite as authors interested: (Junkins and Kim, 1993) dealing with the problem of dynamic and flexible control structure and (Meirotvitch, 1980), (Meirotvitch, 1990) that contributed to the disclosure of the issue of flexible structures.

Based on the examples of rotating cranes and rotational/prismatic joint robots, this work explores simultaneous rotating/translational motions to minimize vibrations on a light one-link manipulator that performs large rotational maneuvers. The basic question investigated is how the motion of independent parts may contribute to reduce the vibration levels of the whole system. As the results achieved are encouraging, it seems feasible to extend the research to other more complex applications. Then, the objectives here are the achievement of suitable system models for a very light flexible manipulator and the synthesis of optimal controllers using the torque applied to the hub where the flexible arm is fixed and the forces applied to the sliding masses as control variables. The case of a single sliding mass has brought surprising results, as shown in (Fleury and Oliveira, 2004). The investigation has been extended to model a mechanism which include any number of sliders and structural modes (Terceiro, 2002). The full approach is firstly introduced in this paper. In all cases, the dynamical models of the structural system have been derived through the Extended Hamilton Principle resulting in a set of coupled integro-differential non linear equations where system parameters are time and space dependent due to changes in the inertia terms. Using substructuring techniques, arm and sliders motions have been separated and systems responses have been expanded in products of spatial and time functions. Many Control techniques (LQR, for example) have been used in order to minimize vibrations induced by the rotational movements, but this reduction resulted dependent on the prescribed motions of the sliders. A bad choice for the slider movements can lead to larger vibrations amplitudes when compared to a situation where the masses remain fixed on the rotating arm (Fleury and Oliveira, 2004). An Optimal Control line of investigation became mandatory to understand the very influence of the composed torque-sliders position controls on the elastic vibrations. Slider trajectories become control variables, among hub torque leading to Optimal Control Problems (OCP). The resulting models are non linear and time variant and analytical solutions are not feasible. Then, optimal arm and slider trajectories are investigated through the use of RIOTS'95 (Schwartz et al, 1997), a computational package based on the Consistent Approximation Theory (Schwartz, 1966). Among many already simulated cases, the results presented in this paper demonstrate the importance and influence of the choice of Objective Functions in System performance. Early results appear in the work of (Oliveira, 2000) when the problem was treated with only one mass sliding. Then, in (Terceiro, 2002) problem has been generalized to different masses has been established, and then using the RIOTS'95 was simulated problem with two sliding masses. In a research work hard, many results have appeared as shown in (Terceiro and Fleury, 2008). Sometimes, the differences seem subtle, and are basically set the simulation time, the performance criterion and the initial guess necessary for the simulation. Even these few changes have produced a wealth of results that are now published and others that need to be studied further. The numerical difficulties inherent problems of this size are increased by the large number of parameters available for analysis. Just a few results are presented and refer to the most interesting cases selected from a broader set of results that were obtained by the choice of all parameters involved and discussed in the previous paragraph.

2 SYSTEM FULL MODEL

As shown in Figure 1, system is composed by a long, slender, flexible beam (the arm) that can rotate in an horizontal plan driven by the torque delivered in a rigid hub. Angular acceleration and disacceleration of the flexible arm should cause large amplitude vibrations. Here, we propose to move some masses (sliders) simultaneously to the arm motion, thus changing rotational inertia properties to minimize arm vibrations, represented by the arm tip excursions. In order to use Hamiltons Extended Principle, kinetic and potential energies of each component, hub, arm and masses must be calculated. Then, the elastic potential energy of the flexible arm, $U(t)$, is given by:

$$U(t) = \frac{1}{2}\int_0^L EI_v \left(\frac{\partial^2 e}{\partial x^2}\right)^2 dx \qquad (1)$$

with **e** the deformation of the arm at a generic point x, I_v is the moment of inertia, b is the width h the height of a tipical section. L is the length of the flexible arm. The total kinetic energy of the arm, T_B,is:

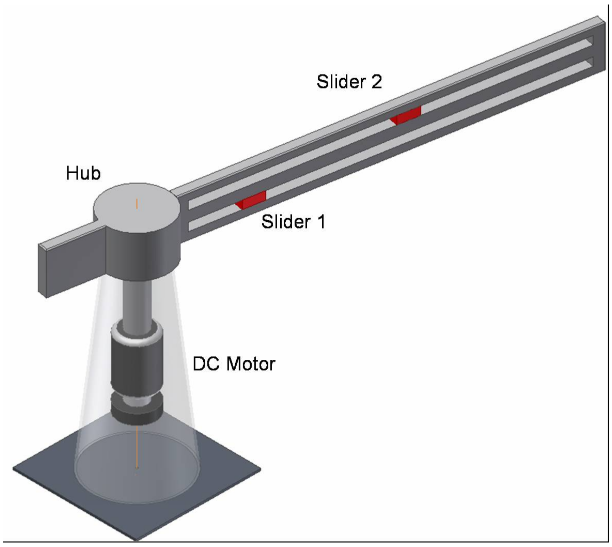


Figure 1: Flexible Arm Model

$$T_B = \int_0^L \frac{1}{2}\rho(e\dot{\theta})^2 dx + \int_0^L \frac{1}{2}\rho(\dot{e}+\dot{x}\theta)^2 dx \qquad (2)$$

with ρ arm linear mass density, θ the angular displacement and $\dot{\theta}$ the angular velocity.

The kinetic energy of the two masses, T_M, is given by:

$$T_M = \frac{1}{2}\sum_{i=1}^{2} M_i\left((\dot{l}_i - e_i\dot{\theta})^2 + (\dot{e}_i + l_i\dot{\theta})^2\right) \qquad (3)$$

l_i is the location on the mass of each mass over the arm.

The kinetic energy of the hub, T_c is:

$$T_c = \frac{1}{2}\frac{M_c L_c^2}{2}\left(\frac{d\theta}{dt}\right)^2 = \frac{1}{2}J_c\left(\frac{d\theta}{dt}\right)^2 = \frac{1}{2}J_c\dot{\theta}^2 \qquad (4)$$

L_C is the radius of the hub, M_C its mass and J_C is its moment of inertia of the cube.

The work of the nonconservative forces, W, is given by:

$$\delta W = \tau\delta\theta + F_1\delta l_1 + F_2\delta l_2 \qquad (5)$$

The first term refers to the virtual work of the applied torque, the other two terms refer to the virtual work of the tangent forces applied to the masses m_1 and m_2 respectively.

Explicitly, τ is the torque applied by the motor to produce rotational movement of the flexible arm, and F_1,

F_2 is the force applied by the sliders m_1, m_2 to reduce the vibration of the flexible arm

The Extended Hamilton's Principle states that between two instants t_1 and $t_2$the system energy follows:

$$\int_{t_1}^{t_2} (\delta L_g + \delta W) dt = 0 \tag{6}$$

with L_g, the Lagrangian of the system, is given by:

$$L_g = T - U = T_B + T_M + T_C - U \tag{7}$$

That is true for the system:

$$\int_{t_1}^{t_2} (\delta T_B + \delta T_M + \delta T_C - \delta U + \delta W) dt = 0 \tag{8}$$

The application of the principle leads to the following model

$$\begin{array}{rcl}
\int_0^L \rho(-x\ddot{e} - (x^2 + e^2)\ddot{\theta})dx & & \\
+M_1 \int_0^L (e_1\ddot{l}_1 - l_1\ddot{e}_1 - e_1^2\ddot{\theta})\Delta l_1 dx & & \\
+M_2 \int_0^L (e_2\ddot{l}_2 - l_2\ddot{e}_2 - e_2^2\ddot{\theta})\Delta l_2 dx & & \\
-(M_1 l_1^2 + M_2 l_2^2 + J_c)\ddot{\theta} + \tau & = & 0 \\
\int_0^L \rho(\ddot{e} - x\ddot{\theta} + e\theta^2)dx - \int_0^L EI_v \frac{\partial^4 e}{\partial x^4} dx & = & 0 \\
M_1 \int_0^L (-\ddot{e}_1 + e_1\dot{\theta}^2 + -\dot{l}_1\dot{\theta})\Delta l_1 dx & = & 0 \\
M_2 \int_0^L (-\ddot{e}_2 + e_2\dot{\theta}^2 + -\dot{l}_2\dot{\theta})\Delta l_2 dx & = & 0 \\
M_1 \int_0^L [e_1\ddot{\theta} + e_1\dot{\theta}^2]\Delta l_1 dx + M_1(l_1\dot{\theta}^2\ddot{l}_1) + F_1 & = & 0 \\
M_2 \int_0^L [e_2\ddot{\theta} + e_2\dot{\theta}^2]\Delta l_2 dx + M_2(l_2\dot{\theta}^2\ddot{l}_2) + F_2 & = & 0
\end{array} \tag{9}$$

with boundary conditions:

$$\begin{array}{ll}
e\Big|_{x=0} = 0 & \frac{\partial^2 e}{\partial x^2}\Big|_{x=l} = 0 \\
\frac{\partial e}{\partial x}\Big|_{x=0} = 0 & \frac{\partial^3 e}{\partial x^3}\Big|_{x=l} = 0
\end{array} \tag{10}$$

3 SUBSTRUCTURE SYNTHESIS

The difficulties of the mathematical analysis of the above problem, which involves all the terms of the interaction energy of each of the parties established because requires solving the equations 9 with boundary condition 10 where all parts of the structure appear mixed when to implementing Extended Hamiltons Principle which involves all the terms of the interaction energy of each of the parties established then the strategy was to consider a system consisting of several substructures, determine the motion equations that govern these substructures and then consider the interaction of each structure with the others and their effects on the structure as a whole. This approach is known as Substructure Synthesis and its roots can be found in papers like (Meirovitch and Kwak, 1991). In our case, the arm-hub has been considered as one substructure and each sliding mass as another ones. Hamiltons Extended Principle has been rewritten for each substructure. After many algebraic manipulations, which include disregarding quadratic terms (Terceiro, 2002), the substructured model is given by:

$$\begin{gathered}
\int_0^L \rho(-x\ddot{e} - (x^2 + e^2)\ddot{\theta})dx - J_c\ddot{\theta} + \tau = 0 \\
\int_0^L \rho(\ddot{e} - x\ddot{\theta} + e\theta^2)dx - \int_0^L EI_v \frac{\partial^4 e}{\partial x^4} dx = 0 \\
\ddot{e}_1 + l_1\ddot{\theta} + 2\dot{l}_1 - e_1\dot{\theta}^2 = 0 \\
\ddot{e}_2 + l_2\ddot{\theta} + 2\dot{l}_2 - e_2\dot{\theta}^2 = 0
\end{gathered} \tag{11}$$

with the same boundary conditions as in equation 10.

The free vibration of the arm, after disregarding some second order terms is given by:

$$\begin{gathered}
-\int_0^L \rho x\ddot{e} dx - (J_B + J_C)\ddot{\theta} = 0 \\
\rho(\ddot{e} - x\ddot{\theta}) + EI_v \frac{\partial^4 e}{\partial x^4} = 0
\end{gathered} \tag{12}$$

A change of variables is then introduced:

$$e(x,t) = z(x,t) - x\theta(t) \tag{13}$$

leading the model to a form like:

$$\begin{gathered}
\int_0^L \rho x\ddot{z} dx + J_c\ddot{\theta} = 0 \\
\rho\ddot{z} + EI_v \frac{\partial^4 e}{\partial x^4} dx = 0
\end{gathered} \tag{14}$$

with boundary conditions:

$$z\Big|_{x=0} = 0 \frac{\partial^2 z}{\partial x^2}\Big|_{x=l} = 0 \tag{15}$$

$$\frac{\partial z}{\partial x}\Big|_{x=0} = 0 \frac{\partial^3 z}{\partial x^3}\Big|_{x=l} = 0$$

All these transformations have been necessary to express the system in coordinates where an expansion on independent orthogonal functions of time and space can be performed:

$$z(x,t) = \sum_{r=1}^{\infty} \phi_r(x)\eta_r(t) \tag{16}$$

This allows separated spatial and time descriptions through the equations:

$$\ddot{\eta}_r(t) + \omega_r^2 \eta_r(t) = 0 \tag{17}$$

$$EI_v \frac{\partial^4 \phi_r(x)}{\partial x^4} - \omega_r^2 \rho \phi_r(x) = 0$$

For the eigenvector equation, 17, admissible solutions are proposed as:

$$\begin{aligned} \phi_r(x) &= a_r sin(\beta_r x) + b_r cos(\beta_r x) \\ &+ c_r sinh(\beta_r x) + d_r cosh(\beta_r x) \end{aligned} \tag{18}$$

The determination of the coefficients is made by solving a linear system. After many algebraic manipulations one can arrive at a standard form:

$$\delta_{rs} = \int_0^L \frac{d^2\phi_s}{dx^2}\frac{d^2\phi_r}{dx^2} dx \tag{19}$$

$$= \frac{1}{EI_v}\left\{\int_0^L \omega_r^2 \rho \phi_s \phi_r dx - J_c \omega_r [\phi_s' \phi_r']_{x=0}\right\}$$

Considering the forced system, one can find:

$$\sum_{r=1}^{\infty}\left[\int_0^L \rho\phi_s\phi_r dx - J_c[\phi_s'\phi_r']_{x=0}\right]\ddot{\eta}_r + \tag{20}$$

$$+EI_v \sum_{r=1}^{\infty}\left[\omega_r^2[\int_0^L \rho\phi_s\phi_r dx - J_c[\phi_s'\phi_r']_{x=0}\right]\eta_r = \tau\phi_s'\Big|_{x=0}$$

Observing the adopted norm, we finally have:

$$\ddot{\eta}_r + \omega_r^2\eta_r = \tau\phi_r'|_{x=0} \tag{21}$$

Then, the interaction between substructures becomes:

$$\ddot{\eta}_r + \omega_r^2\eta_r = \int_0^L F_E(l_i)\phi_r dx + \tau\phi_s'|_{x=0} \tag{22}$$

where $F_E(l_i)$ is the force due to the presence of a sliding mass at this point.

After some manipulations:

$$\begin{aligned} \ddot{\eta}_r + \omega_r^2\eta_r = &-\int_0^L M_i \sum_{s=1}^{\infty}\left[\phi_s\big|_{x=l_i}\ddot{\eta}_s\right]\phi_r dx \\ &-\int_0^L M_i 2\dot{l}_i\dot{\theta}\phi_r dx + \tau\phi_s'\Big|_{x=0} \end{aligned} \tag{23}$$

This equation is very important because it shows the relationship between the vibration modes in time and the influence of motions in space, that is, the interaction between the sliding masses, applied at the point l_i, and the structure and their effects transferred as functions of spatial forms. It also shows the effect of the torque applied to the hub on the flexible robotic arm.

4 STATE SPACE FOR THE DYNAMICAL SYSTEM

In order to design any control strategy, the system model should be written in state space variables. Then, substructure motions are synchronized and normalized to get a set of matrix equations in the form:

$$\ddot{\eta}_r = -[T]^{-1}\left(W\eta_r - 2M_i\dot{l}_i\dot{\theta}\left[\int_0^L \phi_r dx\right] + \tau\dot{\phi}_r|_{x=0}\right) \tag{24}$$

where $[W] = \omega_i^2 \times I_{nxn}$ and $[S] = [T]^{-1}$ is given by:

$$S_{rs} = \begin{cases} \dfrac{M_i + M_i^2 \sum\limits_{k=1,k\neq r}^{p} \phi_k\Big|_{x=l_i} \int\limits_0^L \phi_k dx}{1 + \sum\limits_{k=1}^{p} M_i\phi_k\Big|_{x=l_i} \int\limits_0^L \phi_k dx} & \text{for} \quad r = s \\[2ex] \dfrac{-M_i^2\phi_r\Big|_{x=l_i} \int\limits_0^L \phi_s dx}{1 + \sum\limits_{k=1}^{p} M_i\phi_k\Big|_{x=l_i} \int\limits_0^L \phi_k dx} & \text{for} \quad r \neq s \end{cases} \tag{25}$$

Using Newton's Law on the rotational motion, one can get:

$$\left(J_B + J_C + M_L l_l^2 \ddot{\theta}\right) = \tau + F_{Ej}(l_i)l_i \tag{26}$$

Equation 26 describes the rotational motion, includes the torque applied to the hub and the slider reactions F_{Ej}, at positions l_i and takes into account the beam moment of inertia, J_B. Finally, a state model can be written:

$$
\begin{aligned}
\dot{x}_{2r-1} &= x_{2r} \\
\dot{x}_{2r} &= -\frac{2\left(M_1 x_{2p+4}+M_2 x_{2p+6}\right) x_{2p+2} \sum\limits_{s=1}^{p} S_{rs} \int\limits_{0}^{L} \phi_s dx}{M_1+M_2} \\
&\quad -\frac{\sum\limits_{s=1}^{p} w_s^2 S_{rs} x_{2s-1}}{M_1+M_2}+\frac{\sum\limits_{s=1}^{p} S_{rs} \dot{\phi}_s \Big|_{x=0}}{M_1+M_2} u_i \\
\dot{x}_{2p+1} &= x_{2p+2} \\
\dot{x}_{2p+2} &= \frac{\sum\limits_{r=1}^{p} \sum\limits_{s=1}^{p} \omega_s^2 \phi_r(l_i) S_{rs} x_{2s-1}}{J_B+J_C+M_1 x_{2p+3}^2+M_2 x_{2p+5}^2}+ \\
&\quad \frac{1-\sum\limits_{r=1}^{p} \sum\limits_{s=1}^{p} S_{rs} \dot{\phi} \Big|_{x=0} \phi_r(l_i)}{J_{BC+M_1 x_{2p+3}^2+M_2 x_{2p+5}^2}} \tau \\
&\quad +2 \frac{\left(M_1 x_{2p+4}+M_2 x_{2p+6}\right) x_{2p+2}}{J_B+J_C+M_2 x_{2p+5}^2+M_1 x_{2p+3}^2} \times \\
&\quad \frac{\left(\sum\limits_{r=1}^{p} \sum\limits_{s=1}^{p} \phi_r(l_i) S_{rs} \int\limits_{0}^{L} \phi_s dx-1\right)}{J_B+J_C+M_2 x_{2p+5}^2+M_1 x_{2p+3}^2} \\
\dot{x}_{2p+3} &= x_{2p+4} \\
\dot{x}_{2p+4} &= \frac{u_2}{M_1}+x_{2p+3} x_{2p+2}^2 \\
\dot{x}_{2p+5} &= x_{2p+6} \\
\dot{x}_{2p+6} &= \frac{u_3}{M_1}+x_{2p+5} x_{2p+2}^2
\end{aligned}
\tag{27}
$$

The elements of matrix S are written according to eq. 25 with the necessary adaptations, p indicates the number of vibration modes of the system.

In this model, the state variables represent $\dot{x}_{2r}$ are the normal modes of the system, x_{2r-1} are the velocities of normal modes, x_{2p+1} is the angular motion, x_{2p+2} is the velocity angular motion, and x_{2p+3} and x_{2p+5} are the position of sliders 1 and 2 respectively, and x_{2p+4} and x_{2p+5} are the velocities of sliders.

5 THE OPTIMAL CONTROL PROBLEM

A general Optimal Control Problem (OCP) can be stated as one where the control laws $u_j(t)$, $j = 1, \cdots m$ and the initial and final conditions $x_i(a)$and $x_i(b)$, $i = 1, \cdots, n$ have to be chosen so as to minimize an Index of Performance

$$IP = \Phi(x(a), x(b)) + \int_a^b L(x(t), u(t), t)dt \tag{28}$$

subject to:

1. Dynamic Constraints:

$$\dot{x}_i = f_u(x(t), u(t), t), i = 1, \cdots, n \tag{29}$$

2. Boundary Constraints:

$$\phi_k(x(a), x(b), a, b) = 0, k = 1, \cdots, r \tag{30}$$

Although easily included, this system does not require the use of Control or State Inequality Constraints.

6 ILLUSTRATIVE CASES STUDY

The Indexes of Performance were always chosen as combinations of arm tip displacements and velocities since we are interested in minimizing tip vibrations. For robotic problems, minimum time or minimum control energy are not as relevant as tip vibrations because in the case the rotating arm is carrying a tool or other device, getting the final position with the tool ready to use is the most important figure. Manoeuver time may be incorporated to the Index of Performance (1 or 2 or 3 seconds, or more) and the control efforts to move the sliders are low. The structural problem under investigation corresponds to a fast angular manouever of the arm, from 0 to 45^o in 3s. A flexible arm will vibrate with large amplitudes if the system is not controlled and induced vibrations should be attenuated. Then, in our investigations, the objective is to reduce arm tip displacement and velocity to the smallest levels at the end of the angular manoeuver to guarantee a quick start. To do this, the Index of Performance is a combination of the weighted squares of tip displacement and velocity in the fundamental mode of vibration.

In the next figures the main dynamical arm parameters are addressed and two different IPs are simulated. In each case, the computational scheme performed successfully (assured numerical convergence) and 8 simulation results, considering different slider initial positions, are presented to facilitate comparisons, that is, one slider starts motion from a fixed arbitrary point, while the other has its initial position changed each time a new run is initiated.

Table 1shows the main parameters of the flexible arm.

Table 1: Physical parameters of the Optimal Control Problem

Physical Parameters	
Length of the arm	$L = 0,7m$
Arm Thickness	h = 0,001 m
Arm Width	b=0.0254
Arm Mass	$0,0482$ kg
Arm Linear mass density	$\rho_0 = m/L$
Arm Moment of inertia	$J_v = \rho_0 * \frac{L^3}{3}$
Hub Moment of inertia	$J_c = 1.3510^{-4}$
Aluminum Density	$\rho_B = 27 \cdot 10 kg/m^3$ (Al)
Aluminum Young Modulus	$E = 7.1 \cdot 10^{10}$ PA
Slider 1 mass	$M_1 = 0.05 * m$
Slider 2 mass	$M_2 = 0.05 * m$

In the first case, the sliding masses are arranged as follows: sliding mass 1 is initially put in position $l_1 = 0.3556$, and sliding body 2 has initial positions

$l_2 = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$ for the 8 simulations. The performance index is $IP = \int_0^3 (\mathbf{4x_1^2 + x_2^2})\mathbf{dt}$ All computer simulations were run using the RIOTS (Recursive Integration Optimal Trajectory Solver) 95 package by (Schwartz, 1966). This software solves the OCP allowing the use of 1-st, 2-nd, 3-rd and 4-th order fixed step-size Runge-Kutta integrator and 1-st to 4-th order splines. The optimization problem is solved with a class of conjugate-gradient techniques or with an SQP (Sequential Quadratic Programming) solver. User-defined cost and constraint functions, as well as their symbolic derivatives, are written in C code and dynamically linked to RIOTS. For the second case study, a small modification in the IP is made $IP = \int_0^3 (\mathbf{2x_1^2 + x_2^2})\mathbf{dt}$ Through a relatively heavier weight on the variable x_2, one intends to diminish the tip velocity during the flexible arm rotation.

The variables x_1 represent the vibration of the flexible arm and x_2 its derivative. Thus the criterion established intended to minimize the vibration during rotation. The performance criteria are presented with the intention of showing the difficulties in the numerical solution. Comparing the position of the sliding masses, figures 4 and 9, and the external action (forces and torque) applied, figures 6 and 11, to the flexible arm makes it clear the influence of the parameters in performance criterion to obtain the solution.

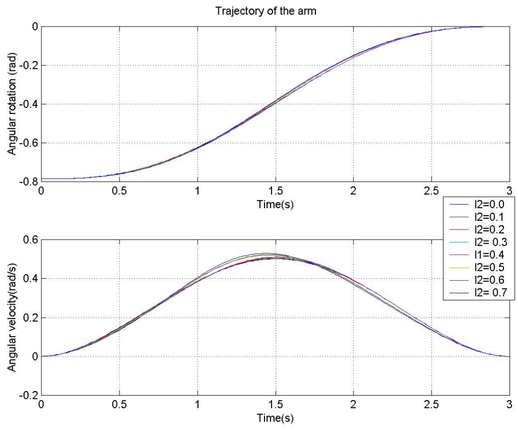


Figure 2: Flexible Arm Motion

Figures 2 and 7 show smooth movements of the rotating structure, as can be observed from the angular displacements, despite high angular velocities, since the structures start from null velocities, accelerate and ceases motion in just 3 seconds. Arm vibrations are quite small and movements are almost rigid ones.

In Figures 3 and 8, the tip vibrations on the end of the flexible arm are shown. Amplitudes are quite small when one considers the large flexibility of the structure.

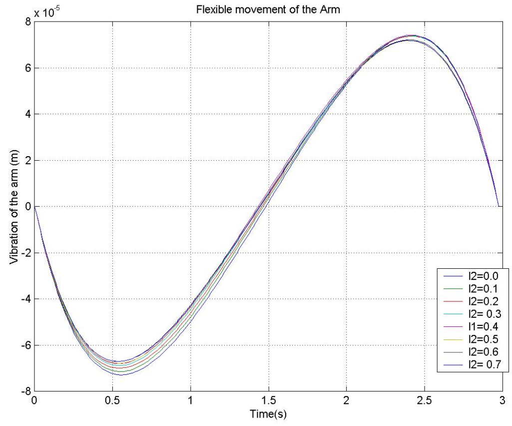


Figure 3: Vibration of the end of the flexible arm

In the cases shown, the smoothness of the curve is determined by the IP system.

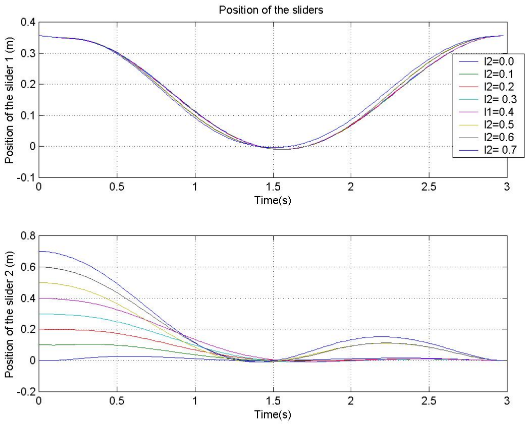


Figure 4: Trajectories of the sliders

Figures 4 and 9 exhibits the trajectories of the two sliders during the manoeuver of the flexible arm. One may observe in these simulations, a movement throughout the flexible arm on the first half of the time interval for the sliding mass m_2. For the mass m_1 , movements are virtually the same in all the simulations, noting that this slider moves toward the inner end of the flexible arm on the first half of the time and returns to its original position in the second half.

In Figures 5 and 10, the velocities of the two sliders are presented. Note that in both cases at the end of movement (T = 3s) their speeds are null, a condition that was deliberately imposed in the simulation scheme.

In Figures 6 and 11, the external forces acting on the flexible arm, due to the mass M_1 and M_2 and the external torque on the hub are shown. Forces to move the sliders are small and the torque is as required to perform the rigid motion.

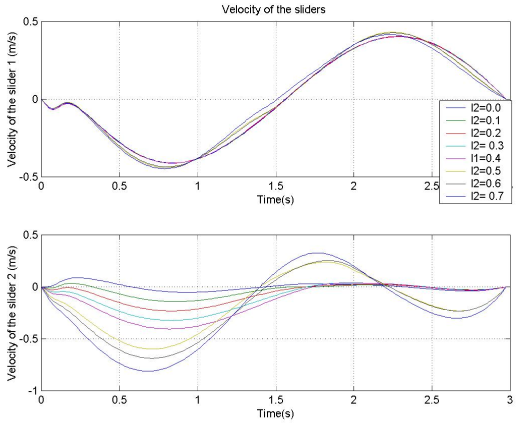


Figure 5: Velocity of the sliders

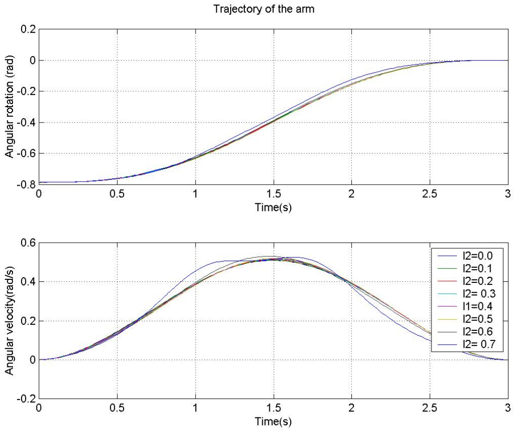


Figure 7: Flexible Arm Motion

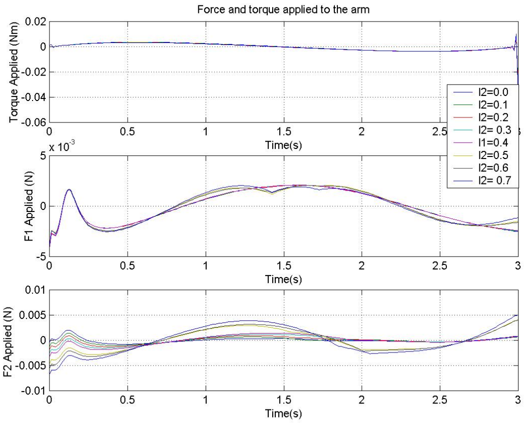


Figure 6: External Forces

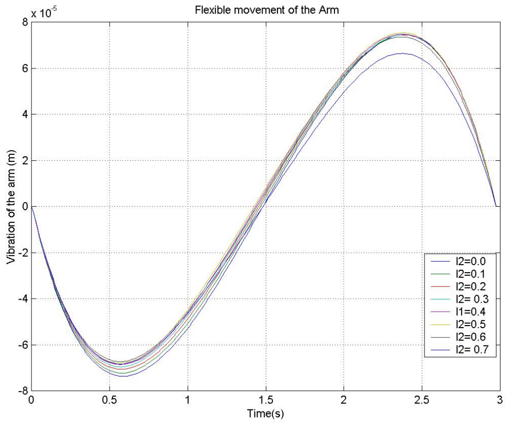


Figure 8: Vibration of the end of the flexible arm

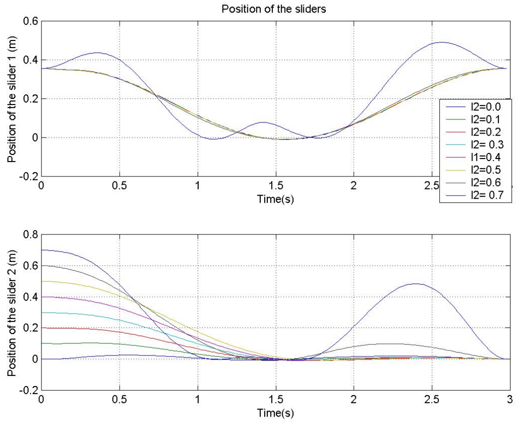


Figure 9: Trajectories of the sliders

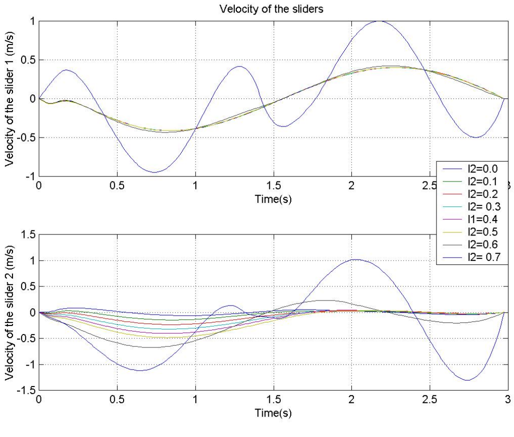


Figure 10: Velocity of the sliders

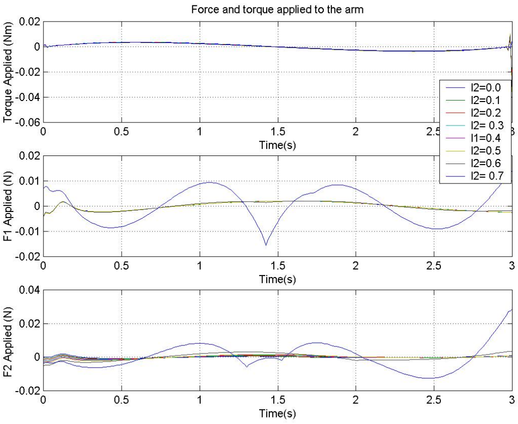


Figure 11: External Forces

7 FINAL COMMENTS

In this paper, we have presented a new way to reduce the vibrations induced on a flexible structure rotating around some axis. Vibration reduction is achieved through the translational motion of mass sliders. The full model of a narrow beam carrying n sliders is deduced and, in the sequence, the equations of motion are simplified by substruturing the system to make it feasible for control design. The non linear nature of the complex problem demands for Optimal Control approaches to find the trajectories the sliders shall perform in order to achieve the best vibration figures. Two slightly different cases considering two sliders have been proposed and simulated to demonstrate the feasibility of the proposed scheme. When comparing the two examples, a small change in the index of performance (other figures have been kept the same) leads to significant change in only one aspect, relative to the region where one of the sliding masses moves, the outer or the inner edge of the flexible arm, leading to the conclusion that the system behaves accordingly, despite the complexity of the motion of each part of the system. The authors intend to implement these and other already published results in an experimental device, already designed but not yet assembled, to confirm the performance of the proposed approaches in a near future.

ACKNOWLEDGEMENTS

Authors gratefully acknowledge CNPq (Brazilian National Research Council) and CAPES (Brazilian Ministry of Education) for finantial support of part of this project.

REFERENCES

Fleury, A. T. and Oliveira, F. R. F. (2004). *Dynamics and Control of a Flexible Rotating Arm Through the Movement of a Sliding Mass.*, pages p.299–318. Dynamical Systems and Control. Chapman&Hall/CRC, USA, (Stability and Control: v.22),. Udwadia, F.E.; Weber, H.I., Leitmann, G.,.

Junkins, J. and Kim, Y. (1993). *Introduction to Dynamics and Control of Flexible Structures*. American Institute of Aeronautics and Astronautics.

Meirotvitch, L. (1980). *Computacional Methods in Structural Dynamics*. Sythoff & Noordhoff.

Meirotvitch, L. (1990). *Dynamics and Control of Structures*. John Wiley & Sons.

Meirovitch, L. and Kwak, M. (1991). Rayleigh-ritz based substructure synthesis of flexible multibody systems. *AIAA Journal*, 29(10):1709–1719.

Oliveira, F. (2000). Controle de vibrações de um braço rotativo flexível pelo movimento de uma massa sobre o braço. Master's thesis, Escola Politécnica da Universidade de São Paulo.

Schwartz, A. (1966). *Theory and Implementation of Numerical Methods Based on Runge-Kutta Integration for Solving Optimal Control Problems*. PhD thesis, U.C. Berkeley.

Terceiro, E. (2002). Use of slider masses to the optimal control of flexible rotating arm vibrations: General case, simulation and results. (in portuguese). Master's thesis, University of São Paulo.

Terceiro, E. and Fleury, A. T. (2008). Optimal vibration control of a rotating beam using moving masses. *EngOpt 2008 - International Conference on Engineering Optimization*, v. 1. Rio de Janeiro, RJ, Brasil : UFRJ.

BIOGRAPHY

Agenor de Toledo Fleury holds a BSc in Mechanical Engineering (ITA/Technological institute of Aeronautics, 1973) and MSc and PhD in Mechanical Engineering (University of São Paulo, 1978 and 1985). He is a Full Professor at FEI University and a Part Time Lecturer at Escola Politecnica, University of São Paulo. His main interests nowadays are in Rehabilitation Robotics, Flexible Structures Control, Optimal Appied Control and Nonlinear Kalman Filter Applications.

Eder Terceiro holds MSc in Mechanical Engineering (University of São Paulo, 2002). He is currently a doctoral student in mechanical engineering at the Escola Politécnica, University of São Paulo. In the current program, his main interest is Flexible Structures Control, Optimal Appied Control.

ON THE PARTIAL GUIDANCE OF AN AUTONOMOUS BLIMP BY VARIOUS FEEDBACK LAWS

Chaker Jammazi[a], **Adel Khadaraoui** [b], **Maâli Zaghdoudi**[c]

[a] Faculté des Sciences de Bizerte. Département de Mathématiques
[b,c] Ecole Polytechnique de Tunisie
Laboratoire d'Ingénierie Mathématique

[a]chaker.jammazi@ept.rnu.tn, [b]adelk989@gmail.com, [c]maali.zaghdoudi@gmail.com

ABSTRACT

In this paper, the problem of partial asymptotic stabilization of the nonlinear autonomous under actuated airship (AUA) by various feedback laws is investigated. It has been shown that the AUA's is not stabilizable via continuous pure-state feedback. This is due to (Brockett 1983), necessary condition. In order to cope with this difficulty, we propose in asymptotically eleven components in finite-time or exponentially, while the remaining one converges.

Keywords: Airship, attitude control, discontinuous controllers, finite-time partial stabilizability.

1. INTRODUCTION

Control problems of aerospace engineering have recently drawn considerable attention in the control community. The rigid spacecraft, the rigid aircraft and the airship are examples of such systems. These systems are presented in cascade structure and have fewer actuators than the system degree of freedom.

For this reason, the tools from linear control theory are not sufficient, and stabilization techniques need to be reconsidered. Indeed, it has been proved by (Sontag and Sussmann 1980), that all nonlinear controlled systems in dimension one cannot be stabilized by continuous feedback laws. As a solution for this problem, the authors proposed piecewise continuous feedback laws. This obstruction to stabilizability in dimension one is generalized for nonlinear control systems by a number of authors. The first one was given by (Brockett 1983), for all controllable systems and (Ryan 1994), only for continuous control systems and (Coron and Rosier 1994), for the stabilizability of systems with drift. (Ryan 1994), proved that Brockett's condition is still necessary for stabilizability by discontinuous feedback laws in Ryan's sense. Also, (Coron and Rosier 1994), proved that Brockett's condition is still necessary for the stabilizability of systems with drift by means of discontinuous feedback laws and the solutions are defined in Filippov's sense. A strong homology necessary condition for stabilizability by dynamic feedback laws was given by (Coron 1990).

An article traced by (Samson 1991), has proved that continuous time-varying feedback laws can be interesting to stabilize many systems which cannot be stabilized by continuous pure-state feedbacks. This has been demonstrated by Coron's results in the famous paper (Coron 1995), which established that most STLC -small-time locally controllable- systems can be stabilized in finite time by continuous time-varying feedback. The obtained result leads to rich research in this area, namely:

- Time-varying periodic controllers: (Beji, Abichou, and Bestaoui 2004; Coron 1992; Coron 1995; Coron 2007; Coron and d'Andréa Novel 1992; Coron and Keraï 1996; M'Closkey and Murray 1997; Morin 2004; Morin and Samson 1997; Morin, Samson, Pomet, Ping, and Jiang 1995; Pettersen and Egeland 1996; Pettersen and Egeland 1999 ; Pettersen and Nijmeijer 2001; Samson 1991; Samson 1995), and the references therein,
- discontinuous controllers: (Astolfi 1996a; Astolfi 1996b; Coron and Rosier 1994; Sontag and Sussmann 1980; Sussmann 1979),
- the partial asymptotic or finite-time stabilization by continuous or discontinuous feedback laws: (Jammazi 2008a; Jammazi 2008b; Jammazi 2010; Jammazi 2011)

In this paper, we will focus our attention on the third approach. It consists of the concept of the Partial Asymptotic Stabilization (PAS). This concept means the asymptotic stabilization with respect to the maximum components of the system; while the remaining components are convergent and not necessarily toward an equilibrium point.

In (Jammazi 2008b), we have developed the backstepping techniques for the partial asymptotic stabilizability. This result was used to solve the partial asymptotic stabilization of many controllable cascaded systems that do not satisfy Brockett's necessary condition. Differentiable stabilizing feedback laws for the rigid spacecraft and for the ship are derived. For both systems, these stabilizing feedbacks make five components asymptotically stable and one component converges; in particular we have improved the (Zuyev 2001), feedbacks for the rigid spacecraft which states that the angular velocity of the third axes is only bounded, and for the ship system we have improved (Wichlund, Sørdalen, and Egeland 1995), feedbacks which states that the yaw angle is only bounded. Moreover, in (Jammazi 2010), we have provided a rigorous formulation of the theory of asymptotic partial stability, respectively, the finite-time partial stability of continuous autonomous systems. Sufficient conditions are derived with applications in control design. For example, we have proved in (Jammazi 2010) that the partial stabilization of the ship can be achieved in finite-time by continuous or discontinuous bounded state feedback laws.

In (Jammazi 2011), we have studied the finite-time partial stability of a prototype system of nonholonomic control systems which is the benchmark knife edge or the unicycle robot system called also the Brockett's integrator. We have proposed various feedback controllers that achieve the partial asymptotic stabilizability, or the finite-time partial stability of the mobile robot. These feedbacks are Hölderian for the rational partial stability, continuous and homogeneous of negative degree or discontinuous and quasi-homogeneous of negative degree for the finite-time partial stability.

The airship is the subject of numerous papers and thesis; (Hygounenc 2001), (Hygounenc 2003), (Zhang and Ostrowski 1999), (Beji and Abichou 2005; Beji, Abichou, and Bestaoui 2004), (Bestaoui 2006) and (Samaali, Abichou, and Beji 2007), and references therein.

As cited in (Samaali, Abichou, and Beji 2004), the problem of adding physical parameters of the blimp into the image plane for the performance of vision-guided control is discussed in (Zhang and Ostrowski 1999).

In (Beji and Abichou 2005), the problem of tracking control for ascent and descent flight with only three controllers is addressed. The authors supposed that roll is totally unactuated.

In (Bestaoui 2006), the problem of generation of characterization nominal trajectories (flight path) to be followed by an autonomous airship is addressed. In (Samaali, Abichou, and Beji 2007), the authors have studied the stabilization with respect to longitudinal and horizontal planes. By using iterative backstepping techniques combined with Lyapunov theory, homogeneity and averaging theorems, the authors have shown that the stabilization is possible via continuous time-varying feedback laws.

In this paper, our objective is to solve the stabilizing control problem of attitude and position for underactuated airship using only three available controls: the main and tail thrusters and the tilt angle of the propellers. The roll is totally unactuated. The same input controls both pitch and surge, while yaw and sway are related.

It was shown in (Beji, Abichou, and Bestaoui 2004), that the stabilization problem of autonomous airship by regular state feedback laws in the usual sense is not possible. As a solution of this problem, the authors have proposed time-varying feedback laws. The proposed method uses the averaging method and homogeneous exponential stability developed in (Morin and Samson 1997).

Note that all papers (Beji and Abichou 2005; Beji, Abichou, and Bestaoui 2004; Bennaceur 2009; Bestaoui 2006; Samaali, Abichou, and Beji 2007) cited here have treated the LSC'AS200 airship (Figure 1).

In fact, introducing the time in these feedback laws produces "undesirable" oscillations of the system around his equilibrium point, (Beji, Abichou, and Bestaoui 2004; Coron 1992 ; M'Closkey and Murray 1997; Morin and Samson 1997; Morin, Samson, Pomet, and Ping Jiang 1995; Pettersen and Egeland 1996; Samson 1991; Samson 1980), for more general systems. To get around the problem of impossibility to stabilize the autonomous airship by pure and regular feedback laws, and to overcome the drawback of the time dependence of these feedback laws, the stabilization of the airship should be solved via static feedbacks in partial asymptotic stabilizability sense.

The obtained results show that we can ensure the asymptotic stabilizability of eleven states of variables, and convergence of the remaining one. In the first approach, by using the backstepping techniques and partial asymptotic stabilizability developed in (Jammazi 2008a), we have shown that the LSC'AS-200 blimp can be stabilized partially exponentially by linear feedbacks. In the second approach, we have proved that the blimp can be stabilized partially in finite-time by means of continuous state feedback laws. However, the airship is an example of system with drift in which the (Coron and Rosier 1994), condition fails to be stabilized by discontinuous feedback laws. For this reason, to get around this obstruction, we have developed discontinuous state feedback laws that make the blimp stable in finite-time with respect to six components (which are the position $(x, y, z)'$ of the blimp in the inertial frame, and the linear velocities $(\mathrm{u}, \mathrm{v}, \mathrm{w})'$ in surge, sway and heave decomposed in the body-fixed frame), this leads by linearization to exponential stability of five components (which are $(p, q, r)'$: angular velocities in roll, pitch and yaw decomposed in the body-fixed frame, and (e_1, e_2) the orientation of principal axis (n_x, n_y)) and therefore the convergence of the orientation angle e_3 with respect to axis n_z. The stabilization by discontinuous feedback law appears significant, despite the presence of chattering phenomenon (Orlov 2009).

This paper is structured in this way: The section 2 contains some mathematical preliminaries. The stabilization strategies of the model of airship by various state feedback laws are the subject of Section 3. The theoretical results are confirmed by simulations in Section 4 and the conclusion is given in Section 5.

Throughout the paper, $|.|$ denotes the Euclidean norm in $\mathbb{R}^n$, $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^{n\times m}$ defined by, $\|A\| = \sup_{i,j} |a_{ij}|$ for $A = a_{ij}, 1 \le i \le n, 1 \le j \le m$, $'$ is the symbol of transposition and sgn is the function "sign" and $A \cong B$ means A is diffeomorph to B.

2. PRELIMENAIRES

The double integrator is a key system that can appears in all underactuated dynamical systems. For this reason, the stabilization of such system is an interesting area of many works (Bhat and Bernstein 1998; Hong, Yang, Cheng, and Spurgeon 2004; Huang, Lin, and Yang 2005; Orlov 2005; Orlov 2009). In this section, we begin with review some results concerning the stabilization in finite-time of the double integrator

$$\begin{cases} \dot{x} = y, \\ \dot{y} = u. \end{cases} \qquad (1)$$

The system (1) can be stabilized by two classes of feedbacks: continuous and discontinuous state static feedback laws which are presented in the following lemmas.

Lemma 1: (Bhat and Bernstein 1997*) The system (1) is finite-time stabilizable under the continuous feedback*

$$u(x,y) = -sgn(y)|y|^{\alpha} - sgn(x)|x|^{\frac{\alpha}{2-\alpha}}, \alpha \in (0,1).$$

Lemma 2: *(Orlov 2009) the system (1) is finite-time stabilizable under the discontinuous feedback*

$$u(x,y) = -a\, sgn(x) - b\, sgn(y), a > b > 0.$$

3. PARTIAL ATTITUDE CONTROL OF AUTONOMOUSUNDERACTUATED AIRSHIP SYSTEM

This section is devoted to studying the complete system of underactuated airship which is the AS-200 by Airspeed Airships, see (Beji and Abichou 2005; Bestaoui 2006), for more details. It was shown in (Beji, Abichou, and Bestaoui 2004) that no continuous time-invarying feedback law which makes the origin of the airship asymptotically stable exists, because the latter system does not satisfy Brockett's condition (Brockett 1983). In order to overcome the Brockett's obstruction, the stabilization of the airship is treated in partial asymptotic stabilization sense.

Equation of motion: The autonomous underactuated airship is a complex nonlinear system described by twelve variables of state and three controls. The model was found in (Beji and Abichou 2005; Beji, Abichou, and Bestaoui 2004):

$$\begin{cases}
\dot{u} = \frac{1}{m_{11}}(X_u u - 2(B - mg)(e_1 e_3 \\
\qquad - e_2\sqrt{1 - e_1^2 - e_2^2 - e_3^2}) \\
\qquad - m_{33} wq + m_{22} v r + \tau_1), \\
\dot{v} = \frac{1}{m_{22}}(Y_v v - 2(B - mg)(e_1 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} \\
\qquad + e_2 e_3) - m_{11} ur + m_{33} pw + \tau_2), \\
\dot{w} = \frac{1}{m_{33}}(Z_w w - (B - mg)(1 - 2(e_1^2 + e_2^2)) \\
\qquad + m_{11} uq - m_{22} vp + \tau_3), \\
\dot{p} = \frac{1}{\Delta}(- L_p I_{33} p + N_r I_{13} r \\
\qquad - 2Bz_b I_{33}\left(e_2 e_3 + e_1\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) \\
\qquad + (I_{33}^2 + I_{13}^2 - I_{33} I_{22})\, qr + I_{13}(I_{11} - I_{22} + I_{33})\, pq \\
\qquad + I_{13}(Y_v - X_u)\, uv + I_{33}(Y_v - Z_w)\, vw - P_2^1 I_{13} \tau_2), \\
\dot{q} = \frac{1}{I_{22}}(M_q q - 2Bz_b(e_2 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} \\
\qquad - e_1 e_3) + (X_u - Z_w)\, uw + I_{13}(p^2 - r^2) \\
\qquad + (I_{11} - I_{33})pr + P_1^3 \tau_1), \\
\dot{r} = \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r \\
\qquad + 2Bz_b I_{13}\left(e_1\sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2 e_3\right) + (-I_{13}^2 \\
\qquad - I_{11}^2 + I_{11} I_{22})\, pq + I_{13}(-I_{11} + I_{22} - I_{33})qr \\
\qquad + I_{11}(X_u - Y_v)\, uv + I_{13}(Z_w - Y_v)\, vw + P_2^1 I_{11} \tau_2), \\
\dot{x} = (1 - 2(e_2^2 + e_3^2))\, u \\
\qquad + 2\left(e_1 e_2 - e_3\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) v \\
\qquad + 2\left(e_1 e_3 + e_2\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) w, \\
\dot{y} = 2\left(e_1 e_2 + e_3\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) u \\
\qquad + (1 - 2(e_1^2 + e_3^2))v \\
\qquad + 2\left(e_2 e_3 - e_1\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) w, \\
\dot{z} = 2\left(e_1 e_3 - e_2\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) u + \\
\qquad 2\left(e_2 e_3 + e_1\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right) v \\
\qquad + (1 - 2(e_1^2 + e_2^2)) w, \\
\dot{e}_1 = \frac{1}{2}\left(p\sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2 r - e_3 q\right), \\
\dot{e}_2 = \frac{1}{2}\left(e_3 p + q\sqrt{1 - e_1^2 - e_2^2 - e_3^2} - e_1 r\right), \\
\dot{e}_3 = \frac{1}{2}\left(-e_2 p + e_1 q + r\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right).
\end{cases} \qquad (2)$$

The constants m_{ij} and I_{ij} are the coefficients of the inertia matrix M supposed to be symmetric and positive definite.

The constants X_u, Y_v, Z_w, L_p, N_r *and* M_q are the aerodynamic coefficients. The vector $\nu := (u, v, w, p, q, r)'$ denotes the linear velocities in surge, sway and heave, and the angular velocities in roll, pitch and yaw, decomposed in the body-fixed frame.

Define the vector $\eta := (\eta_1, \eta_2)$ where $\eta_1 := (x, y, z)'$ is the position of the airship in the inertial frame. The vector $\eta_2 := (e_0, e_1, e_2, e_3)'$ is defined as follows: the vector $(e_0, e_1, e_2, e_3)'$ defined the unit quaternion (i.e. $e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1$), the component e_0 is supposed non negative and given by $e_0 := \sqrt{1 - e_1^2 - e_2^2 - e_3^2}$. The vector $\tau = (\tau_1, \tau_2, \tau_3)'$ denotes the control forces decomposed in the body-fixed frame.

As in (Beji and Abichou 2005), the lighter than air platform used in this paper is the AS-200 by Airspeed Airships (Figure 1), for more description of this type of blimp the reader is referred to (Bestaoui 2006). The blimp's parameters are as follows in the International System Units:

- blimp's total mass: $m = 9.07$, the nacelle mass m_n =1.58,
- added masses $X_x = 1.13, Y_y = 7.25, Z_z = 7.25, K_x = 0, M_y = 8.87, N_z = 8.87$,
- inertial parameters around the principal axes of inertia: $I_x x = 2.19, I_y y = 18.85, I_z z = 18.76$ and $I_x z = 0$,
- inertial terms $I_{11} = I_{xx} + K_x = 2.19, I_{22} = I_{yy} + M_y = 27.73$, $I_{33} = I_{zz} + N_z$ =27.63, $I_{13} = -I_{xz} + K_x = 0.22$,
- Δ term: $\Delta = I_{13}^2 - I_{11} I_{22} = -60.89$,
- positions of input forces F_1 and F_2: $P_1^3 = -1$ and P_2^1=-3,
- aerodynamic coefficients: $X_u = Y_v = Z_w = L_p = N_r = M_q$ =-10
- buoyancy and gravity magnitudes: mg =89 and $B = 72.2$ and $z_b = -0.1$.

Figure 1: The AS-200 Airship (Bestaoui 2006).

Consider the function $\sigma: B(0,1) \sqsubset \mathbb{R}^3 \to \mathbb{R}$ defined by

$$\sigma(\eta_2) = \sigma(e_1, e_2, e_3) = \sqrt{1 - e_1^2 - e_2^2 - e_3^2} \qquad (3)$$

where $B(0,1)$ is the open ball of $\mathbb{R}^3$. A Taylor's expansion on the neighborhood B of $0_{\mathbb{R}^3}$ of σ gives

$$\sigma(\eta_2) = \sigma(0) + \sigma'(0)\eta_2 + g(\eta_2), \qquad (4)$$

where g is a smooth function on the open ball $B(0,1)$ satisfying $g(0) = 0$.

Since $\sigma(0) = 1$, and

$$\sigma'(\eta_2) = ((-e_i)((1 - e_1^2 - e_2^2 - e_3^2)^{-1/2}))_{1 \le i \le 3},$$

then $\sigma'(0) = 0_{\mathbb{R}^3}$

and we get $\sigma(\eta_2) = 1 + g(\eta_2)$. (5)

Thus, the system (2) can be transformed as follows:

$$\dot{\xi} = f(\xi, \tau) + g(\xi),$$

where $\xi \in \mathbb{R}^{12}$ is the state and $\tau \in \mathbb{R}^3$ the control. The function $f(\xi, \tau)$ contains the linear terms of the system (2) where g contains the nonlinear part of the rest of the system.

We begin by studying the system $\dot{\xi} = f(\xi, \tau)$ which is given by

$$\begin{cases} \dot{u} = (\frac{1}{m_{11}} X_u u + 2(B - mg)e_2 + \tau_1), \\ \dot{v} = \frac{1}{m_{22}}(Y_v v - 2(B - mg)e_1 + \tau_2), \\ \dot{w} = \frac{1}{m_{33}}(Z_w w - (B - mg) + \tau_3), \\ \dot{p} = \frac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2B z_b I_{33} e_1 - P_2^1 I_{13} \tau_2), \\ \dot{q} = \frac{1}{I_{22}}(M_q q - 2B z_b e_2 + P_1^3 \tau_1), \\ \dot{r} = \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2B z_b I_{13} e_1 + P_2^1 I_{11} \tau_2), \\ \dot{x} = u, \\ \dot{y} = v, \\ \dot{z} = w, \\ \dot{e}_1 = \frac{1}{2} p, \\ \dot{e}_2 = \frac{1}{2} q, \\ \dot{e}_3 = \frac{1}{2} r. \end{cases} \qquad (6)$$

3.1. Obstruction to stabilizability

We show that (6) cannot satisfy the Brockett's necessary condition for stabilizability.

Proposition 1: *There is no continuous state feedback that can stabilize asymptotically the system (6).*

Proof: Let τ' be the feedback transformation defined by

$$\tau'_3 = \tau_3 - (B - mg),$$

this means that the term $-(B - mg)$ is crushed by a component of control τ_3, and the airship is at position h above the ground. With the new input τ'_3, the system (6) becomes $\dot{\xi} = f(\xi,\tau) :=$

$$\begin{pmatrix} \frac{1}{m_{11}}(X_u u + 2(B-mg)e_2 + \tau_1) \\ \frac{1}{m_{22}}(Y_v v - 2(B-mg)e_1 + \tau_2) \\ \frac{1}{m_{33}}(Z_w w + \tau'_3) \\ \frac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2Bz_b I_{33} e_1 - \\ I_{13} P_2^1 \tau_2) \\ \frac{1}{I_{22}}(M_q q - 2Bz_b e_2 + P_1^3 \tau_1) \\ \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2Bz_b I_{13} e_1 + \\ P_2^1 I_{11} \tau_2) \\ u \\ v \\ w \\ \frac{1}{2}p \\ \frac{1}{2}q \\ \frac{1}{2}r \end{pmatrix}$$

All points M_ϵ in the form $(0,0,*,0,0,\varepsilon,0,0,0,0,0,0)'$ where $\varepsilon \neq 0$ are not in the image of f. Indeed, if it was the case, the equation $f(\xi,\tau) = M_\epsilon$ admits a solution. So, by combining the equation 2 and the equation 4 we get $2(B-mg)e_1 = \tau_2$ and $2Bz_b I_{33} e_1 = -P_2^1 I_{13} \tau_2$ then $2Bz_b I_{33} e_1 = -P_2^1 I_{13}(B-mg)e_1$ which implies $e_1 = \tau_2 = 0$; then $\varepsilon = 2Bz_b I_{13} e_1 + P_2^1 I_{11} \tau_2 = 0$.
We obtain a contradiction. ☐

3.2. First strategy: Partial exponential stabilizability

To get an adequate form of the system (6), we adopt the following transformation

$$u_1 := \frac{1}{m_{11}}(X_u u + 2(B-mg)e_2 + \tau_1),$$
$$u_2 := \frac{1}{m_{22}}(Y_v v - 2(B-mg)e_1 + \tau_2),$$
$$u_3 =: \frac{1}{m_{33}}(Z_w w - (B-mg) + \tau_3).$$

Then (6) is equivalent to

$$\begin{cases} \dot{u} = u_1, \\ \dot{v} = u_2, \\ \dot{w} = u_3, \\ \dot{p} = \frac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2(Bz_b I_{33} + P_2^1 I_{13}(B - \\ \quad mg))\, e_1 - P_2^1 I_{13}(m_{22}u_2 - X_v v)), \\ \dot{q} = \frac{1}{I_{22}}(M_q q - 2(Bz_b + P_1^3(B-mg))e_2 + \\ \quad P_1^3(m_{11}u_1 - X_u u)), \\ \dot{r} = \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2(Bz_b I_{13} + P_2^1 I_{11}(B - \\ \quad mg))e_1 + P_2^1 I_{11}(m_{22}u_2 - Y_v v)), \\ \dot{x} = u, \\ \dot{y} = v, \\ \dot{z} = w, \\ \dot{e}_1 = \frac{1}{2}p, \\ \dot{e}_2 = \frac{1}{2}q, \\ \dot{e}_3 = \frac{1}{2}r. \end{cases} \tag{7}$$

In the sequel, we will be interested in system (7) with respect to $\zeta = (u,v,w,p,q,r,x,y,z,e_1,e_2)'$. In order to apply the backstepping techniques in partial asymptotic stabilizability developed in (Jammazi 2008b), we start by studying the reduced system which is given by

$$\begin{cases} \dot{p} = \frac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2Bz_b I_{33} e_1 \\ \quad + P_2^1 I_{13}((Y_v - m_{22})u_2 - 2(B-mg)e_1)), \\ \dot{q} = \frac{1}{I_{22}}(M_q q - 2Bz_b e_2 + P_1^3((m_{11} - X_u)u_1 + \\ \quad -2(B-mg)e_2)), \\ \dot{r} = \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2Bz_b I_{13} e_1 + P_2^1 I_{11}((m_{22} - \\ \quad Y_v)u_2 + 2(B-mg)e_1)), \\ \dot{x} = u_1, \\ \dot{y} = u_2, \\ \dot{z} = u_3, \\ \dot{e}_1 = \frac{1}{2}p. \\ \dot{e}_2 = \frac{1}{2}q. \end{cases} \tag{8}$$

Stabilization of the system (2)

Proposition 2: *Let* $k_i (i = 1,2,3)$ *three nonnegative reel numbers. Then, with the action of the following feedbacks*

$$v_1 := -k(u + k_1 x),\ v_2 := -k(v + k_2 y),$$
$$v_3 := -k(w + k_3 z), \tag{9}$$

where k *is large enough, the system (2) is eleven locally partially exponentially stable. More precisely, the partial state* $\zeta = (u,v,w,p,q,r,x,y,z,e_1,e_2)' \in \mathbb{R}^{11}$ *is locally exponentially stable and* e_3 *converges.*

Proof: The proof of the proposition comes from ((Jammazi 2010), Corollary 7) which states that if the linearized system is p-partially exponentially stable then the initial system is p-locally exponentially partially stable. In closed loop, the linearization of (2) with respect to ζ around the equilibrium point is given by the system

$$\begin{cases} \dot{u} = -k(u + k_1 x), \\ \dot{v} = -k(v + k_2 y), \\ \dot{w} = -k(w + k_3 z), \\ \dot{p} = \frac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2(B z_b I_{33} + \\ \quad P_2^1 I_{13}(B - mg))e_1 - \\ \quad P_2^1 I_{13}(-k m_{22}(v + k_2 y) - Y_v v)), \\ \dot{q} = \frac{1}{I_{22}}(M_q q - 2(B z_b + P_1^3(B - mg))e_2, \\ \quad + P_1^3(-k m_{11}(u + k_1 x) - X_u u)), \\ \dot{r} = \frac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2(B z_b I_{13} \\ \quad + P_2^1 I_{11}(B - mg)e_1 \\ \quad + P_2^1 I_{11}(-k m_{22}(v + k_2 y) - Y_v v)), \\ \dot{x} = u, \\ \dot{y} = v, \\ \dot{z} = w, \\ \dot{e}_1 = \frac{1}{2} p, \\ \dot{e}_2 = \frac{1}{2} q. \end{cases} \tag{10}$$

Clearly the system (10) is exponentially stable with respect to $(u, v, w, x, y, z)'$. The linear system (10) with respect to $\delta = (p, q, r, e_1, e_2)' \in \mathbb{R}^5$ admits the following set of eigenvalues which are with negative real parts

$\lambda_1 = -0.7597, \lambda_2 = -3.6869 + 0.4501i,$
$\lambda_3 = -3.6869 - 0.4501i, \lambda_4 = -0.0970$ *and* $\lambda_5 = -3.5560.$

Straightforward computations show that (10) is exponentially stable with respect to. Clearly nonlinear part of (2) with respect to ζ vanish when the "uncontrolled part" e_3 is zero. Then, by using, ((Jammazi 2010), Corollary 7), the system (2) is locally exponentially stable with respect to ζ.

Consequently, there exists α > 0 and λ > 0 such that

$$|\zeta(t)| \le \alpha |\zeta(0)| e^{-\lambda t}, t > 0. \tag{11}$$

In particular, we get

$$|p(t)|, |q(t)|, |r(t)|, |e_1(t)|, |e_2(t)| \le \alpha |\zeta(0)| e^{-\lambda t}. \tag{12}$$

Since $\dot{e}_3 = \frac{1}{2}\left(-e_2 p - e_1 q + r\sqrt{(1 - e_1^2 - e_2^2 - e_3^2}\right)$, and $\sqrt{1 - e_1^2 - e_2^2 - e_3^2} \le 1$, then we get

$$|\dot{e}_3| \le \alpha^2 |\zeta(0)|^2 e^{-2\lambda t} + \frac{1}{2} \alpha |\zeta(0)| e^{-\lambda t}. \tag{13}$$

From (13) we easily deduce that $\dot{e}_3(t)$ is Lebesgue integrable and therefore e_3 converges. This completes the proof. □

3.3. Second alternative: Finite-time partial stabilizability

In this section, we give other strategies to stabilize the airship. This alternative is based on the theory of partial stabilization and on continuous feedback laws given in Lemma 1 (respectively, discontinuous feedback laws given in Lemma (2). We begin with the continuous finite time stabilizing feedback laws.

Proposition 3: *Let be* $\alpha \in (0,1)$, *then under the following feedback laws*

$$v_1 := -sgn(\dot{x})|\dot{x}|^\alpha - sgn(x)|x|^{\frac{\alpha}{2-\alpha}},$$

$$v_2 := -sgn(\dot{y})|\dot{y}|^\alpha - sgn(y)|y|^{\frac{\alpha}{2-\alpha}},$$

$$v_3 := -sgn(\dot{z})|\dot{z}|^\alpha - sgn(z)|z|^{\frac{\alpha}{2-\alpha}}, \tag{14}$$

the underactuated system (2) is finite-time stable with respect to $(x, u, y, v, z, w)'$ *and locally exponentially stable with respect to* $(p, q, r, e_1, e_2)'$ *which implies that* e_3 *converges.*

Proof: We consider the system (2) and taking the feedback transformation

$$\bar{\tau}_1 := \frac{1}{m_{11}}(X_u u - 2(B - mg)(e_1 e_3 - e_2\sqrt{1 - e_1^2 - e_2^2 - e_3^2}\ ) - m_{33} w q + m_{22} v r + \tau_1),$$

$$\bar{\tau}_2 := \frac{1}{m_{22}}(Y_v v - 2(B - mg)(e_1 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2 e_3) - m_{11} u r + m_{33} p w + \tau_2),$$

$$\bar{\tau}_3 := \frac{1}{m_{33}}(Z_w w - (B - mg)(1 - 2(e_1^2 + e_2^2)) + m_{11} u q - m_{22} v p + \tau_3), \tag{15}$$

then the dynamic of the states u, v and w become $\dot{u} = \bar{\tau}_1$, $\dot{v} = \bar{\tau}_2$, and $\dot{w} = \bar{\tau}_3$. Then by taking the time derivative respectively of ẋ, ẏ and ż we get

$$\ddot{x} = \left(1 - 2(e_2^2 + e_3^2)\right)\dot{u} + \{\text{nonlinear terms}\}$$

$$= (1 - 2(e_2^2 + e_3^2))\, \bar{\tau}_1 + \{\text{nonlinear terms}\}. \tag{16}$$

Let g_1 be the smooth function defined *by* $g_1(\eta_2) = 1 - 2(e_2^2 + e_3^2)$. Since $g_1(0) = 1 \neq 0$, then there exists a neighborhood V_0 of zero such that $g_1(\eta_2) \neq 0$ for all $\eta_2 \in V_0$. In this case, the system (16) is locally feedback equivalent to

$\ddot{x} = v_1$, where $v_1 = (1 - 2(e_2^2 + e_3^2))\, \bar{\tau}_1 +$ {nonlinear terms}. (17)

Since the functions $g_2(\eta_2) = 1 - 2(e_1^2 + e_3^2)$ and $g_3(\eta_2) = 1 - 2(e_1^2 + e_2^2)$ satisfies $g_2(0) = g_3(0) = 1 \neq 0$, by the same above argument, the dynamic of y and z are locally equivalent to

$$\ddot{y} = v_2 := (1 - 2(e_1^2 + e_3^2))\, \bar{\tau}_2 + \{\text{nonlinear terms}\}.$$

$$\ddot{z} = v_3 := (1 - 2(e_1^2 + e_2^2))\,\bar{\tau}_3 \quad + \quad \{\text{nonlinear terms}\}. \tag{18}$$

Here, these nonlinear terms vanish in the equilibrium point. To summarize, the dynamic equation of (x, y, z) is now in the following double integrator form:

$$\begin{aligned} \ddot{x} &= v_1, \\ \ddot{y} &= v_2 \\ \ddot{z} &= v_3. \end{aligned} \tag{19}$$

Then, according to Lemma 1, by choosing feedbacks given in (14), we get easily x, y and z are stable in finite time and $\dot{x}, \dot{y}$ and $\dot{z}$ are too, which give also u, v and w are stable in finite-time. Then there exists a settling time T such that $\forall\, t \geq T$,

$$x(t) = y(t) = z(t) = u(t) = v(t) = w(t) = 0.$$

In this case, the system (2) becomes

$$\begin{cases} \dot{p} = \dfrac{1}{\Delta}\big(-L_p I_{33} p + N_r I_{13} r + (I_{33}^2 + I_{13}^2 - I_{33} I_{22})\, qr + I_{13}(I_{11} - I_{22} + I_{33}) pq \\ \quad -2\left(e_1 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2\, e_3\right)(B z_b I_{33} + (B - mg) P_2^1 I_{13})\big), \\ \dot{q} = \dfrac{1}{I_{22}}\big(M_q q + I_{13}(p^2 - r^2) + (I_{11} - I_{33})\, pr \\ \quad -2\left(e_2 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} - e_1\, e_3\right)(B z_b + P_1^3 (B - mg))\big), \\ \dot{r} = \dfrac{1}{\Delta}\big(L_p I_{13} p - N_r I_{11} r + (-I_{13}^2 - I_{11}^2 + I_{11} I_{22})\, pq + I_{13}(-I_{11} + I_{22} - I_{33}) qr \\ \quad +2(e_1 \sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2\, e_3)(P_2^1 I_{11}\ (B - mg) + B z_b I_{13})\big), \\ \dot{e}_1 = \dfrac{1}{2}\left(p \sqrt{1 - e_1^2 - e_2^2 - e_3^2} + e_2 r - e_3 q\right), \\ \dot{e}_2 = \dfrac{1}{2}\left(e_3 p + q \sqrt{1 - e_1^2 - e_2^2 - e_3^2} - e_1 r\right), \\ \dot{e}_3 = \dfrac{1}{2}\left(-e_2 p + e_1\, q + r \sqrt{1 - e_1^2 - e_2^2 - e_3^2}\right). \end{cases} \tag{20}$$

The system (20) can be expressed as:

$$\begin{cases} \dot{\delta} = A\delta + S(\delta, e_3), \\ \dot{e}_3 = R(\delta, e_3), \end{cases} \tag{21}$$

where $\delta = (p, q, r, e_1, e_2)'$, S and R represent higher order nonlinear terms and vanish when $\delta = 0$.

The linearized system of (20) with respect to δ is given as follows:

$$\begin{cases} \dot{p} = \dfrac{1}{\Delta}(-L_p I_{33} p + N_r I_{13} r - 2(B z_b I_{33} + P_2^1 I_{13} (B - mg) e_1)), \\ \dot{q} = \dfrac{1}{I_{22}}\big(M_q q - 2(B z_b + P_1^3 (B - mg)) e_2\big), \\ \dot{r} = \dfrac{1}{\Delta}(L_p I_{13} p - N_r I_{11} r + 2(B z_b I_{13} + P_2^1 I_{11} (B - mg)) e_1), \\ \dot{e}_1 = \dfrac{1}{2} p, \\ \dot{e}_2 = \dfrac{1}{2} q. \end{cases} \tag{22}$$

By using MATLAB Toolbox, the linear system (22) around the partial equilibrium point $(p, q, r, e_1, e_2)' = (0,0,0,0,0)'$ admits the following eigenvalues $-0.7597, -3.6869 + 0.4501i$, $-3.6869 + 0.4501i, -0.0970, -3.5560$. Clearly the linearized system is asymptotically stable, and therefore by using partial exponential stability and linearization theorem (Jammazi 2010), the initial system (21) is locally exponentially stable with respect to δ.

Since the function R satisfies the property $R(0, e_3) = 0$, then the state e_3 converges.

Moreover, we have $e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1$, then the state e_1, e_2 and e_3 are Lyapunov stable. Therefore, the system (20) is locally exponentially stable with respect to the partial state $(p, q, r, e_1, e_2)'$, stable with respect to $(p, q, r, e_1, e_2, e_3)'$ and the "uncontrolled" state e_3 converges. This achieves the proof. □

The airship is an example of system with drift in which the Coron and Rosier's condition fails to be stabilized by discontinuous feedback laws (Coron and Rosier 1994). In order to overcome this obstruction, the next proposition introduces the finite-time partial stabilizability by discontinuous feedback laws. By the same argument as in the proof of Proposition 3 we show the following proposition.

Proposition 4: *Let be $\alpha \in (0,1)$, then under the following feedback laws*

$$\begin{aligned} v_1 &= -a_1 sgn(x) - b_1\ sgn(\dot{x}), \\ v_2 &= -a_2\, sgn(y) - b_2\ sgn(\dot{y}), \\ v_3 &= -a_3\, sgn(z) - b_3\ sgn(\dot{z}), \end{aligned} \tag{23}$$

where $a_i > b_i > 0$, the underactuated system (2) is finite time stable with respect to $(x, u, y, v, z, w)'$ and locally exponentially stable with respect to $(p, q, r, e_1, e_2)'$ and therefore e_3 converges.

Now, we are ready to give the open question.

Open Question: Is the feedbacks proposed in the section 4 are robust with respect to measurement noise on the state variables and with respect to unmodeled dynamics ?

4. SIMULATION RESULTS

The performances of our feedback laws are tested by numerical simulations on the nonlinear model of airship. The advantage our method resides in obtaining a static stabilization. Moreover, the state variable, which is not “controllable” converges, which makes it possible to avoid the oscillation of the system in the neighborhood of the equilibrium point. For space reason, the simulations of the exponential stabilizability are omitted. Only the simulations of the finite-time partial stabilizability by continuous feedbacks are considered.

4.1. Second approach: Finite-time partial stabilizability by continuous feedback laws.

In this strategy we have used the initial condition:

$$(u^0, v^0, w^0, p^0, q^0, r^0, x^0, y^0, z^0, e_1^0, e_2^0, e_3^0)' = (0.3, 0.5, -1.5, 0.2, 0.1, 0.5, 0.5, -0.2, 0.5, -0.5, 0.5, 0.5)',$$

and the feedbacks

$$u_1 := -0.2\, sgn(u)|u|^{\frac{1}{3}} - 0.2\ \ sgn(x)|x|^{\frac{1}{5}}$$
$$u_2 := -0.2\, sgn(v)|v|^{\frac{1}{3}} - 0.2\ \ sgn(y)|y|^{\frac{1}{5}}$$
$$u_3 := -0.2\, sgn(w)|w|^{\frac{1}{3}} - 0.2\ \ sgn(z)|z|^{\frac{1}{5}}.$$

This simulation shows the finite time stability of $(u, v, w, w, y, z)'$. The asymptotic stability of $(p, q, r, e_1, e_2)'$ and convergence of $e_3\ to \simeq 0.21$.

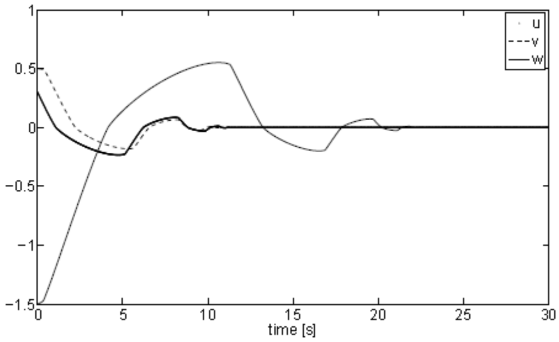


Figure 2: Velocities u and v and w

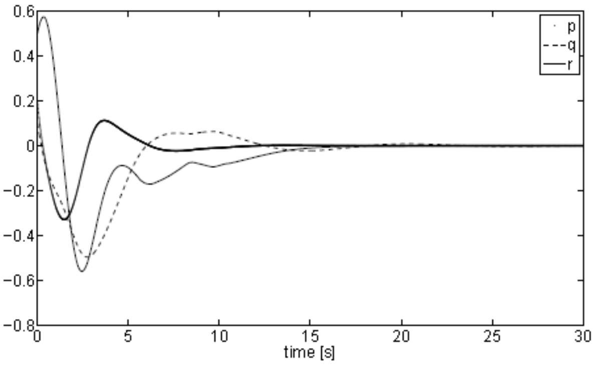


Figure 3: Velocities p and q and r.

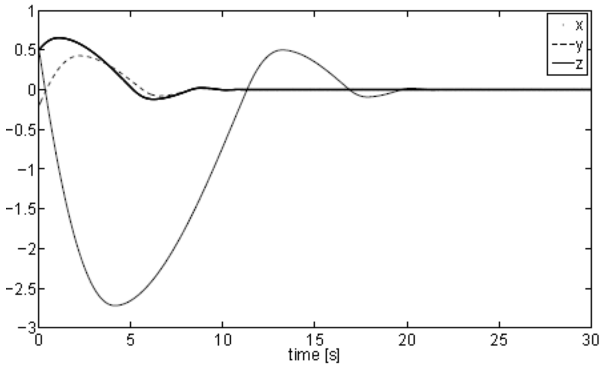


Figure 4: Positions x and y and z.

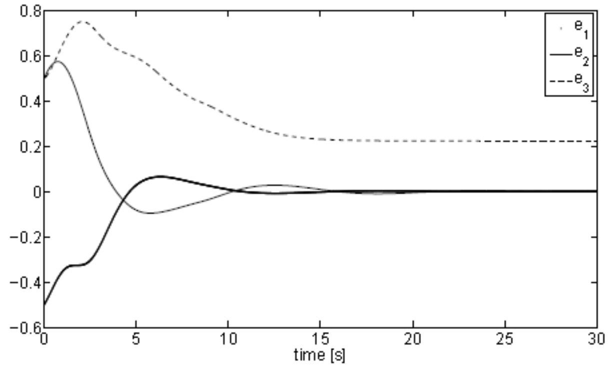


Figure 5: Positions e_1, e_2 and convergence of e_3.

5. CONCLUSION

The model of airship cannot be stabilized by continuous pure state feedback laws, this due to Brockett necessary condition. To overcome this problem, various controllers are proposed to study the position and the partial attitude of the airship; summarizing, these feedbacks makes eleven states of variables asymptotically stable once only one variable remains convergent. In the airship model, the "uncontrolled part “e_3” is the yaw angle which leads the system to revolve around the axis n_z attached to the frame airship. We have shown with the action of our feedback that the airship is asymptotically stable without taking into consideration its orientation with respect to axis n_z, since the latter angle converges. Clearly this stabilization seems sufficient.

REFERENCES

Astolfi., A., 1996a. *Asymptotic stabilization of nonholonomic systems with discontinuous control.* Thesis (PhD). Swiss Federal Institute of Technology Zurich.

Astolfi, A., 1996b. Discontinuous control of nonholonomic. *systems. Systems and control letters,* 27, 37–45.

Beji, L. and Abichou, A., 2005. Tracking control of trim trajectories of a blimp for ascent and descent flight manoeuvres. *Inter. J. Control,* 78 (10), 706–719.

Beji, L. Abichou, A. and Bestaoui, Y., 2004. Position and attitude control of an underactuated autonomous airship. *International Journal of Differential equations and Applications,* 8 (3), 231–255.

Bennaceur, S., 2009. *Modélisation et commande d'engins volants flexibles*. Thesis (PhD). Université d'Evry.

Bestaoui, Y., 2006. Nominal trajectories of an autonomous under-actuated airship. *International Journal of Control, Automation, and Systems,* 4 (4), 395–404.

Bhat, S. P. and Bernstein, D. S., 1997. Finite-time Stability of homogenoues systems. *Procceding of the American Control Conference,* pp. 2513–2514. June, (Albuquerque, New Mexico).

Bhat, S.P. and Bernstein, D. S. 1998. Continuous finite time stabilization of the translational and rotational double integrators. *IEEE Trans. Automatic Control*, 43 (5), 678–682.

Brockett, R. W., 1983. Asymptotic stability and feedback stabilization. *Differential geometric control theory, Progress in Math,* 27, 181–191.

Coron, J.-M., 1990. A necessary condition for feedback stabilization. *Systems and Control Letters,* 14 (3), 227–232.

Coron, J.-M., 1992. Global asymptotic stabilization for controllable systems without drift. *Math Control Signals Systems,* 5, 295–312.

Coron, J.-M., 1995. Stabilization in finite time of locally controllable systems by means of continuous time varying feedback laws. *SIAM J.* Control and Optimization, 33 (3), 804–833.

Coron, J.-M., 2007. *Control and Nonlinearity:* Mathematical Surveys and Monographs.

Coron, J.-M. and Keraï, E.Y., 1996. Explicit feedbacks stabilizing the attitude of a rigid spacecraft with two torques. *Automatica,* 32, 669–677.

Coron, J.-M. and Rosier, L., 1994. A relation between continuous time-varying and discontinuous feedback stabilization. *J. Math. Systems Estimation and Control,* 4, 67–84.

Hong, Y. Yang, G. Cheng, D. and Spurgeon, S., 2004. Finite time convergent control using terminal sliding mode. *Journal of control theory and applications*, 2, 69–74.

Huang, X. Lin, W. and Yang, Bo., 2005. Global finite time stabilization of a class of uncertain non linear systems. Automatica, 41, 881–888.

Hygounenc, E., 2003. *Modélisation et commande d'un dirigeable pour un vol autonome*. Thesis (PhD). Université de Paul Sabatier de Toulouse.

Jammazi, C., 2008a. Backstepping and Partial Asymptotic Stabilization. Applications to Partial Attitude Control. *Inter. J. Contr. Automation, and Systems*, 6 (6), 859–872.

Jammazi, C., 2008b. Finite-time partial stabilizability of chained systems. C. R. Acad. Sci. Paris. 975-980, Ser. I (346) (17-18), 975–980.

Jammazi, C., 2010. On a sufficient condition for finite time partial stability and stabilization: Applications. *IMA J. Mathematical Control and Information*, 27 (1), 29–56.

Jammazi, C., 2011. A discussion on the Hölder and finite time partial stabilizability of Brockett's Integrator. *ESAIM: Control, Optimisation and Calculus of Variations, doi:10.1051/cocv/2010101.*

M'Closkey, R. T. and Murray, R. M., 1997. Exponential stabilization of driftless nonlinear control systems using homogeneous feedback. *IEEE Trans. Autom.Cont,* 42, 614-628.

Morin, P. and Samson, C., 1997. Time-varying exponential stabilization of a rigid spacecraft with two control torques. *IEEE Trans. Autom. Cont.*, 42:528–534.

Morin, P. Samson, C. Pomet, J-B, and Z-Ping Jiang., 1995. Time-varying feedback stabilization of the attitude of a rigid spacecraft with two controls. *Systems and Control Letters,* 25, 375–385.

Hong, Y. Yang, G. Cheng, D. and Spurgeon, S.,2004. Finite time convergent control using terminal sliding mode. *Journal of control theory and applications*, 2, 69–74.

Huang, X. Lin, W. and Yang, Bo., 2005 Global finite time stabilization of a class of uncertain nonlinear systems. *Automatica*, 41, 881–888.

Orlov, Y., 2005. Finite time stability and robust control synthesis of uncertain switched systems. *SIAM J. Control Optim,* 43 (4), 1253–1273.

Orlov, Y., 2009. *Discontinuous systems - Lyapunov Analysis and Robust Synthesis under Uncertainty Conditions.* London : Springer-Verlag.

Pettersen, K.Y. and Egeland, O., 1999. Time-varying exponential stabilization of the position and attitudeof an underactuated autonomous underwater vehicle. *IEEE Trans. on Automatic Control,* 44 (1), 112–115.

Pettersen, K. Y. and Nijmeijer, H., 2001. Underactuated ship tracking control: theory and expriments. *Int.J. Control,* 74 (14), 1435–1446.

Ryan, E.P., 1994. On Brockett's condition for smooth stabilizability and its necessity in a context of nonsmooth feedback. *SIAM J. Contr. Optim.* 32, 1597–1604.

Samaali, S. Abichou, A. and Beji, B., 2007. Longitudinal and horizontal stabilization of an unmanned blimp. *IJVAS,* 5 (1-2), 138–157.

Samson, C., 1995. Control of chained systems: Application to path following and time-varying point stabilization of mobile robots. *IEEE Trans. Autom. Cont.*, 40 (1), 64–77.

Sontag, E.D. and Sussmann, H.J., 1980. Remarks on continuous feedback. *IEEE CDC, Albuquerque,* 2, 916–921.

Sussmann, H.J., 1979. Subanalytic sets and feedback Control. *J. Differential Equations,* 31 (1), 31–52.

Zuyev, A.L., 2001. On partial stabilization of nonlinear autonomous systems: Sufficient conditions and examples. *In Proc. of the European Control Conference ECC'01,*pp. 1918–1922. Porto (Portugal).

Zhang, H. and Ostrowski, J.P., 1999. Visual servoing with dynamics: control of unmanned blimp. *Proceedings of IEEE conference on Robotics and Automation.* May 10-15, In Michigan USA Detroit.

Coron, J.-M. and d'Andréa Novel, B., 1992. Smooth stabilizing time-varying control laws fo a class of nonlinear systems. Applications to mobile robots. . In: Michel Fliess. IFAC Nonlinear Control Systems Design (Bordeaux, France, 1992), 413–418.

Hygounenc, E., 2001. Modélisation et commande d'un ballon de dirigeable souple de petite taille. Rapport LAAS 01170, CNRS-LAAS.

Morin, P., 2004. Stabilisation de systèmes non linéaires critiques et application à la commande de véhicules. Habilitation à diriger des recherches. L'université de Nice-Sofia Antipolis.

Pettersen, K.Y. and Egeland, O., 1996. Exponential stabilization of an underactuated surface vessel. In Proc. 35th IEEE Conf. on Decision and Control,Kobe, Japan.

Samson, C., 1991. Velocity and torque feedback control of a nonholonomic cart In Proceeding of International Workshop on Nonlinear and Adaptive Control, volume 162, chapter Advanced Robot Control, pages 125–151. Springer-Verlag.

Wichlund, K.Y. Sørdalen, O. J. and Egeland, O.,1995. Control properties of underactuated vehicles. In Proceedings of the IEEE Int. Conf. on Robotics and Automation, (Nagoya, Japan), pages 2009–2014.

INTERVAL APPROACH FOR ROBUST FAULT DIAGNOSIS

M. S. Jha [(a)], G. Dauphin-Tanguy [(a)], B. Ould Bouamama [(b)]

LAGIS, UMR CNRS 8219,
[(a)] Ecole Centrale de Lille, BP 48, 59651 Villeneuve d'Ascq cedex, France
[(b)] Polytech-Lille, 59650 Villeneuve d'Ascq, France

mayank-shekhar.jha@ec-lille.fr, genevieve.dauphin-tanguy@ec-lille.fr,
belkacem.ouldbouamam@polytech-lille.fr

ABSTRACT

It is seen that methodology for diagnosis of uncertain systems using the Bond Graph (BG) model in Linear Fractional Transformation (LFT) form is an effective way to detect the fault by systematic generation of robust adaptive thresholds. There are limitations associated, many of which are imposed due to the manner uncertain parameters can be treated. In this work, a new method of generating robust and adaptive thresholds is developed, where uncertain parameters can be treated as intervals that vary between certain upper and lower bounds. The developed method proposes to generate the envelopes, using the interval extension form of Analytic Redundancy Relations (ARRs) in continuous time domain. The interval extensions of ARR functions are obtained by replacing the uncertain parameters with their interval equivalent. A very optimum range of such a function can be obtained by taking into account parameter variation in the interval. Superior and inferior values of the range determine the thresholds that form the envelope. The methodology developed is then implemented over an uncertain model of DC Motor wherein cases of parameter deviation with time, multiple uncertain parameters are considered.

Keywords: Uncertain parameters, Bond-Graph, BG-LFT, Interval, Fault diagnosis.

1. INTRODUCTION

Dynamic systems can be represented by continuous state space models as in (1),

$$\dot{x} = f(x,u,d,\theta_n) \qquad (1)$$

$$y = g(x,u,d,\theta_n)$$

where $x \in \Re^n$, $u \in \Re^m, d \in \Re^l, y \in \Re^r, \theta \in \Re^p$, denote respectively the state vector, the control input vector, the disturbance vector , the output vector and the vector regrouping the model parameters.

The system parameters may vary around their nominal value in certain range. For example, a resistor element R, of nominal value 10Ω, may have manufacturing tolerance as 10Ω±1Ω. A mechanical friction coefficient will increase from its nominal value (which may be zero or otherwise), to a value permissible for normal functioning of the system because of thermal effects and ageing of the materials.

To represent uncertainty in a parameter, different forms can be used. Additive uncertainty as represented in (2) is equivalent to (3) in an interval form. As generalization, a non symmetrical interval of this form is represented as in (4). Multiplicative uncertainty can be represented as shown in (5).

$$\theta = \theta_n \pm \Delta\theta \, , \Delta\theta > 0 \qquad (2)$$

$$\theta = [\theta_n - \Delta\theta, \theta_n + \Delta\theta] \qquad (3)$$

$$\theta = [\theta_n - \Delta\theta_1, \theta_n + \Delta\theta_2] \qquad (4)$$

with

$\Delta\theta_1 \ge 0$, $\Delta\theta_2 \ge 0$ $\Delta\theta_1 \ne \Delta\theta_2$

$$\theta = \theta_n(1 \pm \delta\theta) \, , \; \delta\theta = \frac{\pm\Delta\theta}{\theta_n} \qquad (5)$$

Robust diagnosis of uncertain systems has been an interesting, widely studied area of recent research works (Patton et al. 1989, Djeziri et al, 2007). Efficient and optimal methods for robust fault detection and isolation are required. Dauphin-Tanguy et al (1999), Djeziri et al (2006, 2007) highlight the method of using Bond Graph in Linear Fractional Transformation (LFT) approach as a powerful tool for this purpose. When using Analytical Redundancy Relationships (ARRs) for Fault Detection and Isolation (FDI), uncertainties in parameters generate envelopes around the nominal trajectories defining a domain corresponding to thresholds inside which the behaviour of the system can be considered as "normal" or "non-faulty". To avoid false alarms, the thresholds must be formed accurately and optimally.

The determination of ARRs on a bond graph model in preferred derivative causality is done by elimination of unknown variables contained in the structural constraints of junctions 0 and 1. The equations of power balance on the junctions constitute the ARRs candidate (Ould Bouamama et al. 2005). The bond graph model in

LFT form, obtained from multiplicative form of the uncertainties, allows the generation of the ARRs $r_{j,}$ with j=1,2…$n_{r.}$ The residual $r_{j,}$ is composed of two completely separated parts:

- a nominal part r_{j_n} , of the residual r_j ,with j=1,2…n_r, as shown in (6). Here n_r is the number of residuals.

$$r_{j_n} = \Phi_j \left\{ \begin{matrix} Se, Sf, De, Df, De_m, Df_m, R_n, C_n, I_n, TF_n, \\ GY_n, RS_n \end{matrix} \right\} \tag{6}$$

where TF_n and GY_n are respectively the nominal values of TF and GY moduli. R_n; C_n; I_n and RS_n are the nominal values of elements R; C; I and RS, De and Df being the measured variables and De_m and Df_m are the dualised signal sources. The method is well developed in Djeziri et al. (2006)).

- an uncertain part b_j, serves for the calculation of adaptive thresholds and sensitivity analysis, as shown in (7) and (8).

$$b_j = \Sigma w_i \tag{7}$$

$$w_i = \Theta_i \left\{ \begin{matrix} Se, Sf, De, Df, De_m, Df_m, R_n, C_n, I_n, TF_n, \\ GY_n, RS_n, \delta_R, \delta_I, \delta_C, \delta_{TF}, \delta_{GY}, \delta_{RS} \end{matrix} \right\} \tag{8}$$

where, δ_R ,δ_I, δ_C, δ_{RS}, δ_{TF}, δ_{GY} are values of multiplicative uncertainties as shown in (6).

The threshold being defined on residual r_j equation (9), it generates an envelope around nominal residual as in equation (10).

$$a_j = \Sigma | w_i | \tag{9}$$

$$-a_j < r_j < a_j \tag{10}$$

Because of the definition of the thresholds itself as in Eq. (9), using the absolute values, the envelope determines a domain around the nominal trajectory with symmetrical upper and lower frontiers which is efficient when the uncertain parameters are defined as $\theta_n - \Delta\theta \le \theta \le \theta_n + \Delta\theta$ (case of manufactured components with a defined tolerance). It may induce an overestimation of the envelope if the bounds are not symmetrical as $\theta_n - \Delta\theta_1 \le \theta \le \theta_n + \Delta\theta_2$ with $\Delta\theta_1 \ne \Delta\theta_2$, or when the uncertainty appears in one side only as in case of irreversible deviation with $\theta_n \le \theta \le \theta_n + \Delta\theta$ (case of time deviation).

It is seen, even though BG-LFT approach offers an effective method for fault detection by adaptive threshold generation that it is not successful in accounting all various types of uncertainty found commonly. As such, treatment of uncertain parameters as intervals offers much scope and better methods for FDI.

In this work, the objective is to treat the uncertain parameters as intervals and use them to generate adaptive thresholds. The new method presented here generates adaptive thresholds and forms envelopes, using the interval extension form of the ARRs in continuous time domain which helps in fault detection and identification. Besides this section of Introduction, paper is divided into four other sections. Section 2 introduces the method of treating the uncertain parameters, features of INTLAB as the software tool to carry out simulations is given, interval extensions of real functions are briefly discussed to introduce the interval extensions of ARRs which are subsequently used to generate the thresholds. Further, the methodology is proposed .In section 3, the methodology is implemented over a model of standard direct current (DC) motor example. Section 4 contains the simulations and results. Finally, conclusions are drawn in the final section 5.

2. INTERVAL APPROACH FOR ROBUST DIAGNOSIS

The interval analysis was initially developed to hold account on the inaccuracies of the numbers. These inaccuracies come from data resulting from a chain of instrumentation or from computing tool. Interval analysis in the past one decade has been exploited to a great extent in many disciplines of engineering. Because of interval arithmetic's power to bound ranges of functions, interval arithmetic has been most successful in solution of nonlinear systems and global optimization (Hansen E.R, 1992). Much information can be found in the literature available, concerned with many disciplines of science and maths where it is used to obtain rigorous proofs or results (Jaulin, Keiffer, Didrit and Walter, 2001).

Early work on treatment of uncertain parameters as intervals and subsequent usage for diagnosis is found in works of Adrot et al, (2000). The approach, called bounded approach, represented these uncertainties by a set of possible values for which only their bounds were known. Ragot et al., (1997) proposed an interval technique for the detection and the isolation of sensor faults in the case of a static linear model .The case is treated for dynamic systems by Armengol. et al, (2003). D. Maquin and J. Ragot (Ragot et al, 2003a) treated the problem of data validation in the case of certain systems with uncertain measurements through interval approach. A method of data validation is developed based on bounded approach, for dynamic uncertain linear systems, where all the state variables of the system are measured (Ragot et al, 2003b). Here the proposed method treats the case where the state variables of the system are partially measured. The state estimation method is described which uses interval approach.

Finally, the method is used to detect, isolate and correct the data affected by faults. In the works of Fagarasan (2004), Rinner and Weiss (2004), interval calculation laws are used to generate the exact estimate of the output, bounds of the estimates are computed using traditional numerical integration techniques from the uncertain parameter interval vertices, assuming that monotonicity property holds. Thus, the envelopes generated, are primarily by the estimation of state or parameter.

2.1. INTLAB – The Choice

Many software tools are successful in implementation of interval arithmetic, prominent of them being, INTLAB (Rump, 1998), Sun's Forte C++ compilers with interval support (Sun Microsystems, 2001), filib++ (Lerch et al, 2000) and C-XSC (Hofschuster, 2001).Our choice to implement the interval computations and carry out simulations is INTLAB.

INTLAB (Rump, 1998; 1999) is a well-designed interval toolbox for the interactive programming environment MATLAB (The Math Works, 2001). It is integrated with MATLAB, providing immediate access to an extensive range of tools that lets us develop algorithms, analyze and visualize simulations. It allows the more traditional infimum-supremum as well as the midpoint-radius representations of intervals. That is to say, parameter variation with different value of deviations may be treated by declaring it as an interval by infimum-supremum .Variations with same deviations may be declared using 'midrad' notation. As example $\theta = [\ \theta_n - \Delta\theta_1,\ \theta_n + \Delta\theta_2\ ]$ may be declared as *infsup* $(\theta_n - \Delta\theta_1,\ \theta_n + \Delta\theta_2\ )$. When $\theta = [\ \theta_n - \Delta\theta,\ \theta_n + \Delta\theta\ ]$, interval may be declared in midpoint-radius format as *midrad(*θ_n, $\Delta\theta$). Here nominal value θ_n is the centre and deviation $\Delta\theta$ is the radius in positive and negative direction.

Every computation using INTLAB is rigorously verified to be correct, including input and output. All the algorithms written in INTLAB, may be implemented in MATLAB. Since it is implemented in MATLAB, it can also be used by SIMULINK for simulations. Mathematical routines written in MATLAB code can be modified and run in INTLAB. INTLAB offers predefined problem solving routines for dense and sparse systems of linear and nonlinear equations and eigenvalue problems. A multi precision interval arithmetic, a slope arithmetic as well as routines automatic differentiation are also included (Hargreaves, 2002). The main features of INTLAB can be explored through demo files included as .m files in the package. To work with systems in real time, Real-Time Windows Target™(RTWT) provides a real-time engine for executing Simulink® models on a Microsoft® Windows® PC .Thus, using INTAB will facilitate easy interfacing of systems involving interval computations with Simulink and thus, in real time with RTWT. However as direct libraries for interval variables are not available in SIMULINK, its execution in RT-WT for real-time application may face some major difficulties.

2.2. Interval Value Extensions of Real Functions

Interval-valued functions are obtained by selecting a real-valued function f and computing the range of values $f(x)$ takes as x varies through some interval X. By definition, the result is equal to the set image $f(X)$. Now, *interval valued extensions* of functions are obtained by *extending* a given real-valued function f by applying its *formula* directly to interval arguments (Moore, 1966).

Consider the real-valued function $f$given by (11). The entity is a formula, not a function. Now the formula (12) is taken that describes function (11) and applied to the interval arguments. The resulting interval-valued function is the *interval extension_*of the function as in (13).

$$f(x) = 1 - x, \qquad x \in R. \tag{11}$$

$$f(x) = 1 - x \tag{12}$$

$$F(X) = 1 - X, \quad X=[\underline{X}, \overline{X}] \tag{13}$$

This way the domain is enlarged to include non-degenerate intervals and the degenerate intervals x=[x, x]. $F = F\ (X_1 ,..., X_n\ )$ is called *inclusion isotonic* (Jaulin et al,2001) if equation (14) is satisfied when

$$Y_i \subseteq X_i \text{ for } i = 1,...,n$$

$$F\ (Y_1 ,..., Y_n\ ) \subseteq F\ (X_1 ,..., X_n\ ). \tag{14}$$

If F is an inclusion isotonic interval extension of f, then according to the *Fundamental Theorem of Interval Analysis* (Moore, 1979),

$$f(X_1 ,..., X_n\ ) \subseteq F\ (X_1 ,..., X_n\ ) \tag{15}$$

This theorem is never violated in cases of inclusion isotonic functions. Now, all rational interval functions are inclusion isotonic. Thus, an interval value of F contains the range of values of the corresponding real function $f$when the real arguments of $f$lie in the intervals.

This forms a mean for the finite evaluation of upper and lower bounds on the ranges of values of real rational functions.

2.3. Interval Extensions of the Analytic Redundancy Relations (ARRs)-The Approach

The finite evaluation of upper and lower bounds on the ranges of values of real rational functions becomes very useful to find the range of interval extension of ARRs and thus, generate the thresholds. The analytic redundancy relations (ARRs) are essentially functions of measurement of state variables with physical parameters as coefficients. They are treated as rational functions, then their interval extension obeys the above relation as in Eq. (15).

Our approach is based on the above principle. The ARRs are obtained from bond-graph model in a preferred derivative causality. These ARRs are functions of known variables (sensor measurements, control inputs) with system uncertain parameters as coefficients. Consider an ARR as,

$$r_j = f_j(y,u,\theta_n) \tag{16}$$

with *j*=1,2…..n_r , where n_r is the total number of ARRs.

The interval extension of the ARRs is obtained by treating the ARR as a rational polynomial and putting intervals of the coefficient uncertain parameters as,

$$R_j = F_j(y, u, [\theta]) \tag{17}$$

The upper bound and lower bound of R_j are evaluated as the value of θ varies within the interval $[\theta]$ and from Eq.(15), give the envelope such that

$$r_j \subseteq R_j\,, \text{ whenever } \quad \theta \in [\, \theta_n - \Delta\theta_1, \theta_n + \Delta\theta_2 \,] \tag{18}$$

These upper and lower bound values form the thresholds and hence, the envelope is generated.

Remarks:

- The envelope generated by R_j guarantees to contain the residual r_j only when all the considered parameters are within their respective interval bounds.
- If the residual r_j goes out of the envelope then it surely means that either one or more of the uncertain parameter has gone beyond its considered interval range.
- When one or more, *but not all* considered parameters go beyond their interval range, the residual may remain inside the envelope. So, the envelop guarantees to contain the residual only when *all* the parameters considered in ARR as intervals are within their bound.

Thus, in case of multiple parameter uncertainties, set of combination of envelops should be considered on various ARRs to identify the fault occurred.

3. APPLICATION ON DC MOTOR

This section presents the application of the proposed approach on a DC Motor model with uncertain parameters. Firstly, the specification of DC motor is given after which, BG model of same in derivative causality is used to generate the ARRs. Intervals extensions of ARRs are considered to generate the robust envelop. Importantly, cases of parameter deviation, ageing, or a catastrophic fault are considered through simulations.

3.1. Description of the DC Motor System

The model parameters are as resistance of stator, R = 2.4 Ω; inductance of the stator, L = 0.44 H; coefficient of couple, k_e = 0.14; moment of inertia of rotor, J = 0.08 kg m^2; coefficient of friction of motor shaft, f = 0.01 NS/ m ;the inputs being the voltage of the inductor U(t)=220 V, the load torque $\tau(t)$ =5 Nm; and the observed outputs being current of inductor i_m(t) and angular velocity of the motor shaft $\omega_m(t)$.

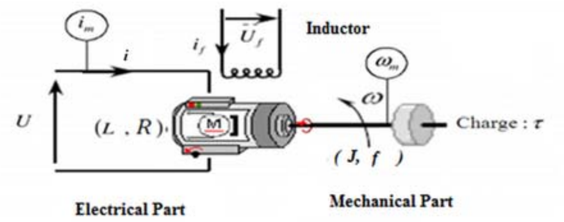


Figure 1: DC motor system representation.

The system can be modelled by bond graph in integral causality and can be represented in state space form as:

$$\begin{cases} \begin{bmatrix} \dot{i} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} -\frac{R}{L} & -\frac{F_l}{L} \\ \frac{F_l}{J} & -\frac{f}{J} \end{bmatrix} \begin{bmatrix} i \\ \omega \end{bmatrix} + \begin{bmatrix} \frac{1}{L} & 0 \\ 0 & -\frac{1}{J} \end{bmatrix} \begin{bmatrix} U \\ \tau \end{bmatrix} \\ y = \begin{bmatrix} i_m \\ \omega_m \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ \omega \end{bmatrix} \end{cases} \tag{19}$$

3.2. Derivation of Analytic Redundancy Relations

The bond graph model of the system in derivative causality (BGD) is given in Fig.2. The electrical motor is characterized by the resistance R and inductance *L*. The mechanical part of the motor is characterized by friction *f*, inertia *J* and the load torque τ. The conversion of electrical energy into mechanical energy is represented by a transformer GY.

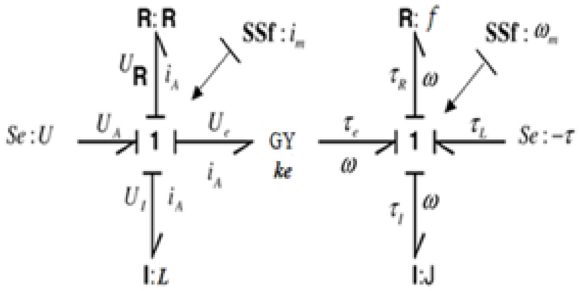


Figure 2: Bond graph model of DC motor in derivative causality with dualized sensors.

Here, the sensors have been dualized as corresponding sources of signal (SS). For simulation purpose, the measurements of sensors can be obtained from the BGI model, Eq.(19). But for unknown variable elimination to generate an ARR, the initial point is the source of signal. This covering of causal paths leads to well known oriented graphs. Any ARR candidate can be derived very easily from this model by considering the balance of powers on both junctions. The algorithm can be consulted in (Samantaray and Ould Bouamama, 2008).They are obtained as equations (20) and (21).

$$r_1 : U - R_n i_m - L_n \frac{di_m}{dt} - k_{en}\omega_m = 0 \qquad (20)$$

$$r_2 : k_{en} i_m - f_n \omega_m - J_n \frac{d\omega_m}{dt} - \tau = 0 \qquad (21)$$

Note that here all the parameters considered are with nominal values.

3.3. Interval Extensions of the ARRs

Next, the interval extension of r_1 and r_2 are considered by replacing the nominal parameters by their interval counterparts.

$$R_1 : U - [R] i_m - [L] \frac{d(i_m)}{dt} - \omega_m [ke] = 0 \qquad (22)$$

$$R_2 : i_m [k] - [f] \omega_m - [J] \frac{d(\omega_m)}{dt} - \tau = 0 \qquad (23)$$

Here, $[\theta]$ denotes [θ_n- $\Delta\theta_1$, θ_n+ $\Delta\theta_2$] and deviations in either direction may be equal (a special case). Interval bounds are considered on all the parameters. This can be done selectively too, depending upon which parameter we want to consider for envelope. Thus, R_1 as Eq. (22) and R_2 as Eq. (23) are the interval extensions of r_1 and r_2. They are evaluated by INTLAB in offline mode.

By definition, R_1 and R_2 are not rational polynomials as they have terms with differential operator such as $\frac{di_m}{dt}$ in R_1 and $\frac{d\omega_m}{dt}$ in R_2. However, if the values of such a term (say $\frac{d\omega_m}{dt}$) is obtained externally, then they can be directly substituted in the expression during evaluation by INTLAB. In our case, such expressions are evaluated “externally” in SIMULINK where the differential operation on i_m in (22) and on ω_m in (23) is done in continuous time domain. The obtained values are then “sent” to INTLAB where the interval expressions (22) and (23) are evaluated as a rational expression.

R_1 and R_2 can thus be treated as rational interval polynomials. This is essential, to have them inclusion – isotonic which is a necessary condition to satisfy the fundamental theorem of interval analysis.

4. SIMULATIONS AND RESULTS

Parameters are considered selectively, in different cases with each having interval bound as θ = [θ_n- $\Delta\theta_1$, θ_n+ $\Delta\theta_2$] or $\theta \in$ [θ_n- (0.1*θ_n) , θ_n+ (0.1*θ_n)], for normal (non-faulty functioning).In following cases, $\Delta\theta_1$=$\Delta\theta_2$=$\Delta\theta$. But, in general such is not the case. Other cases when $\Delta\theta_1 \neq \Delta\theta_2$ can also be treated similarly and efficiently. It is not done here for sake of clarity.

Note: In the following cases parameter deviation such as $\Delta\theta_e$ denotes the parameter deviation considered to form the envelopes (subscript ‘e’ denotes deviation to form envelopes) and is always considered as

$\Delta\theta_e$ =0.1*θ_n .

4.1. Case 1: Only one parameter uncertain for envelope generation

The envelope is generated by considering interval on frictional parameter, *f* only. Thus, the interval extension of residual r_2 is obtained as R_2 in (24).

$$R_2 : i_m ke(t) - [f]\omega_m - J_n \frac{d(\omega_m)}{dt} - \tau = 0 \qquad (24)$$

The parameter deviation for envelope generation Δf_e, is taken as Δf_e =0.1* f_n so that, $[f]$= [f_n- Δf_e, f_n + Δf_e]. The envelope in Fig.3, Fig.4 and Fig.5 is deduced from the interval extension expression (24). Nominal parameters are considered.

In Fig 3, residual r_2 corresponds to Δf =0, the ideal condition, residual is zero. In Fig 4, residual r_2 corresponds to Δf =0.08* f_n , parameter deviates but below interval bound of envelope (Δf_e =0.1* f_n), residual is below upper threshold and bound by the envelope. In Fig 5, residual r_2 corresponds to Δf=0.15* f_n, the damage (faulty) condition, residual is well above the threshold, beyond the envelope.

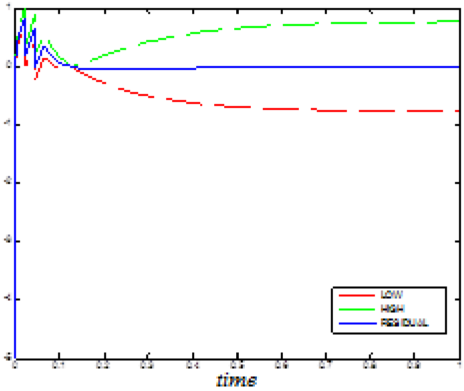


Figure 3: Residual r_2 with Δf=0

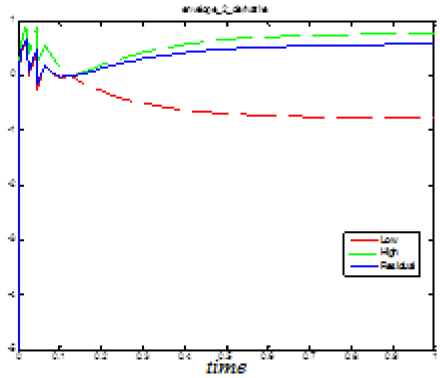


Figure 4: Residual r_2 with Δf=0.08*f_n

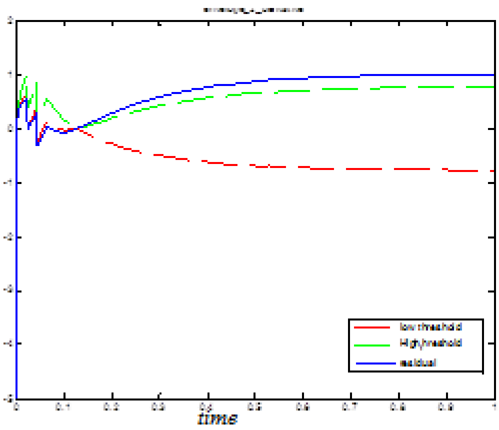


Figure 5: Residual r_2 with $\Delta f=0.15*f_n$

4.2. Case 2: A Parameter (friction) variation with time

In this case, the parameter is considered to vary with time. So, after a finite time, it goes outside the permissible limit. It is shown in Fig.6. Envelope is built by considering interval on frictional parameter, $\Delta f_e=0.1*f_n$. The residual corresponds to situation when the parameter is time varying, and goes out of interval limit after a certain time ie. $f=\Theta(t)$. It can be seen, if and when a parameter deviates outside its interval limit. This way, a slow ageing (a case of slow parameter variation with time) of a parameter can be identified.

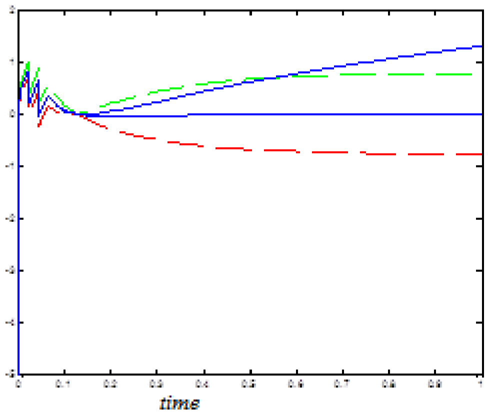


Figure 6: Parameter variation with time, Residual r_2 in blue.

4.3. Case 3: Abrupt-sudden change in a parameter

We consider the friction here to show a sudden, non-called for variation. A pulse is introduced as variation in friction (something similar associated with dry friction behaviour). The pulse variation can easily be detected by the residual. If the pulse amplitude is greater than that of interval limit, then it causes the residual to go outside the envelope. It is depicted in Fig.7. Here, the pulse is introduced at time 0.5 sec approx. in the frictional parameter, f. It is well reflected in the residual, causing it to go outside the envelope depending on the amplitude. Thus, fault is detected at instant t=0.5 s. At t=0.82 s, the pulse amplitude is just equal to the threshold limit

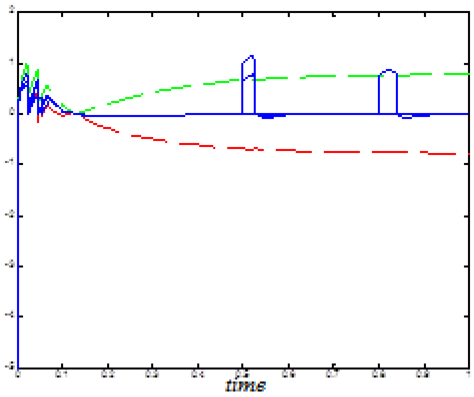


Figure7: Abrupt behaviour of a parameter

4.4. Case 4: Two parameters uncertain for envelope generation

In such cases, envelopes are built considering all the parameters that are likely to deviate. Envelope is built by considering interval on both friction f and ke. This generates the envelopes as seen in Fig.8 and Fig.9. In both figures envelope is generated considering, $\Delta ke_e=0.1*ke_n$, and $\Delta f_e=0.1*f_n$.

In Fig 8, residual r_2 corresponds to $\Delta f=0$, $\Delta ke=0$ the ideal condition, residual is zero. In Fig.9, the residual corresponds to, $\Delta ke=0.08*ke_n$, and $\Delta f=0.08*f_n$ (both below the interval limit.) The residual remains successfully inside the envelope. In Fig.10, the residual r_2 is drawn for $\Delta ke=0.15*ke_n$, and $\Delta f=0.15*f_n$ (both above the interval limit.) The residual goes outside the envelope. So, it is seen that when both parameter go outside their limit, then residual goes out of the envelope too.

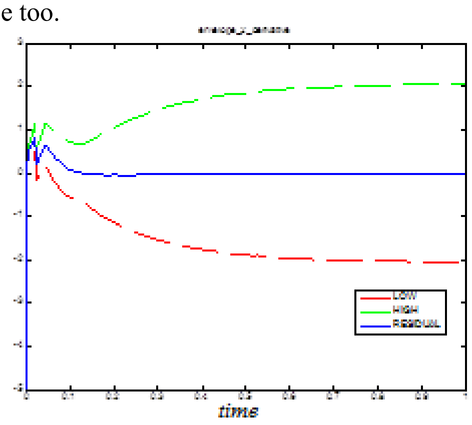


Figure 8: Residual r_2 with $\Delta f=0$, $\Delta ke=0$

The envelope generated, guarantees to contain the residual, only when all the considered parameters are within their interval bounds. Its converse is however, not true.

Conversely, (a) if one of them goes beyond its interval limit (say friction f. not ke) , then the residual might go outside. and equivalently,(b) one of the parameter may go beyond its interval limit(say *ke*, not friction), yet the residual may be inside the envelope.

The former situation (a) is alright, as it however, results in fault detection as residual goes out. But, latter situation (b), can result in missed alarms.

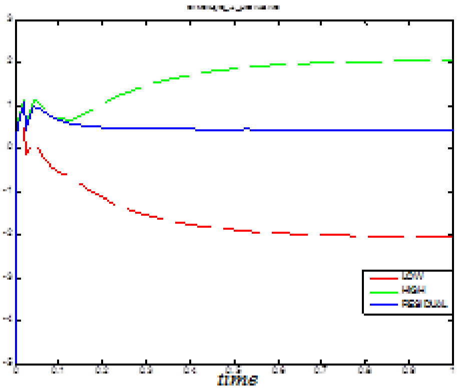


Figure 9: Residual r_2 with $\Delta ke = 0.08 * ke_n$, $\Delta f = 0.08 * f_n$

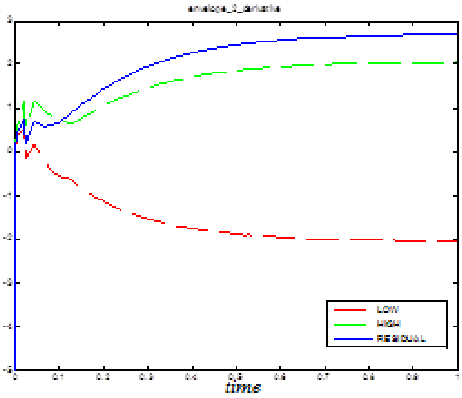


Figure 10: Residual r_2 with Δke $= 0.15 * ke_n$, $\Delta f = 0.15 * f_n$

In Fig.11, envelope is generated by considering, the two parameters friction *f* and *ke* (like in previous three figures.).One of the parameter, *ke* goes faulty, $\Delta ke = 0.12 * ke_n$ and friction *f* is nominal. $\Delta f = 0 * f_n$. The residual however, remains inside. This is due to the relative influence of the parameters and is linked to their order of magnitude.

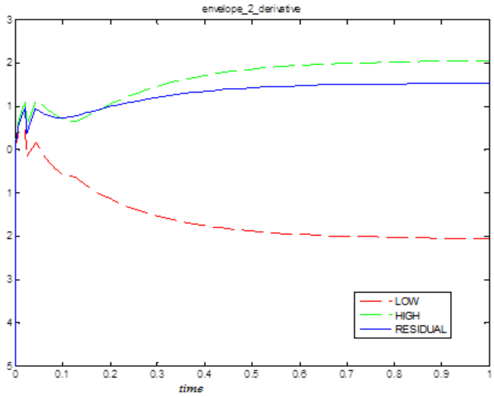


Figure11: Envelope considers both *f* and *ke* ($\Delta ke_e = 0.1 * ke_n$, and $\Delta f_e = 0.1 * f_n$). Residual r_2 corresponds to situation when, $\Delta ke = 0.12 * ke_n$, $\Delta f = 0 * f_n$.

To diagnose such cases, set of envelopes are built as shown in Fig.12.Env 1- by considering only ‘suspected’ variable that will deviate *ke* and Env 2- by considering both *ke* and *f*. Then, as shown in Fig.12 the residual will not be included in Env 1- showing *ke* deviates and is included in Env -2, showing *f* doesn’t deviate.

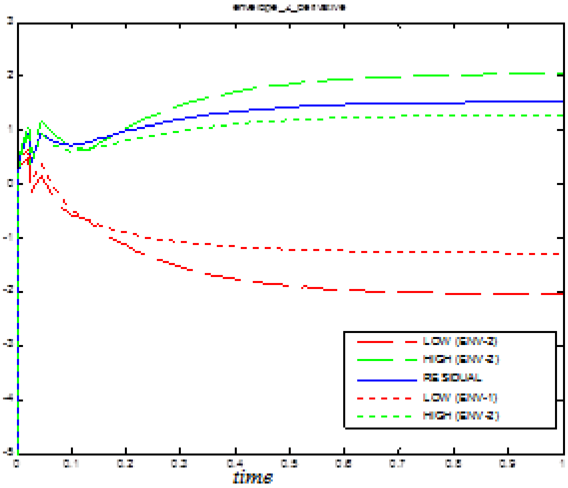


Figure 12: Env 1-considers only *ke* as $\Delta ke_e = 0.1 * ke_n$, Env-2 considers both *ke* and *f* as $\Delta ke_e = 0.1 * ke_n$, and $\Delta f_e = 0.1 * f_n$. Residual r_2 has $\Delta ke = 0.12 * ke_n$, $\Delta f = 0 * f_n$.

5. CONCLUSIONS

This method of fault diagnosis by generation of thresholds from the interval value extensions of analytical redundancy relations (ARRs) is robust and efficient. There are no limitations in terms of nonlinearities. The envelopes guarantee to contain the residual when all the parameters considered for the envelope, vary within their interval range. System with single or multiple uncertain parameters can be dealt effectively, by creating a series of envelopes. Parameters are treated as intervals for threshold generation, and thus a parameter with variations different in either side or with deviation that varies with time, can be treated easily, which is not possible with the BG-LFT approach. This renders a much more freedom in treating the uncertain parameters accurately. This results in efficiently optimized envelopes.

REFERENCES

Adrot, O., (2000), Diagnostic à base de modèles incertains utilisant l’analyse par intervalles: l’approche bornante, Doctorat de l’Institut NationalPolytechnique de Lorraine, Specialité automatique, Vandoeuvre-les Nancy, France.

Armengol,J. Vehi,J., Angel Sainz, M., and P.J. Herrero,(2003) ‘Fault detection in a pilot plant using interval models and multiple sliding time windows’, Safeprocess - Washington - USA, 729–734.

Dauphin-Tanguy, G., Sie´ Kam, C., (1999). 'How to Model Parameter Uncertainies in a Bond Graph Framework'. ESS'99, Erlangen. Germany. pp. 121-125

Djeziri, M. A., Merzouki, R., Ould Bouamama, B., Dauphin Tanguy, G., (2006). 'Fault Detection of Backlash Phenomenon in Mechatronic System with Parameter Uncertainties using Bond Graph Approach', Proceeding of the 2006 IEEE International Conference on Mechatronics and Automation Luoyang, China. pp. 600-605.

Djeziri, M. A., Merzouki, R., Ould Bouamama, B., Dauphin Tanguy, G., (2007). ' Bond Graph Model Based For Robust Fault Diagnosis', Proceeding of the 2007 American Control Confer- ence New York City, USA. pp. 3017-3022.

Fagarasan, I., Ploix, S. & Gentil, S. (2004). Causal fault detection and isolation based on a set-membership approach. Automatica, 40, 2099,2110. 27, 60

Hansen, E. R., (1992). Global Optimization Using Interval Analysis. Marcel Dekker, Inc., New York,

Hargreaves, G.I.,(2002) ,*Interval Analysis in MATLAB.* Master's thesis, Department of Mathematics, University of Manchester.

Hofschuster, W., Kramer, W., Wedner, S., Wiethoff, A,(2001).: C-XSC 2.0 – A C++ Class Library for Extended Scientific Computing, Preprint 2001/1, Wissenschaftliches Rechnen/Softwaretechnologie, Universit¨at Wuppertal.

Jaulin, L., Keiffer, M.,, O Didrit, O.,, Walter, E.,(2001) *Applied Interval Analysis*. Springer-Verlag, Berlin.

Lerch, M., Wolff von Gudenberg,(2000) J.: filib++, Specification, Implementation, and Test of a Library for Extended Interval Arithmetic, RNC4 proceedings..

Moore, R.E.,(1966), *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ.

Moore, R. E.,(1979) *Methods and Applications of Interval Analysis*. SIAM, Stud. Appl. Math. 2, Philadelphia.

Ould Bouamama, B., Samantaray, A.K., Staroswiecki, M., Dauphin-Tanguy,G., (2005).'Derivation of constraint relations from bond graph models for fault detection and isolation'.in:Proc. ICBGM'03,Simulation Series, vol. 35, no. 2, pp.104-109, ISBN. 1-56555-2571.

Ould Bouamama, B., and Samantaray, A. K., and (2008) "*Model-based Process Supervision. A Bond Graph Approach*". Springer Verlag, Series: Advances in Industrial Control, 490 p. ISBN:978-1-84800-158-9, Berlin

Patton R, J., Frank, P.M., and Clark, R.N., 1989, Fault diagnosis in dynamic systems : theory and application, Prentice Hall, New Jersey.

Ragot, J., Boukhris, A., and Mandel, D., (1997) 'A propos de l'algèbre des intervalles. application à la validation de données', Rencontres Francophones sur la Logique Floue - Lyon- France,

Ragot, J and Maquin, D.,(2003a), 'Validation de données issues de systèmes de mesure incertains', Journal Européen des Systèmes Automatisés RS série JESA numéro spécial Applications des outils de calculs ensemblistes,37, n9.

Ragot, J and Alhaj Dibo, M., Maquin, D., (2003b) 'Validation de données par approche bornante', Journées Doctorales d'Automatique - Valenciennes-France, 347–352,

Rinner, B. & Weiss, U. (2004). Online monitoring by dynamically re¯ning imprecise models. IEEE Transactions on Systems, Man, and Cybernetics, Part B34, 1811{1822. 60, 66

Rump, S., M., (1998), INTLAB - INTerval LABoratory, in Csendes, T. (Ed.), Developments in Reliable Computing, KluwerAcademic Publisher, Dordrecht, pp. 77-104.

Rump, S., M.,(1999), Fast and Parallel Interval Arithmetic, Bit, Vol. 39, No. 3, pp. 534-554, 014.

Sun Microsystems,(2001): C++ Interval Arithmetic Programming Reference (Forte Developer 6 update 2), Sun Microsystems, July 2001.

The MathWorks, Inc. (Publisher),(2001),: MATLAB, The Language of Technical Computing.

DYNAMIC MODELLING OF MULTI SECTION BIONIC MANIPULATOR: APPLICATION TO ROBOTINO-XT

Sonam Behl[a], Coralie Escande[b], Pushparaj Mani Pathak[c*], Rochdi Merzouki[d], B.K. Mishra[e]

[a,c,e] Mechanical and Industrial Engineering Department, Indian Institute of Technology, Roorkee, India.
[b,d] LAGIS FRE CNRS, 3303, Polytech'Lille, avenue Langevin, 59655, Villeneuve d'Ascq, France.
(a)(sonambehl5007@gmail.com), (b)Ecoralie.escande@googlemail.com, (c)*(pushpfme@iitr.ernet.in),
(d)(rochdi.merzouki@polytech-lille.fr), (e)(bhanufme@iitr.ernet.in)
*Corresponding author

ABSTRACT

This paper deals with the Bond Graph dynamic modeling of a multi section bionic manipulator applied to Bionic Handling Assistant robot (BHA). This bio-inspire flexible manipulator is comprised of two bending sections, one rotating section and one compliant gripper. Each bending section is comprised of three bellows, which are actuated pneumatically. To simulate the bellows behavior, differential equation for a pipe conveying fluid has been obtained and finite element representation of the mathematical model is carried out. This finite element approach has been linked to the bond graph modeling approach. Similarly remaining of the bellows have been modeled and are linked to each other by putting suitable boundary conditions to obtain the dynamic model of the manipulator as a whole. Thus, this model can be used to design an adequate control for the whole BHA robot. This model will be validated through simulation and experimental tests.

Keywords: Bionic Manipulator modeling, FE Bond Graph Model, Pneumatic actuator bond graph

1. INTRODUCTION

Bionic engineering started growing last decade, by finding the best correlation between biological and mechanical motions. According to the considered missions (Siciliano and Khatib 2009), with the biological inspired robots, the mechanism that is patterned after principles found in humans or nature is being reproduced. It also refers to artificial devices implanted into humans, replacing or extending normal human functions with certain accuracy. Two classifications of bionic robots have been done by Trivedi et.al (2008) with large spectrum of applications, one concerns the soft robots such as Pneumatic Artificial Muscle robots (PAM), and the other concerns the hard robots with soft capabilities such as Multi-Section Backbones Tubes robots (MSBT). Bionic Handling Assistant (BHA) from Festo (2011) can also be classified in the latter category. It is 3-sections pneumatically actuated with 3 backbones tubes per section. The flexibility of this bellow is provided by the Polyamide material fabricated by Rapid Prototyping method and gives a great maneuverability of 11 Degrees of Freedom (DOF).

To carry out dynamic modeling of such kind of manipulator, the various sections of manipulator arm need to be studied in detail. Numerous papers deal with various aspects of bellows, such as stresses due to internal pressure and axial deflection, dynamic analysis and static analysis and fatigue life estimations.

Andersson (1964, 1965) derived correction factors relating the behavior of the bellows convolution to that of a simple strip beam. This approach had subsequently been the basis of standards and other publications presenting formulae for hand-calculation for bellows.

The most comprehensive and widely accepted text on bellows design is however the Standards of the Expansion Joint Manufacturers Association (1998). The EJMA standards are found to comply better with experience, and are therefore recommended everywhere. Ting-Xin et al. (1990) and Osweiller (1989) did comparisons of the EJMA standards with finite element. The EJMA organization itself has also conducted research to verify the standards.

This paper presents the dynamic model of a single section of multi-section bionic manipulator for the 'Bionic Handling Assistant' robot of Festo called Robotino XT. This robot is omni directional mobile robot with bionic manipulator as shown in Figure 1. (Festo, 2011).

Figure 1: Robotino XT Robot

2. VARIOUS SECTIONS OF MANIPULATOR ARM

Robotino XT shown in Figure 1 (Festo, 2011), is a combination of a mobile Omni directional robot called

Robotino and a small version of the 'Bionic Handling Assistant' (BHA) mounted on it. This small bionic trunk has 8 degrees of freedom (DOF) and can be divided in several parts as shown in Fig. 2: two bending parts, a rotation part, and a compliant gripper. Each bending part is composed of three flexible backbones tubes, while the rotation part is actuated by two flexible backbone tubes in form of arc. The gripper works with the same principle of the rotation part in another plane. The backbones tubes are actuated with compressed air, where the position and the orientation are controlled by differentiating the pressures.

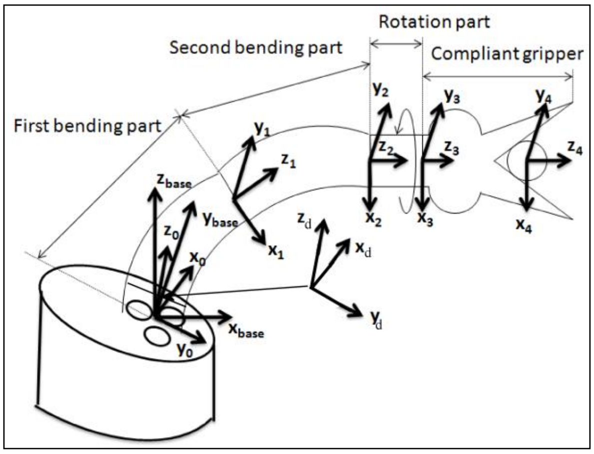


Figure 2: Model of the Manipulator arm

3. DYNAMIC MODELLING OF BELLOWS

To carry out dynamic modeling of the manipulator arm, the different sections of the arm were taken into account and studied thoroughly. Now, the bending sections are made up of bellows. According to Borman et al. (2000), special geometry of the bellows makes it very flexible in overall bending but very stiff against radial deflections. This means that significant transverse deflections and rotations of cross-sections perpendicular to the bellows axis can take place, with oval ling remaining negligible. Thus, although being far from a beam in appearance, a beam representation of the bellows should be relevant. Thus, each bellow has been modeled as a fictitious pipe, through which fluid is flowing.

3.1. Assumptions

Assumptions that have been taken while modeling the bellow as a beam are as follows:

- Plane sections originally normal to the longitudinal axis of the beam remain plane and normal to the deformed longitudinal axis upon bending.
- The beam is just undergoing deformation due to bending.
- The strains in the longitudinal, radial and torsional directions are assumed to be negligible.
- The effect of shear deformation on the dynamic response of the beam is ignored.
- The material is assumed to be isotropic and elastic.

3.2. Mathematical Model

The differential equation governing the bending vibrations of the bellow can be obtained from the general differential equation for bending vibration of pipe conveying fluid, given by Broman et al. (2011)

$$E_p I_p \frac{\partial^4 w}{\partial x^4} + (\rho_p A_p + \rho_f A_f)\frac{\partial^2 w}{\partial t^2} - \rho_p A_p (1 + \alpha \frac{E_p}{G_p}) \frac{\partial^4 w}{\partial x^2 \partial t^2} + 2\rho_f A_f V \frac{\partial^2 w}{\partial t \partial x} + (p\pi {R_m}^2 + \rho_f A_f V^2)\frac{\partial^2 w}{\partial x^2} = F \tag{1}$$

Where E_p is the young's modulus of elasticity of the pipe, I_p is the area moment of inertia for the pipe cross-section, w is the transverse displacement, x is the length of the pipe, ρ_p is the density of the pipe material, A_p is the area of cross section of the pipe, ρ_f is the density of the fluid, which is flowing inside the pipe, A_f is the flow area, t is the time, α is the shear coefficient, G_p is the polar modulus of elasticity of pipe, V is the flow velocity, F is the transverse force acting on the pipe.

The influence of shear deformation on bending of bellows is assumed to be very small (Morishita et al., 1989). Thus the shear coefficient can be neglected from Equation (1). Thus the equation reduces to:

$$E_p I_p \frac{\partial^4 w}{\partial x^4} + (\rho_p A_p + \rho_f A_f)\frac{\partial^2 w}{\partial t^2} - \rho_p A_p \frac{\partial^4 w}{\partial x^2 \partial t^2} + 2\rho_f A_f V \frac{\partial^2 w}{\partial t \partial x} + (p\pi {R_m}^2 + \rho_f A_f V^2)\frac{\partial^2 w}{\partial x^2} = \text{F} \tag{2}$$

To obtain the mathematical formulation of this model by FEM, first step is to convert the equation (2) into its weak formulation. We consider a beam element, whose two consecutive nodes are x_e and x_{e+1}. Each node has two degrees of freedom associated with it. Thus we obtain as follows:

$$\int_{x_e}^{x_{e+1}} v[E_p I_p \frac{\partial^4 w}{\partial x^4} + (\rho_p A_p + \rho_f A_f)\frac{\partial^2 w}{\partial t^2} - \rho_p I_p \frac{\partial^4 w}{\partial x^2 \partial t^2} + 2\rho_f A_f V \frac{\partial^2 w}{\partial t \partial x} + (p\pi {R_m}^2 + \rho_f A_f V^2)\frac{\partial^2 w}{\partial x^2} - F] \tag{3}$$

In the above equation, v is the weight function. Simplifying the equation (3), we obtain:

$$\int_{x_e}^{x_{e+1}} \frac{\partial^2 v}{\partial x^2}.E_p I_p \frac{\partial^2 w}{\partial x^2} + \int_{x_e}^{x_{e+1}} v \frac{\partial^2 w}{\partial x^2}(p\pi {R_m}^2 + \rho_f A_f V^2)$$

$$+\int_{x_e}^{x_{e+1}} v[(\rho_p A_p + \rho_f A_f)\frac{\partial^2 w}{\partial t^2} - \int_{x_e}^{x_{e+1}} v.\rho_p I_p \frac{\partial^4 w}{\partial x^2 \partial t^2}$$

$$+\int_{x_e}^{x_{e+1}} v[2\rho_f A_f V \frac{\partial^2 w}{\partial t \partial x}] = \frac{\partial\ v}{\partial x}\{E_p I_p \frac{\partial^2 w}{\partial x^2}\}_{x_e}^{x_{e+1}}$$

$$-v[E_{\mathrm{p}}\mathrm{I_p}\frac{\partial^3 w}{\partial x^3}]_{x_e}^{x_{e+1}} \tag{4}$$

So the stiffness matrix [K] can be obtained from the following part of the equation (4)

$$\int_{x_e}^{x_{e+1}} \frac{\partial^2 v}{\partial x^2}.E_p I_p \frac{\partial^2 w}{\partial x^2} + \int_{x_e}^{x_{e+1}} v.\frac{\partial^2 w}{\partial x^2}(p\pi {R_m}^2 + \rho_f A_f V^2) \tag{5}$$

The Mass Matrix [M] can be obtained from part of equation (4), given as

$$\int_{x_e}^{x_{e+1}} v[(\rho_p A_p + \rho_f A_f)\frac{\partial^2 w}{\partial t^2} - \int_{x_e}^{x_{e+1}} v.\rho_p I_p \frac{\partial^4 w}{\partial x^2 \partial t^2} \tag{6}$$

The damping matrix can be identified as:

$$\int_{x_e}^{x_{e+1}} v[2\rho_f A_f V \frac{\partial^2 w}{\partial t \partial x}] \tag{7}$$

Now for the model of the form [M]{X}+[K]{X} = F we have ignored the damping effect and also we have assumed mass matrix as the lumped mass matrix. For the lumped mass matrix, the inertial effects in transverse as well as the rotary degree of freedom have been considered. Thus, mass matrix is as follows:

$$[M] = \frac{\rho_p A_p l}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & l^2/12 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & l^2/12 \end{bmatrix} \tag{8}$$

In equation (8), l is the length of one beam element. To obtain the stiffness matrix, we have considered Equation (4) and have written the weight function, v as ϕ_i and the dependent variable, w as ϕ_j where ϕ_i and ϕ_j are the shape functions for the beam element and w_j are the various degrees of freedom of a beam element. Thus, $[K_{ij}]$ matrix can be obtained from the following expression:

$$K_{ij} = \int_{x_e}^{x_{e+1}} \frac{\partial^2 \phi_i}{\partial x^2}\frac{\partial^2 \phi_j}{\partial x^2} E_p I_p w_j + \int_{x_e}^{x_{e+1}} \phi_i \frac{\partial^2 \phi_j}{\partial x^2}$$

$$(p\pi {R_m}^2 + \rho_f A_f V^2)w_j \tag{9}$$

If h_e is the length of single beam element, then shape functions of the beam element are as follows:

$$\begin{aligned} \phi_1 &= 1 - 3\left(\frac{\bar{x}}{h_e^2}\right) + 2\left(\frac{\bar{x}}{h_e^2}\right)^3 \\ \phi_2 &= -\bar{x}\left(1 - \frac{\bar{x}}{h_e}\right)^2 \\ \phi_3 &= 3\left(\frac{\bar{x}}{h_e^2}\right)^2 - 2\left(\frac{\bar{x}}{h_e^2}\right)^3 \\ \phi_4 &= -\bar{x}\left(\left(\frac{\bar{x}}{h_e^2}\right)^2 - \left(\frac{\bar{x}}{h_e^2}\right)\right) \end{aligned} \tag{10}$$

The stiffness matrix thus obtained is:

$$[K] = \frac{EI}{l^3}\begin{bmatrix} 12 & -6h_e & -12 & -6h_e \\ -6h_e & 4h_e^2 & 6h_e & 2h_e^2 \\ -12 & 6h_e & 12 & 6h_e \\ -6h_e & 2h_e^2 & 6h_e & 4h_e^2 \end{bmatrix} +$$

$$\frac{p\pi {R_m}^2 + \rho_f A_f V^2}{30h_e}\begin{bmatrix} -\frac{6}{5h_e} & \frac{11}{10} & \frac{6}{5h_e} & \frac{1}{10} \\ \frac{1}{10} & \frac{-2h_e}{15} & -\frac{1}{10} & \frac{h_e}{30} \\ \frac{6}{5h_e} & -\frac{1}{10} & -\frac{6}{5h_e} & -\frac{11}{10} \\ \frac{1}{10} & \frac{h_e}{30} & -\frac{1}{10} & \frac{-2h_e}{15} \end{bmatrix} \tag{11}$$

4. DYNAMIC MODELLING OF MANIPULATOR ARM

Three bellows have been fixed onto a plate to form a bending section as explained in Fig. 2. Two such bending sections together form the manipulator arm of the Robotino XT. Thus, by putting suitable initial and boundary conditions, the dynamic model can be modeled as shown in Fig. 3. The various parameters considered for simulation are shown in Table 1.

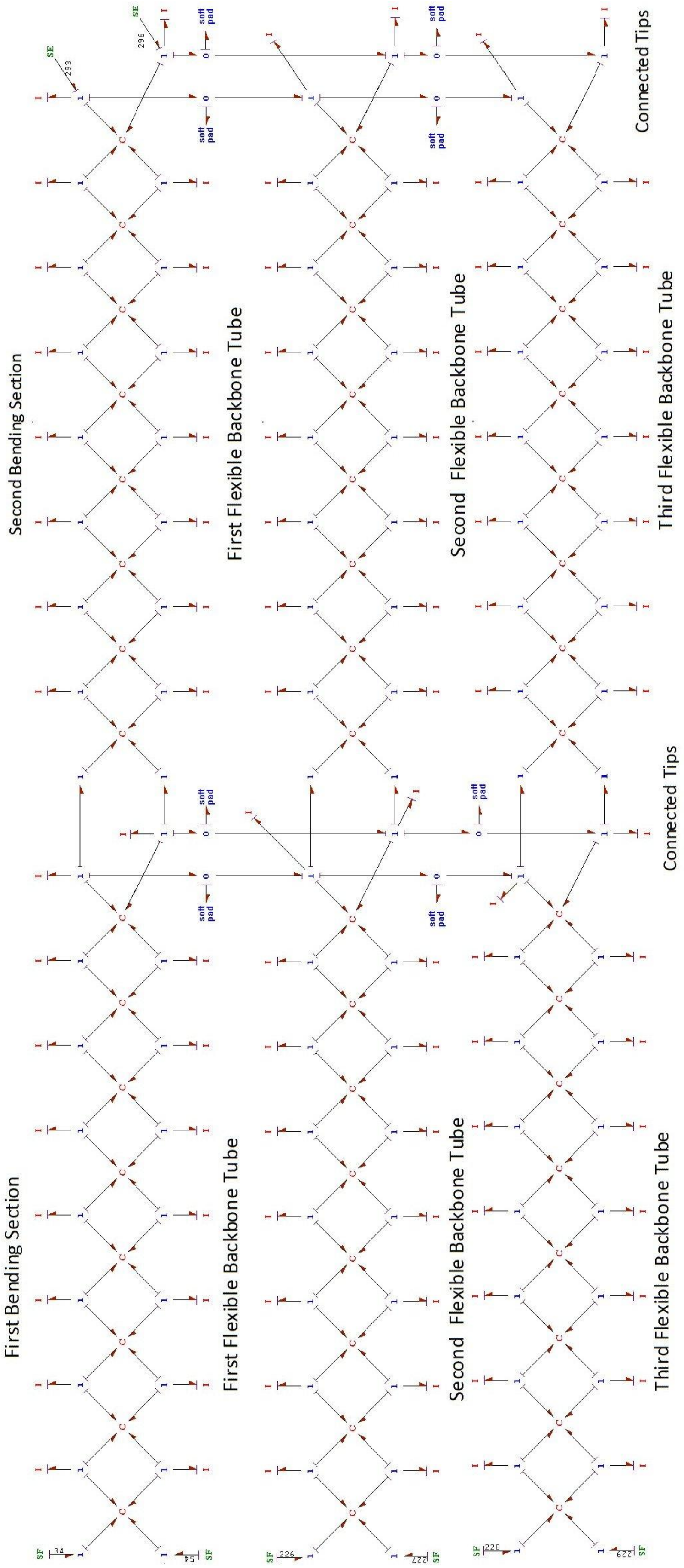


Fig. 3: Bond Graph model of the Manipulator Arm

Table 1: Parameters for modeling of manipulator arm

Parameters	**Symbol**	**Value**	**Units**
Young's Modulus of Elasticity of equivalent pipe	E	1,55,000	N/m^2
Mean Radius of the equivalent pipe	R_M	0.0341	m
Wall thickness of the equivalent pipe	SP	0.0007	m
Mass per unit length	M	1.643	Kg/m
Density of fluid	ρ_f	1.225	Kg/m^3
Length of equivalent pipe	L	0.1014	m
Pressure in the first bellow of first bending section	PA1	500	N/m^2
Velocity of flow in the first bellow of first bending section	V	6	m/sec
Transverse force acting on the tip of first bending section	F_{verti1}	4	N
Pressure of the fluid in third bellow of first bending section	PA2	600	N/m^2
Velocity of the fluid in second bellow of first bending section	V_2	6.4	m/sec
Pressure of the fluid in third bellow of first bending section	PA3	800	N/m^2
Velocity of fluid in third bellow of first bending section	V_3	6.8	m/sec
Pressure of fluid in first bellow of second bending section	PB1	500	N/m^2
Velocity of fluid in first bellow of second bending section	V_4	6	m/sec
Pressure of the fluid in the second bellow of second bending section	PB2	700	N/m^2
Velocity of the fluid in second bellow of second bending section	V_5	7.0	m/sec
Pressure of the fluid in the second bellow of second bending section	PB3	900	N/m^2
Velocity of the fluid in third bellow of second bending section	V_6	7.4	m/sec
Transverse Force acting on the tip of the second bending section	F_{verti2}	-4	N

Thus the bond graph consists of six Rayleigh Beam elements. In the model, the end tips of bellows of first bending section are connected to starting points of the corresponding bellows of second bending section. Also, the bellows of each section are connected to each other.

Under the influence of parameters stated above, the behavior of the tip is shown in Figure 4. Figure 5 (a) - (c) shows the animation frames for the motion of the six flexible backbones.

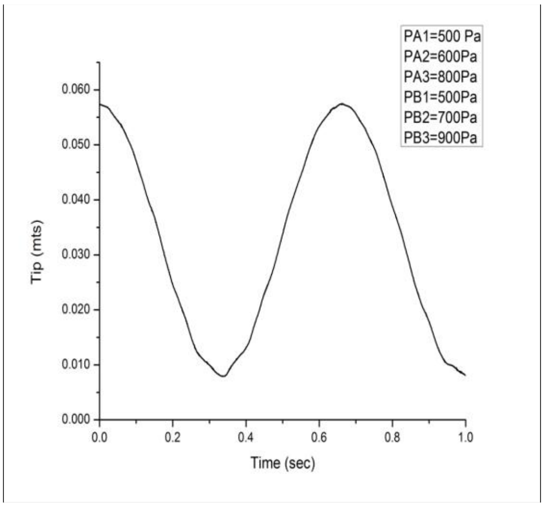
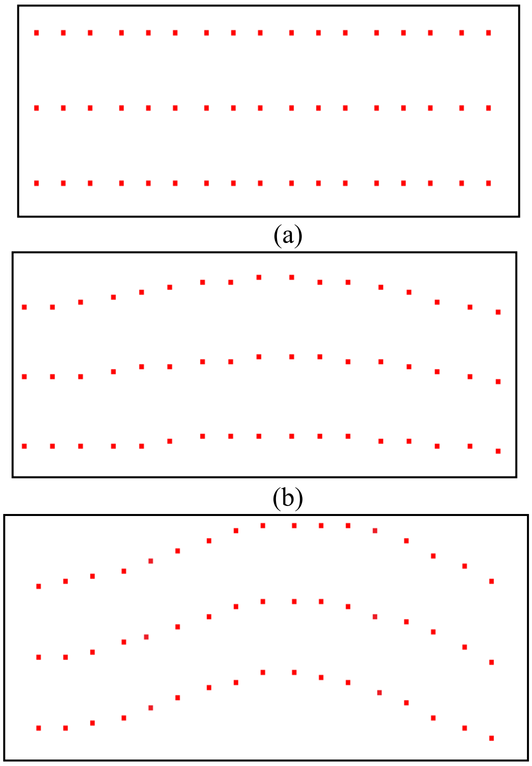


Figure 4: Tip response of manipulator arm

In the above graph, PA1, PA2, PA3 are the pressures of the fluid in first bending section and PB1, PB2, PB3 are the pressures of fluid in the bellows of second bending section.

Also, the following frames describe the positions of the various nodes of the modeled beam elements.


(a)

(b)

(c)

Figure. 5: (a), (b) and (c): Animation frames for motion of six backbones

In Fig. 5, nodes of six Rayleigh beams have been shown. Each beam is divided into nine elements, thus having nine nodes into it. Here only one such cycle of vibration has been shown. One cycle is assumed to be of 0.60 seconds. Simulation results have been carried out for 1s.

5. CONCLUSION

The compressed air is flowing in each bellow of both the bending sections with different velocities and different pressures, giving rise to different stiffness. This different stiffness of bellows give rise to a powerful movement. To obtain the dynamic model of the manipulator arm, geometry of bellow is studied in detail and is found that the differential equation governing the bending vibrations of the bellow can be obtained from the general differential equation for bending vibrations of a pipe conveying fluid. To simulate the bellow behavior, finite element representation of differential equation is done and stiffness matrix can be derived. This FE approach can be linked to bond graph modeling approach to obtain the dynamic model of manipulator arm as a whole. Once the dynamic model of this class of bionic manipulator has been synthesized, it can be used to design an adequate control for the whole Bionic Handling Assistant.

REFERENCES

Andersson, W. F., 1964. Analysis of Stresses in Bellows – Part I Design Criteria and Test Results, *Atomic International, NAA-SR-4527, United States Atomic Energy Commission*.

Andersson, W. F., 1965, Analysis of Stresses in Bellows – Part II Mathematical, *Atomic International, NAA-SR-4527, United States Atomic Energy Commission*.Siciliano B. and Khatib W., 2009. *Springer Handbook of Robotics*, ISBN 978-3-5423957-4, Springer.

Broman, G.I., JoÈnsson, A.P., Hermann, M.P., 2000. Determining dynamic characteristics of bellows by manipulated beam Finite elements of commercial software, *International Journal of Pressure Vessels and Piping* 77, pp 445-453

EJMA: *Standards of the Expansion Joint Manufacturers Association*, seventh edition, 1998.

Morishita, M., Ikahata, N., Kitamura, S., 1989. Dynamic Analysis Methods of Bellows Including Fluid-Structure Interaction, in Metallic Bellows and Expansion Joints, The 1989 *ASME Pressure Vessels and Piping Conference - JSME* Co-Sponsorship, Honolulu, Hawaii, July 23-27, 1989, pp 149-158.

Osweiller, F., 1989. Design of an Expansion Joint by a Finite Element Program – Comparison with the EJMA Standards, in Metallic Bellows and Expansion Joints – 1989, pp 87-94, The 1989 *ASME Pressure Vessels and Piping Conference -* JSME Co-Sponsorship, Honolulu, Hawaii, July 23-27.

Ting-Xin, L., Bing-Liang, G., Tian-Xiang, L., 1990. Natural Frequencies of U-Shaped Bellows, *International Journal of Pressure Vessels and Piping 42*, pp 61-74.

Trivedi, D., Rahn, C. D., Kier, W. M., and Walker, I. D., 2008. Soft robotics: Biological inspiration, state of the art, and future research, *Applied Bionics and Biomechanics,* 5(3), pp. 99. 117

www.festo.com/cms/en_corp/9617.htm, 2011

www.festo.com/cms/en_corp/11367.htm, 2011

BIOGRAPHY

Sonam Behl:
She is M. Tech CAD, CAM & Robotics student at Mechanical and Industrial Engineering Department, IIT Roorkee. Her area of interest is bionic manipulators modeling and control.

Coralie Escande:
She is PhD student at Polytech'Lille, Villeneuve d'Ascq, France. Her area of interest is bionic manipulators.

Pushparaj Mani Pathak:
He is Assistant Professor at Mechanical and Industrial Engineering Department, IIT Roorkee. His area of interest is space robotics, walking robot, bionic manipulators modeling and control.

Rochdi Merzouki:
He is Professor at Polytech'Lille, Villeneuve d'Ascq, France. His area of interest is Fault diagnosis and isolation, bionic manipulators, intelligent vehicles.

B.K. Mishra:
He is Professor at Mechanical and Industrial Engineering Department, IIT Roorkee. His area of interest is Finite Element Modeling and control of flexible structures.

Autohrs’ Index