Volume 12, Number 5
October 2011 (click for next issue)
Submission Policy, Template and Copyright Form

ISSN 1473-804x Online
ISSN 1473-8031 Print

International Journal of Simulation
Systems, Science & Technology

A publication of the United Kingdom Simulation Society

Editorial Policy and Editorial Board | Submission Policy and Author's Instructions

Nanosimulator for Analysis of Mosfet at Nanoscale

Pragya Kushwaha[1] , Amit Chaudhry[2] , Garima joshi[3]

[1] University Institute of Engineering and Technology, Panjab University, Chandigarh, India.pragya189@gmail.com

[2,3]University Institute of Engineering and Technology, Panjab University,Chandigarh,India. amit_chaudhry01@yahoo.com, joshi_garima5@yahoo.com

***Abstract* - At nanoscale dimensions (below 100nm) the behavioral characteristics of MOSFET change due to quantum mechanical effect and deviate from the normal equations. In this paper we have proposed a simple Graphical user interface model-NANO SIMULATOR . This simulator is based upon modified equations to predict the curves of nanoscale MOSFETs. It has been developed to study the Current-Voltage characteristics and the model parameters i.e. threshold voltage,flat band voltage, mobility of nanoscale MOSFETs specifically at 90nm, 65nm, 45nm,32nm technology nodes(or CMOS Process). This GUI based NANO SIMULATOR also helps in studying various types of leakage currents at 90nm, 65nm, 45nm, 32nm technology nodes.**

Keywords - I-V Curve, Leakage Current Curve, Pushbutton, popup menu, MatLab GUI

I. NOMENCLATURE

Vgs- Gate to Source voltage
Vds- Drain to Source voltage
Vth- Threshold voltage
Vfb- Flat band voltage
Jtn- Direct gate tunneling in n-channel MOSFET current density
Jsdn- Source Drain Extension tunneling current density
Jtpoly-Leakage current density due to poly gate
Jfn- Fowler Nordheim Tunneling
Eox- Dielectric constant of gate oxide
EOT- Equivalent Oxide Thickness
Cox - Oxide capacitance
μ - Mobility

II. INTRODUCTION

As the transistor gate length drops to 45 nm and below and the gate oxide thickness drops to 1 nm physical limitations, such as off-state leakage current and reduction in drive currents, make geometric scaling an increasingly challenging task. Moreover, a combined study of I-V Curves on the basis of channel length scaling along with the leakage currents on the basis of gate oxide thickness scaling at various technology nodes by a GUI-SIMULATOR has not been reported yet in the existing literature. In this paper ,we present our GUI-SIMULATOR to study MOSFET at 90nm,65nm and 45nm technology nodes. This simulator is built upon MATLAB-*Graphical user interface* and the curves which this simulator will provide are also calculated by MATLAB Tool.

MatLab software environment with its own Graphical User Interface (GUI). The GUI is used to analyze the I-V characteristics of MOSFET at nanoscale and displays the curves and parameters such that a user can visualize the voltage and current variations as well as different type of leakage currents. The user interface also plays a vital role in controlling the application's performance to suit the user's requirements. The objective of this work is to make a easy understanding of MOSFET at nanoscale using a simple simulator with simple MOSFET equations. This simulator has the ability to take the input from user, and generates the characteristic curve and corresponding parameters. This paper is organized as follows. In section III, NANOSIMULATOR's structure is shown by flow diagram in fig.1. Section IV describes nanoscale MOSFET modeling quations. Section V, explains how we can install the simulator. Section VI, explains modeling results for leakage currents. Section VII, describes modeling results for input-ouput characterstics. A summary and future work is presented in section VIII.

III. SOFTWARE STRUCTURE

The proposed simulator is built in MatLab software environment with its own GUI. MatLab code here used is in 'm-file' at backend because it is more flexible and have fewer limitations. User end code is in ".pfile", so that our original file can be saved from unwanted disturbances. The GUI is a user friendly interface that can provide a user with the access to perform I-V analysis of MOSFET. This simulator can perform the following processing:

- Id-Vds Curve
- Id-Vgs Curve
- Jtn-Vgs Curve
- Jsdn-Vgs Curve
- Jtpoly-Vgs Curve
- Jfn-Vgs Curve
- Jfn-Eox Curve
- Jtn-EOT Curve
-

Fig. 1 shows the flowchart of the simulator , where the code has been written using MatLab code to obtain the results on the GUI.

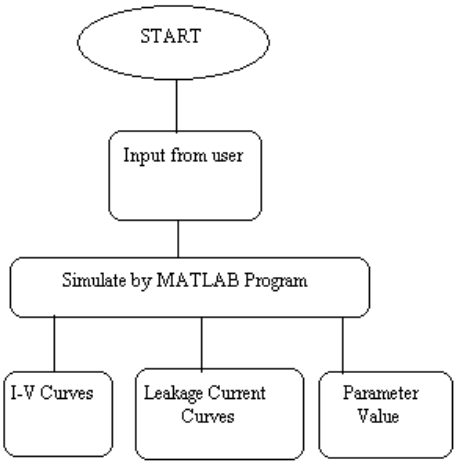


Fig. 1 Flowchart of the proposed simulator

IV. NANOSCALE MOSFET MODEL

Here we are giving analytical MOSFET's I-V model for circuit simulator. This approach is necessary for circuit design. Analytical model is of two types[4],one is physically based,the other is continuity between different operation regions.Since Quantam Mechanical Effects have significant influences on the threshold voltages[2],on gate capacitance attenuation[3] due to finite inversion layer thickness.

Modified equation for Threshold voltage:

Initially the threshold voltage will be equation 1

V_th=V_fb+phi_s_qm+sqrt(2*q*epsilon_si*N_a*phi_s_qm) /C_ox; (1)

Threshold voltage in linear region due to QME is given by equation 2

$V\_th\_qm_{linear}$=V_fb+phi_s_qm+phi_c_{linear}+sqrt(2*q*epsilon_si*N_a*(phi_c_{linear}+phi_s_qm))/C_ox; (2)

Threshold voltage in saturation region due to QME is given by equation3

V_th_qm_sat= V_fb+phi_s_qm+phi_c_sat+sqrt(2*q*epsilon_si*N_a*(phi_c_sat+phi_s_qm))/C_ox; (3)

Where

alpha_t= ((3*m_e*q^2*N_a*d_cl)/(2*epsilon_si*h_^2))^(1/3);

phi_s = (q*N_a*(d_cl)^2)/(2*epsilon_si);

phi_s_qm = (q*N_a*((d_cl+9/(4*alpha_t))^2))/(2*epsilon_si);

V_fb = -0.56-phi_b; Flat Band Voltage

phi_c_{linear} = (V_dsat)/2;

phi_c_sat = V_dsat/2;

Modified equation for mobility:

Due to QME and Short Channel Effects mobility equations are different from classical mobility.

Mobility in linear region is given by equation 4

u_eff = u_o./(1+theta*(V_gs-$V\_th\_qm_{linear}$)); (4)

Mobility in saturation region is given by equation 5

u_eff_sat = u_o./(1+theta*(V_gs-V_th_qm_sat)); (5)

where

u_o = 0.06; low field surface electron mobility

E_cro_{linear} = v_sat./u_o; Vertical electric field component in linear region

E_cro_sat = v_sat./u_eff_sat; Vertical electric field component in saturation region

theta = u_o/(2*t_ox*v_norm); %constant

v_norm = 2.2e7 ;v_sat = 1e5; %unit: m/s

v_sat is saturation velocity

Modified equation for Bulk Deplition region charge:

Without QME and SCE the value of Q_{BO} is given by equation6

Q_dep = q*N_a*d_cl (6)

Without QME and SCE the value of Q_{BO} is given by equation7

Q_bo_qm = q*N_a*(d_cl+(9/(4*alpha_t))) (7)

Where

d_cl = sqrt((2*epsilon_si*2*phi_b)/(q*N_a))

alpha_t=((3*m_e*q^2*N_a*d_cl)/(2*epsilon_si*h_^2))^(1/3)

h = 6.626e-34; %unit: J.s

h_ = h/(2*pi);

epsilon_o = 8.854e-12; %unit: F/m

epsilon_si = 11.80*epsilon_o;

epsilon_sio2 = 3.90*epsilon_o;

Modified equation for I_{DS} and V_{dsat}:

V_dsat = (L*E_cro_{linear}*((sqrt(1+(2*(V_gs-V_th))/((L*E_cro_{linear})*eta)))-1));

I_dst = (W*u_eff*C_t/L).*(1./I_2).*I_6;

I_dsat = ((W*u_eff2*C_t/L).*(1./Iss_1))*Is_6;

Where

eta = 1-((Q_bo_qm)/(2*phi_s_qm*C_ox))

C_t = (1/C_ox+1/C_inv)^(-1)

C_inv is capacitance of inversion layer

Modified equation for Jtn and V_{GS}:

A = exp(-gamma*sqrt(q*E_b));

B = exp(q*(phi_s-phi_b-E_g/2)/(k*T));

C = 1-exp(-q*V_gs/(k*T));

E = 1.+(gamma*k*T)./(2.*sqrt(q*E_b));

F = 4*pi*q*m_e*((k*T)^2)/(h^3);

J_t = (F.*E.*C.*B.*A)*1e-4;

Where

m_o =9.11e-31 %unit: Kg

m_ox = 0.32*m_o;

m_e = 0.19*m_o;

h = 6.63e-34 ; %unit: J-s

q = 1.6e-19 ; %unit: C

phi_b = ((k*T)/q)*log(N_a/n_i); %eV

gamma = (4*pi*t_ox*1e-9*sqrt(2*m_ox))/h;

N_a=4e24 ; %unit: m-3

n_i= 1.45e16 ; %unit: m-3

k = 1.38e-23 ; %unit: J/K

T = 300 ; %unit: K

Modified equation for Jtpoly and V_{GS}:

A = exp(-gamma.*sqrt(q.*E_b));

B = exp((phi_s-phi_b-(E_g*q)/2)/(k*T));

C = 1-exp(-V_gs./(k*T));

E = 1+(gamma*k*T)./(2.*sqrt(q.*E_b));

G = (k*T)^2;

F = 4*pi*m_e*q^2/(h^3);

J_tpoly = F*G.*E.*C.*B.*A;

Modified equation for Jfn and V_{GS}:

A = (q^3.*m_n.*El_ox.^2)

B = 16*pi^2*m_o*h_*phi_ox

C = 4.*sqrt(2*m_e).*(phi_ox^(3/2))

D = 3.*h_.*q.*El_ox

J_fn = ((A./B).*exp(-C./D)).*10^-4;

Where

phi_ox= q*3.1-del_phi;

m_o =9.11e-31; %unit: Kg

m_e = 0.4*m_o;

m_n= 0.19*m_o;

V_ox =(V_gs-V_fb)-phi_s;

El_ox= V_ox/(T_ox*1e-9);

Eg = 1.12; %unit: eV

V. HOW TO INSTAL NANOSIMULATOR

The display results are obtained using the proposed MatLab GUI. Fig. 3 is the MatLab GUI window that shows the input from the user, analysis of characteristic curves, parameter analysis. Copy two files from the given CD namely NanoSim.p and NanoSim.fig .Paste these files in a particular directory and set the path of the MatLabDirectory. Type "run NanoSim" on Command Window of MATLAB Press 'Enter' and the simulator will be displayed as shown.

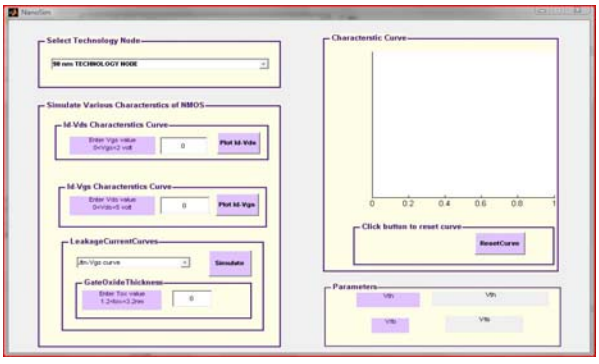


Fig. 2 Nano Simulator

VI .MODELING RESULTS FOR LEAKAGE CURRENTS

An analytical model for gate direct tunneling current density is modeled using MATLAB codes, the results drawn are being presented here.

A)Gate Direct Tunneling Current Density in n-MOSFET

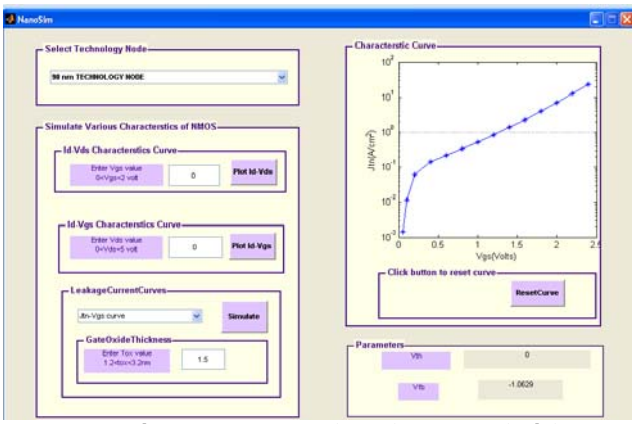


Fig 3.Tunneling Current Density Jt(A/cm2) versus Vgs (Volts) in n-MOSFET at Various Oxide Thicknesses(i.e. tox=1.5nm)[6]

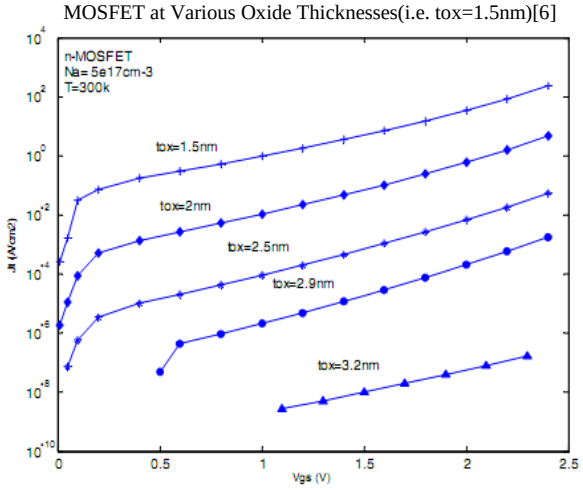


Fig 4: Plot of fig3 is matching with ref[6]

B)Edge Direct Tunneling Current Density

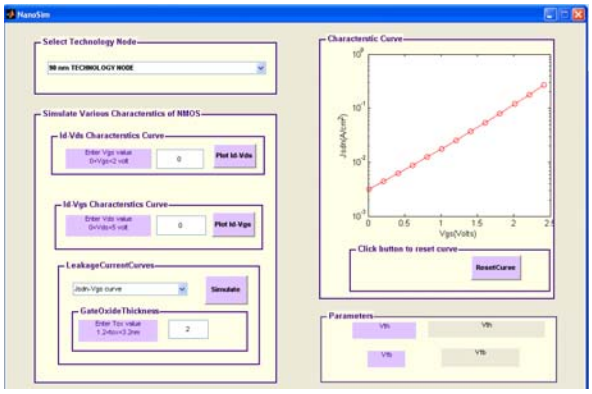


Fig 5.SDE Tunneling Currents in n-MOSFET at 2nm Oxide Thickness[6]

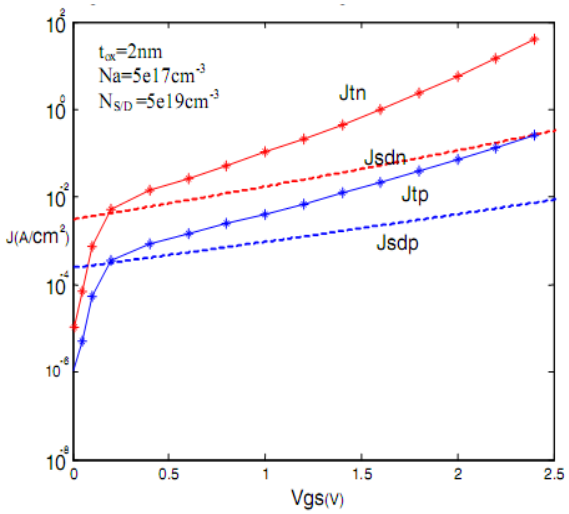


Fig 6: Plot of fig5 is matching with ref[6]

VII .MODELING RESULTS FOR I/O CHARACTERSTICS

Vth, Cox, μ are the parameters which undergo a change due to Short Channel Effects and so the conventional equations are invalid at Nanoscale and we include Quantum Mechanical effects[1]. And by the addition of QME effects we got such input-output characterstic curves as shown in Fig7&Fig8 at 90nm technology node.

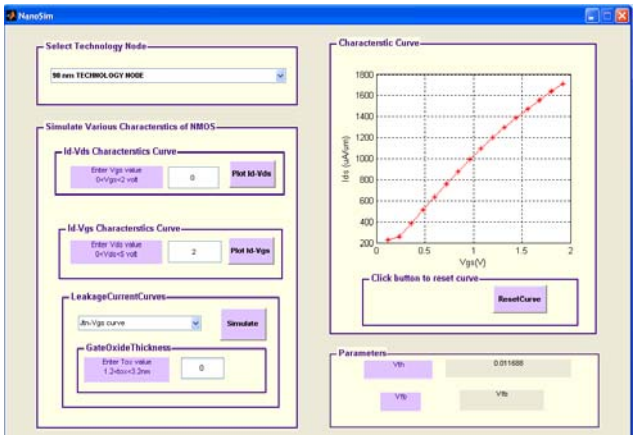


Fig7: Id-Vgs Curve

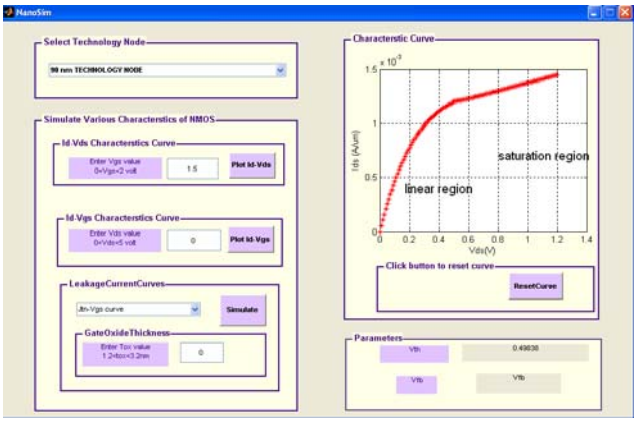


Fig8: Id-Vds Curve at Vgs=1.5v [5]

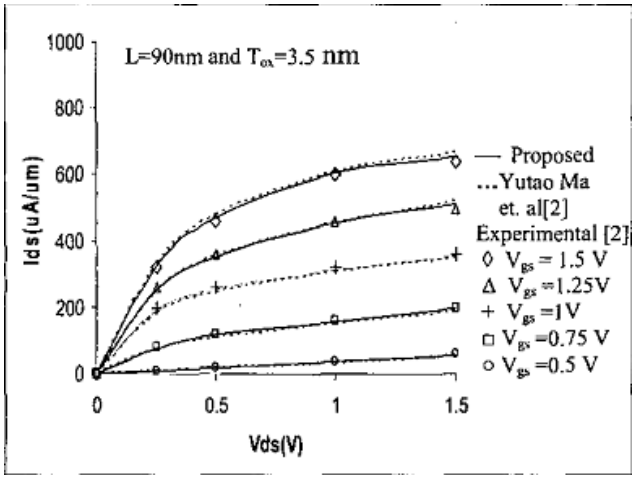


Fig9: : Plot of fig8 is matching with ref[5]

VIII. CONCLUSION

A simulator for Leakage Current detection and classification as well as I-V characterstics using MatLab program has been developed. At all technology nodes I-V Curves were successfully visualized. The Leakage Current detection, Input-Output characterstic results are displayed to a user via MatLab graphical user interface (GUI). The significant features of the developed GUI are the ability to graphically display the results by taking inputs from user side,hence this tool is user friendly. At backend we have done calculation by considering short channel effects and QME effects hence the developed simulator provides accurate, efficient and reliable curves and parameter values to study. In future this simulator will be available to our users online with some special functions like online help and also more curves can be studied by this (i.e.Doping and Temperature effects on nanoscale mosfet).

IX. REFERENCES

[1] J. Mracek, "Understanding Quantum Mechanical Tunneling in MOSFET Gate Oxides", Journal of Physics 334, Article no. 0017, Dec 2007.

[2] Tomasz Janik, Bogdan Majkusiak. "Influence of carrier energy quantization on threshold voltage of metal-oxide-semiconductor transistor", J.Appl.Phys. ~0175, nolo, pp5 186-5 190, 1994.

[3] R. Versari, B. Ricco . "Scaling of maximum capacitance of MOSFET with ultra-thin oxide." Electronics Letters v34, n22, p 2175-2176, 1998.

[4] Yutao Ma, Litian Liu, Lilin Tian, Zhiping Yu* and Zhijian Li "Comprehensive Analytical Charge Control and I-V Model of Modern MOSFET's by Fully Comprising Quantum Mechanical Effects" 0-7803-6279-9/00/$10.00 0 2000 IEEE

[5] Manisha Pattanaik and Swapna Banerjee"A New Approach to Model the I-V Characteristics for Nanoscale MOSFETs" 0-7803-7765-6/03/$17.00 02003 IEEE.

[6] Garima Joshi , D.N. Singh, Member IEEE, Sharmelee Thangjam"Effect of Temperature Variation on Gate TunnelingCurrents in Nanoscale MOSFETs" Nanotechnology, 2008. NANO '08. 8th IEEE Conference on.

BIOGRAPHIES

Er.Pragya Kushwaha was born in1988. She is a student of M.Tech Microelectronics U.I.E.T(P.U.) Chandigarh, India. She obtained her Bachelor Degree in Engineering (Electronics and Communication) from R.K.G.I.T. Ghaziabad, India, in 2009. Currently, her research interest includes low power and high speed memories at nanoscale.

Dr. Amit Chaudhry was born in India, in 1976. He completed his Ph.D (Microelectronics) in 2010 from Panjab University, Chandigarh, India. He joined Panjab University, University Centre for instrumentation and microelectronics October, 2002. He was responsible for teaching and research in VLSI and microelectronics to post graduate students. His research areas include device modeling for sub 100nm MOSFETS. He is a life member of various societies in the area of microelectronics. Currently he is a Senior Assistant Professor, University Institute of Engineering and Technology, Panjab University, Chandigarh. He has more than 18 publications in national and international Journals/Conference proceedings.

Garima Joshi was born in India, in 1981. She is M.E (Electronics and Communication Engineering) from UIET, Panjab University, Chandigarh, India. Her area of research includes modeling and simulation of Nanoscale MOSFETS. Currently she is working as Assistant Professor (Electronics and Communication Engineering) in UIET, Panjab University, Chandigarh, India. She has 5 publications in international Journals/Conference proceedings.

Bond Graph Based Model for Robust Fault Diagnosis

Rafika El Harabi
Unité de Recherche Modélisation, Analyse et Commande des Systèmes (MACS),
ENIG, Rue Omar Ibn Elkhatab, 6029, Tunisie.
e-mail: rafikaharabi@yahoo.fr

Mohamed Naceur Abdelkrim
Unité de Recherche Modélisation, Analyse et Commande des Systèmes (MACS),
ENIG, Rue Omar Ibn Elkhatab, 6029, Tunisie.
e-mail: naceur.adelkarim@enig.rnu.tn

***Abstract*— In this paper, robust Fault Detection and Isolation (FDI) design in nonlinear uncertain dynamic system, with chemical and thermodynamic phenomenon, is addressed. The methodology using a Bond Graph (BG) representation in linear fractional transformation (LFT) form is shown to be a valuable tool for developing dynamic threshold generators and achieving robustness against model uncertainty in combination with sensitivity to faults. The proposed FDI method is illustrated through an equilibrated reaction occurred in a continuous reactor coupled with a heat exchanger. Simulations are given to support the theoretical development and demonstrate the potential of the developed procedure.**

Keywords - bond graph, chemical reactors, FDI design, dynamic threshold generators

I. INTRODUCTION

Due to the growing complexity of automatic control systems, there is an increasing demand for fail-safe operation, fault diagnosis (FD) and fault tolerance (FT). The early detection of system malfunctions and faults as well as the isolation of their origin have become an important issue in advanced control system design. Much attention has been paid to the design of robust fault detection and isolation systems (see for instance [1]).

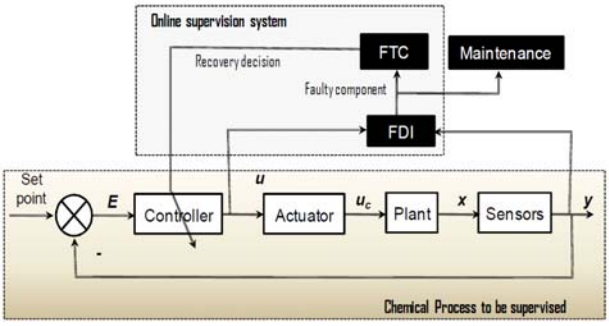


Figure 1. Supervision scheme in process engineering.

Supervision of chemical reactors is a difficult task (as shown in Figure 1). This is due to several factors, such as the transient operation conditions, the various uses of these reactors and the evolution of the state variables which is nonlinear. The evolution of some parameters (the activation energy, pre-exponential factor, specific enthalpy) is nonstationary, which changes according to the condition variation inside the reactor [2]. It is this fact that has motivated our research in this paper.

Furthermore, due to the strong nonlinearities and parameters uncertainties in the chemical systems, their modeling is often complex and therefore less developed in the literature. The graphical modeling such as the bond graph tool becomes significant in this case, because it is appropriate for multiphysics modeling of complex and uncertain systems, as it is given in [3]. However, this tool can be used for residuals generation and monitorability analysis of uncertain systems [4].

The aim of presented paper is the design and analysis of a robust diagnosis scheme for nonlinear chemical processes taking into account the parameter uncertainties, described by coupled pseudo Bond Graph models using LFT form, when the secondary events (secondary reaction, hazard event of thermal runaway…ect) appear in chemical reaction. Thus, due to the energetic and multi physical properties of the Bond Graph, the whole of nonlinear model, structural analysis, residual with adaptive thresholds generations, and residual sensitivity analysis, can be synthesized using only one tool.

Section 2 gives a brief review of based element of coupled Bond Graph. The third section presents uncertain bond graph modeling and linear fractional transformations using in the chemical processes. In the fourth section, the bond graph LFT modeling of the chemical reaction in presence of parameter uncertainties is given. This section describes also the robust ARRs generation algorithm and the residual analysis. The developed methodology is applied for pseudo bond graph model based FDI of a continuous reactor coupled with a heat exchanger in section five. Finally some conclusions are drawn.

II. BASIC ELEMENT OF COUPLED BOND GRAPH

Bond graph models are network type models which are composed of multiport related by power bonds representing the (acausal) identity between pair of conjugated variables (named effort and flow) whose product is the instantaneous

energy flow between the multiport elements. The multiport elements represent storage (**C-element**) (as compliance for instance or volume), inertia (**I-element**) (electrical inductance and mechanical inertia), energy dissipation (**R-element**) (electrical, mechanical or thermal friction), balance and continuity equations (the **0- and 1-junctions**) or inter-domain coupling (the **TF transformer** and **GY gyrator** elements). Finally to reproduce the architecture of the global system to be modeled, bond graph elements (R, C, I,..) are interconnected by a "0" junctions when they have a common effort and by "1" junction if their flow is the same.

In process engineering processes, several phenomena (chemical, thermal and fluidic) are coupled. In addition to matter transformation phenomena, chemical and electrochemical processes involve additional complexity in the modeling task, since the mass that flows through the process carries the internal energy which is stored in it, and which is thus transported from one location to another in a non-dissipative fashion. Power variables are thus in vectorial form:

$$E=\left[e_h\ e_t\ e_c\right]^t,\ \mathrm{F}=\left[f_h\ f_t\ f_c\right]^t \qquad (1)$$

where e_h, e_t and e_c represent respectively the thermal effort (specific enthalpy h or the temperature T), the hydraulic effort (the pressure P), and the chemical effort (the chemical potential μ, chemical affinity A or the concentration c). f_h, f_t and f_c represent respectively the thermal (or entropy) flow (by conduction $\dot{Q}$ or by convection $\dot{H}$ i.e. enthalpy flow), hydraulic flow (mass flow $\dot{m}$ or volume flow $\dot{V}$) and chemical flow (molar flow $\dot{n}$).

Consider a thermofluid process (Figure 2 (a)) which consists of a pump (considered as a flow source) fulfilling a heated tank where a bottom pressure is measured by the sensor P_m, and the average temperature of the fluid is indicated by T_m. The coupled bond graph model in integral causality is given by Figure 2 (b). The two ports C_{ht} represents the coupled thermal and hydraulic energy of the stored fluid (considered here in under saturated state) is decoupled into thermal and hydraulic capacity C_t, and C_h. $Sf:\dot{Q}_{in}$, $Sf:\dot{m}_{in}$, and $Se:T_{in}$ represent, respectively, thermal flow source, inlet mass flow, and the temperature of the incoming fluid (considered constant).
The coupling is modelled by the fictive R_c element in the thermal bond.

Another complexity can be added taking into account transformation of matter in chemical phenomena. The corresponding bond graph model is given by Figure 3. The mixture of mass flow $\dot{m}_{in}$ is considered multicomponent with n species. The n transformers with $1/Mi\ \left[Kg.mole^{-1}\right]$ as modulus used to transform massic flow $\dot{m}_{in}$ to molar flow $\dot{n}_{i,in}$ of i^{th} specie:

$$\dot{n}_{i,in}=\frac{\dot{m}_{in}}{M_i} \qquad (2)$$

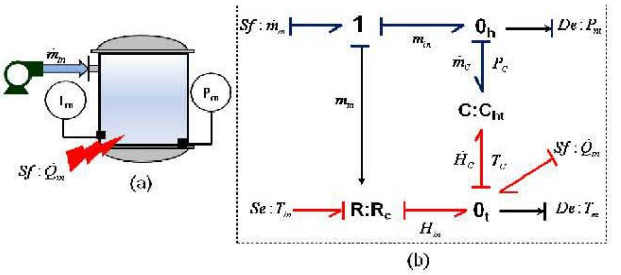


Figure 2. Heated tank (a) and its BG model in integral causality (b)

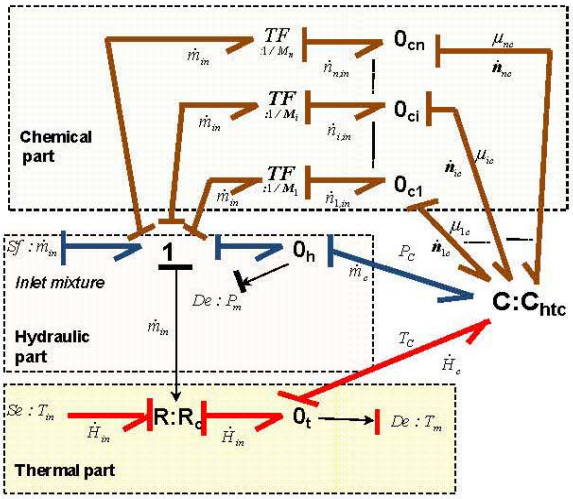


Figure 3. Bond graph model with three coupled energies.

The state equation form $x=f(x,u)$ well suited for control analysis, can be systematically deduced from a bond graph in a linear or non linear form. The input vector u is represented in a bond graph by the sources (Se and Sf), the measured variables are effort and flow detectors. The state vector is composed by the energy variables stored by C (general displacement) and I elements (impulse). The state vector does not appear on the Bond Graph, but only its derivative: The dimension of the state vector is equal to the number of C and I elements in integral causality. In the given dynamic model, there are $n+2$ state variables: $x=\left[m_c\ H_c\ n_{cn}..n_{cn}\right]$. They represent storage of number of mole for $n+2$ species, total mass m_c and internal energy of the mixture H_c.

III. Uncertain Coupled Bond Graph

A. *Uncertain Bond Graph Interest*

Various bond graph based qualitative and quantitative [5], FDI approaches have been developed to detect and

isolate faults in single or piece-wise single energy domains, but none deal with FDI of coupled (energetic and transformation phenomena) nonlinear systems.

Among recent works that deal with parameter uncertainties modeling using bond graph approach, in [6], the authors proposed to construct in a systematic manner a bond graph from another bond graph using standard interconnection form, which is called the associated incremental bond graph (IBG).

In [7] and [8], authors proposed two methods for modeling uncertainties by using bond graph approach, applied on Electromechanical and thermodynamic systems (vehicle, test bench and steam generator) . The first method is based on describing parameter uncertainties as bond graph elements, and the second method introduces the LFT form for uncertainties modeling. Here, this problem is addressed using the linear fractional transformation (LFT) paradigm.

After the pioneering work of Oster and Perelson, it has been mainly used for membrane processes some reaction processes and some electrochemical processes [9], [10]. Bond graph modeling has been used for hydraulic and thermal domain in chemical reactor but not for monitoring and observing kinetic and thermodynamic evolution of chemical mixture. Thus, uncertain bond graph modeling of chemical reaction are not treated until now in literature and diagnosis of chemical reaction is an open research work.

B. BG-LFT Form

The principle of the uncertainties representation using LFT consists in building the uncertain system in the form of a looping between the increased invariant system M, whose parameters are defined perfectly, and a block of uncertainty, noted Δ, gathering various uncertainties, Figure 4. Setting of LFT form requires that the system must be reachable and observable. These properties are necessary conditions for the monitoring ability of the system.

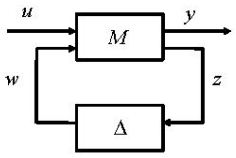


Figure 4. LFT representation.

The interconnection structure induces the following state equations:

$$\begin{cases} w = -\Delta\, z \\ \dot{x} = Ax + B_1 w + B_2 u \\ z = C_1 x + D_{11} w + D_{12} u \\ y = C_2 x + D_{21} w + D_{22} u \end{cases} \quad (3)$$

where $x \in \Re^n$ the state vector, $u \in \Re^m$ the inputs vector, $y \in \Re^p$ the outputs vector. $w \in \Re^l$ and $z \in \Re^l$ are ,respectively, the auxiliary input and output vectors. n, m, l et p are positive entireties. A, B_1, B_2, C_1, C_2, D_{11}, D_{12}, D_{21}, D_{22} are appropriate ranks matrices.

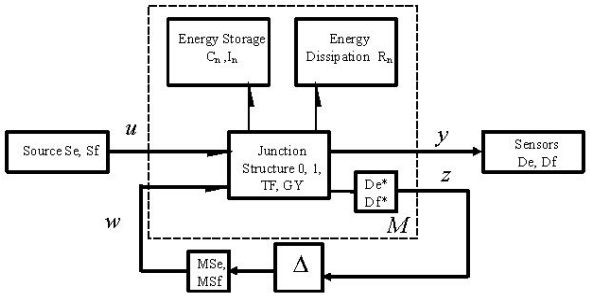


Figure 5. BG-LFT representation.

In LFT bond graph representation, parameters uncertainties are represented under multiplicative form at the level of bond graph component. The method consists in replacing each uncertain element by its BG-LFT. BG-LFT representation is shown in Figure 5.

The advantage of approach BG-LFT compared to an approach of LFT state is summarized in two points: complexity in the model construction and the uncertainties structure on the model [4].

Modeling of bond graph elements $i \in \{R, I, C, TF, GY\}$ in the LFT form consists in decoupling the nominal element $i_n \in \{R_n, I_n, C_n, TF_n, GY_n\}$ part from its uncertain part $\delta_i i_n \in \{\delta_R R_n, \delta_I I_n, \delta_C C_n, \delta_{TF} TF_n, \delta_{GY} GY_n\}$, with δ_i is a multiplicative uncertainty on the parameter i . In the combined BG-LFT representation, the parameter uncertainties are explicitly represented under multiplicative form for each bond graph element. The additive uncertainties of the parameters are related to their multiplicative values by the following relations: $\delta_i = \dfrac{\Delta i}{i_n}$, where Δi is the additive uncertainty values on the bond graph element i .

The principle of this modeling consists in representing the influence of the parameter uncertainty, by a fictive effort or flow input ($MSe : w_i$ or $MSf : w_i$), modulated by $\delta_i (i_n e_i)$ or $\delta_i (i_n f_i)$. Details on this modelling procedure are given in [8].

In chemical processes, to explain the modelling in LFT form using the bond graph, let us consider the Multiport □ , we know that the R-elements dissipate power and that this power comes out as heat. So including thermal effects, an R-element becomes an irreversible and power conserving structure. It is denoted as multiport □ (see Figure 6 (a)). So power can flow only as indicated by the half arrows, and not backwards. In other words it cannot become negative. So, when we are not interested in thermal effects, we speak of

R-elements and multiport-R, otherwise of multiport- □ . Regarding the multiport □ , it can have bonds with several strands as shown on Figure 6 (b). With multiport □ , irreversibility and energy conservation of multiport R are as follows: with several strands, only the sum of the non thermal bonds must be positive, but in single strands power can become negative as long as it is more positive in others. One can also say that power in the thermal bond must by always positive.

Figure 6. Element (a) and Multiport □

In a chemical reaction, the product of chemical affinity A by the global reaction rate J is a power. The thermal loss (transformation of chemical to thermal energy) is modeled by an active resistance Multiport □ (a resistance which generates entropy) [11]. The multiport absorbs chemical power $A \times J$ and product an equivalent quantity in thermal power $T \times S$. Thus, a RS-field is used as a link between the mass and energy parts of the reactor vessel subsystem. It is a two-port element connecting the molar and energy balances. The characteristic law of Multiport □ in resistance causality with uncertainty can be written as follows:

$$\begin{aligned} \dot{S} &= \Phi(\square_n, A)(1+\delta_{1/\square}) \\ &= \Phi(\square_n, A) + \delta_{1/\square}\Phi(\square_n, A) \\ &= \dot{S}_n + \dot{S}_{inc} \end{aligned} \quad (4)$$

The effort A is known at the entry of the multiport □ . $\dot{S}_n$, $\dot{S}_{inc}$, $\delta_{1/\square}$ represent, respectively, the nominal value, the multiplicative uncertainty. $\Phi(\square_n, A)$ is the global reaction rate and is written as

$$\Phi(\square_n, A) = r_f\left(1-\exp\left(\frac{A}{RT_r}\right)\right)V \quad (5)$$

where the reaction $r_f = k_0 \exp\left(\frac{-E_a}{RT_r}\right)\prod C_i^{\nu_i}$ represents non linear term which depends on reactional temperature T_r, according to Arrhenius equation, and concentrations ($C_i = \frac{n_i}{V}$). R is a universal gas constant, E_a is the activation energy of the reaction and k_0 is the pr-exponential factor.

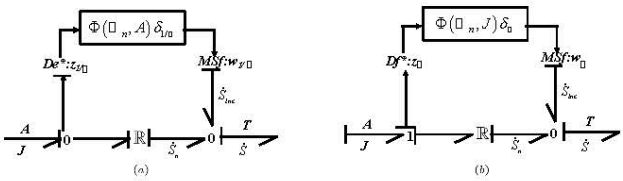


Figure 7. Multiport □ in resistance causality using the LFT form(a) Multiport □ in conductance causality using the LFT form (b)

The characteristic law of □ element in conductance causality is given as follows:

$$\begin{aligned} \dot{S} &= \Phi(\square_n, J)(1+\delta_{\square}) \\ &= \Phi(\square_n, J) + \delta_{1/\square}\Phi(\square_n, J) \\ &= \dot{S}_n + \dot{S}_{inc} \end{aligned} \quad (6)$$

The flow J is known at the entry of the multiport □ .

In the next section will be considered the use of coupled bond graph for FDI design.

IV. FAULT INDICATORS GENERATION FROM BOND GRAPH

In this paper, a bond graph methodology is used to synthesis a robust FDI method for nonlinear system in presence of parameter uncertainties, Figure 8.

Parametric uncertainties are explicitly appears on the BG, one can automatically generate the robust ARRs for the uncertain system by decoupling the nominal and the uncertain parts; residuals correspond to the ARR nominal part, while the residual thresholds represents the ARR uncertain parts.

The main advantages of the bond graph model in LFT form for robust diagnosis are given as follows:

- Introduction of the uncertainties on the nominal model, does not affect the causality and the structural properties of the BG elements;
- Representation of all uncertainties (i.e. structured and unstructured);
- Uncertain part is perfectly separated from the nominal part;
- Parameter uncertainties are easily evaluated.

This FDI method is summarized by the following steps: i) Modeling of studied system using bond graph tool with standard LFT form; ii) Generation of Analytical Redundancy Relations (ARRs) from the uncertain model by decoupling the nominal and the uncertain parts. Residuals correspond to the ARR nominal part, while their adaptive thresholds represent the ARR uncertain parts; ii) Residual' sensitivity analysis is done by using the ARR uncertain part.

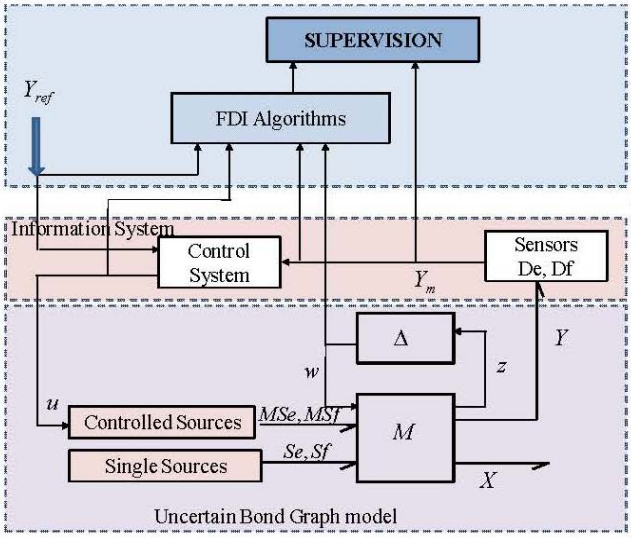


Figure 8. Representation of the robust FDI scheme using bond graph tool.

V. CASE STUDY: A CHEMICAL REACTOR

A. Process Description

Let us consider an adiabatic Continuous Stirred Tank Reactor, where the exothermic reversible reaction is occurred. This reaction is defined as follows:

$$n_A A \xrightarrow[k_r]{k_f} n_C C \qquad (7)$$

where ν_i (for i=A, C) are the stoichiometric coefficients. In our case these coefficients are equal to one.

The technological diagram of reactor system is depicted in Figure 9. The supply system (component A) consists of a storage tank and a pump. The level regulation is guaranteed by the means of a PI regulator acting on a centrifugal pump which supplies continuously the tank.

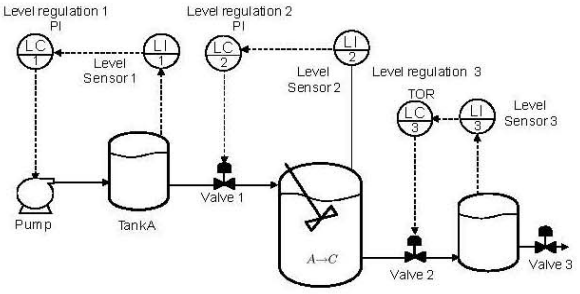


Figure 9. Technological diagram of the process.

The level controller in the reactor is ensured by a regulator which acts on a valve at the reactor input. The tank containing the components (C, A) is controlled in level by a regulator which acts a self-closing valve on the outlet side of the reactor.

B. Word Bond Graph model

The modelling hypotheses are, the reactor is perfectly stirred so that temperature and concentrations of different chemical species are homogeneous in all the reaction mixture, the reaction mixture is composed of one homogeneous liquid phase, and no phase change is considered, the volume of the liquid in the tank is constant. For illustration of developed method and because of limited space, we consider only the main component of the system: reactor vessel especially chemical domain.

The word bond graph model is presented in Figure 10. This model is decomposed into several modules, linked by a pair of pseudo power variables (effort-flow). To simplify the process modelling, we introduced bond graph model of reactor vessel which is composed of several parts corresponding to multi-energy domains.

The used pseudo power variables (effort-flow) are: pressure-mass flow $(P,\dot{m})$, temperature-enthalpy flow $(T,\dot{H})$ in the case of convection, and temperature-thermal flow $(T,\dot{Q})$ in the case of conduction, chemical potential-molar flow $(\mu,\dot{n})$, chemical affinity-reaction velocity (A,J).

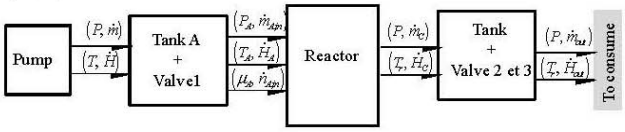


Figure 10. Word pseudo-bond graph of the chemical reactor.

C. Pseudo Bond Graph model

The bond graph model is given (Figure 11). This part includes chemical subsystem in reactor vessel. The bond graph transformers $TF:1/\nu_A$ and $TF:\nu_C$ represent a chemical transformation. Their modulus is the stoichiometric coefficients (the chemical affinity A represent the driving force in reactor vessel).

In the chemical domain the 0-junctions represent the molar balance of each component (A, C). 1-junction is used to represent the equality of the molar reaction flows of the different substances involved.

Thus, a □ -field is used as a link between the mass and energy parts of the reactor vessel subsystem. It is a two-port element connecting the molar and energy balances.

Modelling of bond graph elements $i \in \{C_A, C_C\}$ and multiport $\{\square\}$ in the LFT form consists in decoupling the nominal element $i_n \in \{C_{A,n}, C_{C,n}, \square_n\}$ part from its uncertain part $\delta_i i_n \in \{\delta_1 C_{A,n}, \delta_2 C_{C,n}, \delta_3 \Phi(\square_n, A)\}$, with δ_i is a multiplicative uncertainty on the parameter i.

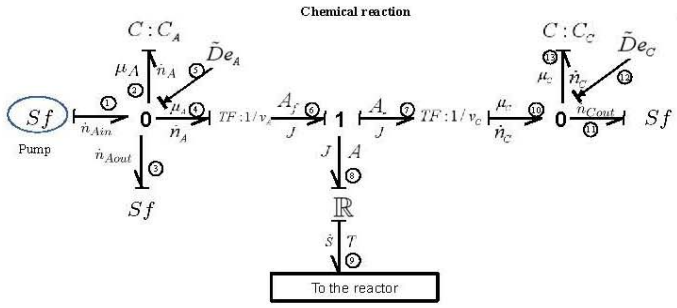


Figure 11. BG determinist model in preferred derivative causality of chemical domain

The determinist and uncertain bond graph model of the chemical domain are respectively given in Figures 11 and 12. The symbols $\tilde{D}e$ and $\tilde{D}f$ correspond to virtual sensors. They are used to distinguish the real measurements from the fictive ones.

The storage of chemical energies is modelled by the bond graph elements $C:C_A$ and $C:C_B$. Then the following equation is deduced from the junction 0 of the bond graph determinist model in derivative causality:

$$\begin{cases} \dot{n}_A = \dfrac{V}{RT}\exp\left(\dfrac{\mu_A-\mu_A^0}{RT}\right)\dot{\mu}_A \\ \dot{n}_C = \dfrac{V}{RT}\exp\left(\dfrac{\mu_C-\mu_C^0}{RT}\right)\dot{\mu}_C \end{cases} \quad (7)$$

where μ_A^0 is standard chemical potential.

In general case, the previous equations becomes

$$\dot{n}_i = C_i\frac{d\mu_i}{dt} \quad (8)$$

where $\dot{n}_i$ is the reaction output' molar flow. μ_i is chemical potential inside the reaction. C_i represents the chemical capacity of the reaction and can be expressed as follows:

$$C_i = \frac{V}{RT}\exp\left(\frac{\mu_i-\mu_i^0}{RT}\right) \quad (9)$$

The relation between $C_{i,n}$ and δ_{C_i} is given by the following expression:

$$C_i = C_{i,n} + \delta_{C_i}C_{i,n} \quad (10)$$

where $C_{i,n}$ is the nominal value of C_i.

The modulated input w_i $(i=2,3)$ in Figure 12 corresponds to an effort variable deduced from δ_{C_i} and expressed by the following equation:

$$w_i = -\delta_{C_i}C_{i,n}\frac{d\mu_i}{dt} \quad (11)$$

w_i is taken with a negative sign, because it is considered as a fictive flow input' source.

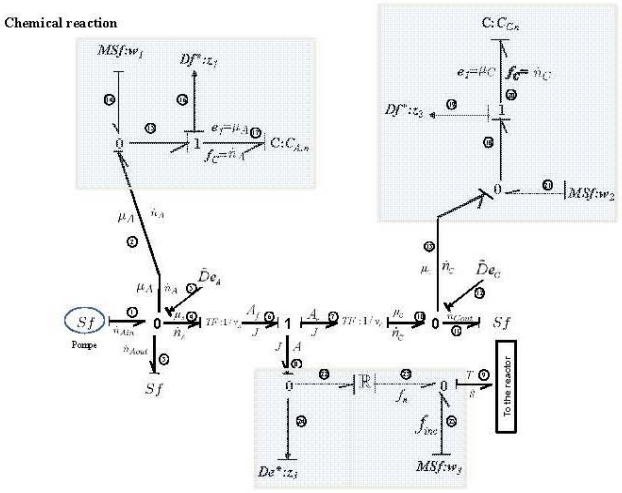


Figure 12. BG-LFT uncertain model in preferred derivative causality.

D. *Design of supervision system*

A method to derive ARR from bond graph models by applying the causality inversion algorithm, have been presented in Djeziri, Merzouki, Ould Bouamama and Dauphin Tanguy (2007), which use structural and causal properties.

1) *Determinist ARRs Generation*

The ARRs are deduced from junctions 0 that contain detectors on the nominal bond graph model of Figure 12. The unknown variables f_2 and f_4 are eliminated using covering causal paths from detectors to unknown variables.

From first junction 0

$$r_1 = f_1 + f_4 - f_3 - f_2 = 0 \quad (12)$$

with $f_1 = Sf = \dot{n}_{Ain}$, $f_3 = Sf_{out}$, $f_2 = C_A\dfrac{de_2}{dt}$

$$f_4 = v_A f_6 = v_A J = v_A\Phi(\square, A) = v_A r_f\left(1-\exp\left(\frac{A}{RT_r}\right)\right)V$$

where $\Phi(\square, A)$ is given by the equation (5).

f_2 is calculated (eliminated) from the following causal paths $f_2 \to \Phi_{C_A} \to e_2 \to \tilde{D}e_A : \mu_{A,m}$

The first ARR, r_1, is deduced from equation (12) and is given as

$$\begin{aligned} r_1 &= \dot{n}_{Ain} + v_A\Phi(\square, A) - Sf_{out} - C_A\frac{d\tilde{D}e_A}{dt} \\ &= \dot{m}_{in}\frac{C_{Ain}}{\rho_A} + v_A\Phi_R(n_A, m, H) - \dot{m}_{out}\frac{n_A}{m_{mé}} - \dot{n}_{A,m} = 0 \end{aligned} \quad (13)$$

with $m_m = \rho V = \rho S L_m$ (S sectional surface of the reactor) and $\dot{n}_{A,m} = \frac{V}{RT_r}\exp\left(\frac{\mu_{A,m} - \mu_A^0}{RT_r}\right)\dot{\mu}_{A,m}$.

From second junction 0

$$r_2 = f_{10} - f_{11} - f_{13} = 0 \tag{14}$$

and from the constraint in equation (14), the second ARR, r_2, is given by

$$\begin{aligned} r_2 &= f_{10} - f_{11} - f_{13} = v_C \Phi(\square, A) - Sf_{out} - C_C \frac{d\tilde{D}e_C}{dt} \\ &= \dot{m}_{out}\frac{n_C}{m_{m\acute{e}}} - v_C r_f\left(1 - \exp\left(\frac{A}{RT_r}\right)\right)V - \dot{n}_{C,m} = 0 \end{aligned} \tag{15}$$

The fault in chemical domain (appearance of secondary event: release of toxic or explosive material, etc.) related to transformer phenomenon can be detected by using the first and the second ARR.

2) Robust ARRs' Generation

In this section, the ARRs are generated for nonlinear systems, using bond graph approach in the LFT form. The aim of the robust diagnosis for the presented chemical reaction is to detect and isolate a chemical fault situation (appearance of secondary reaction when the reaction takes place; undesirable product and runway of the reaction) in presence of parameter uncertainties. This fault corresponds to the increase of the reaction velocity and chemical affinity, which is distinguished from the parameter uncertainties.

The chemical reaction model in the LFT form with derivative causality, after sensors dualization is given in Figure 12. The fictive inputs $w_i\ (i=1,...,3)$ are related with the fictive outputs $z_i\ (i=1,...,3)$ and expressed in the system of (16)

$$\begin{cases} w_1 = -\delta_1 z_1 ; z_1 = \dot{n}_A = C_A \dfrac{d\tilde{D}e_A}{dt} = C_A \dfrac{d\mu_A}{dt} = \dfrac{V}{RT} e^{\left(\frac{\mu_A - \mu_A^0}{RT}\right)} \dfrac{d\mu_A}{dt} \\ w_2 = -\delta_2 z_2 ; z_2 = \dot{n}_C = C_C \dfrac{d\tilde{D}e_C}{dt} = C_C \dfrac{d\mu_C}{dt} = \dfrac{V}{RT} e^{\left(\frac{\mu_C - \mu_C^0}{RT}\right)} \dfrac{d\mu_C}{dt} \\ w_3 = \delta_3 \Phi(\square_n, A)\, z_3 ; z_3 = e_6 = A_f = v_A \mu_A \\ \qquad = v_A\left(\mu_A^0 + RTLog\left(\dfrac{\int \dot{n}_A dt}{V}\right)\right) \end{cases}$$

where δ_1 and δ_2 represent, respectively, multiplicative uncertainties on the energy accumulation of reactant A and product C (leads to uncertainties in heat-storage capacity). δ_3 is the multiplicative uncertainty on the reaction velocity (leads to uncertainties in activation energy, pre-exponential factor, enthalpy...).

The two parts of the ARRs generated from chemical reaction model with parameter uncertainties of Figure 12 are given by equations (12) and (15) where r_1 and r_2 represent the ARRs nominal parts that describe the system operating. a_1 and a_2 (the ARRs uncertain part) represent the intake reduced by the parameter uncertainties such as flow or effort which affect the residuals. It is described by the sum of fictive input values and is used to calculate the normal operating thresholds.

$$\begin{cases} a_1 = |w_1| + |v_A w_3| \\ a_2 = |w_2| + |v_C w_3| \end{cases} \tag{17}$$

Uncertain ARR part cannot be quantified perfectly, it is evaluated to generate a normal operation' threshold which satisfies the following inequality:

$$-a \le ARR \le a \tag{18}$$

E. Simulations results

The chemical system is instrumented with the following sensors. The mixture temperature inside the tank De : Tm1, the level inside the tank De : Lm1, the flow sensor (Df : Fm1) is used to measure the amount of mixture leaving the tank. The water flow in the cooling circuit can be measured using the flow sensor (Df : Fm2). The output control signal of each controller is considered as a known value.

Figure 13 show respectively the residuals r_1 and r_2 without faults.

Fault scenario: Appearance of the secondary reaction: It is supposed now for example that the cooling system is never failing and that the exits of the regulators and the sensors are always correctly measured. A sudden appearance of secondary product occurs between 30 and 60 min. Indeed, to stop the evolution of the secondary reaction and to eliminate these effects in real-time, it is necessary to add a reagent able to eliminate the undesirable products. As can be seen in Figure 14, the appearance of undesirable product is detected perfectly by the residual evolution. The fault is detected perfectly, as it is alarmed by two residuals r_1 and r_2, and not by the other residuals. The thresholds of normal operation are given with dot lines.

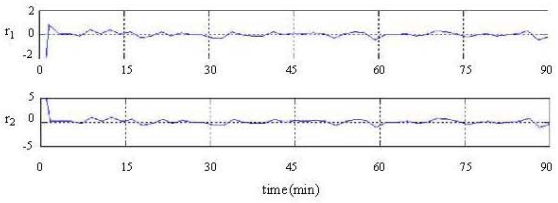


Figure 13. Residual evolutions of normal system.

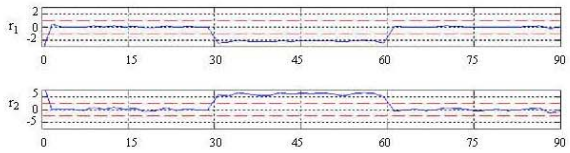


Figure 14. Residual evolutions of faulty system.

In order to explain the appearance of secondary reaction; for example, appearance of undesirable product E which modify reaction dynamics. Namely an unmodeled side reaction, is added to the simulation model; in detail the reaction scheme becomes:

$$n_A A \rightarrow n_C C + n_E E \quad (19)$$

and consequently the mass balance and the fault indicator will be modified. The RRAs (r_1 and r_2) should add the term $v_E \mu_{E,m}$ and can not be equal to zero.

VI. CONCLUSION

In this paper, a robust FDI with respect to parameter uncertainties is given using bond graph modelling approach in the LFT configuration. The robust ARR are generated directly from a bond graph model. This approach is study for complex systems where numerical values of parameters are not available. The obtained results are validated using real process (continuous reactor). The proposed FDI method witch can detect kinetic and thermodynamic drift of chemical reactors due to appearance of secondary reaction. The performances obtained are acceptable.

REFERENCES

[1] M. Blanke, M. Kinnaert, J. Lunze and M. Staroswiecki, Diagnosis and Fault-Tolerant Control. Springer Verlag, 2006.

[2] O. Levenspiel, Chemical Reaction Engineering. John Wiley and Sons, 1999.

[3] R. El Harabi, B. Ould Bouamama, M. El Koni Ben Gayed, M.N. Abelkrim, Robust Fault Dagnosis on Chemaical System by Using Uncertain Bond Graph Model, Proc. 8th IEEE International Multi-Conference on Systems, Signals and Devices (SSD 11), 22-25 Mars 2011, Sousse, Tunisia, 2011, pp.82. ISBN 978-1-4577-0412-3.

[4] M.A. Djeziri, R. Merzouki, B. Ould Bouamama, and G. Dauphin Tanguy, Bond graph model based for robust fault diagnosis. Proc. of the 2007 American Control Conference, New York, 2007. USA 3017-3022.

[5] M.R. Maurya, R. Rengaswamy and V. Venkatasubramanian, A signed directed graph and qualitative trend analysis-based framework for incipientfault diagnosis, Chemical Engineering Research and Design, vol. 85, no. 10, pp.1407-1422, 2007.

[6] W. Borutzky and G. Dauphin-Tanguy, Incremental bond graph approach to the derivation of state equations for robustness study, simulation. ModellingPractice Theory, vol.12, pp. 41-60, 2004.

[7] C. Sié Kam and G. Dauphin-Tanguy, Bond graph models of structured parameter uncertainties,.. Journal of the Franklin Institute, vol. 342, pp. 379-399, 2005.

[8] M.A. Djeziri, B. Ould Bouamama, and R. Merzouki, Modeling and robust FDI of steam generator using uncertain bond graph model. Journal of Process Control, vol.19, pp. 149-162, 2009.

[9] F. Couenne, C. Jallut, B. Maschke, P.C. Breedveld and M. Tayakout, Bond graph modelling for chemical reactors, Math. Comput. Modell. Dyn. Syst, vol. 12, pp.159, 2006.

[10] A.R. Khaled, , B.O. Bouamama and A. Nakrachi, Generic bond graphs procedure for chemical reactions modelling, computational engineering in systems applications. IMACS IEEE "CESA'06", vol. 1, pp. 412-417, 2006.

[11] J. Thoma, and B. Ould Bouamama, Modelling and Simulation in Thermal and Chemical Engineering: A Bond Graph Approach. Springer Engineering, 1999.

A Scalable Relational Database Approach for WebService Matchmaking

Deepak Chenthati
Teradata(R&D)India Pvt Ltd
Hyderabad, India
chvcdeepak@gmail.com

Hrushikesha Mohanty
Department of CIS
University of Hyderabad
Hyderabad, India
hmcs_ hcu@yahoo.com

Avula Damodaram
Department of CSE
JNTU Hyderabad
Hyderabad, India
damodarama@gmail.com

***Abstract* —Web services have a potential to enhance Business to Business collaboration by exposing compatible services to a service of higher dimension. Composition of web services essentially requires matching of their conversation protocols in addition to service requirements. For modeling and matching of conversations, the usability of formal structure like Finite State Machine (FSM) is well studied. However we are in opinion that for storing FSMs in graphs and retrieving them is not scalable for high cost in terms of memory and time. As in near future, as millions of web services will be available this scalability problem is a challenge to research communities. In order to cope up with this problem this paper proposes a relational model to store FSM models and retrieve for service composition.**

Keywords - webservice, RDBMS, matchmaking

I. INTRODUCTION

A confluence of Internet and computing technologies have given rise to a new computing paradigm that host several services for use by remote users over World Wide Web(WWW). These services use SOA: Services Oriented Architecture that not only lists the services offered on web but also provides a mechanism for their interactions. Interactions among services are necessary for making of a service of higher granularity by making use of services of lower granularity. This phenomena of service composition on synthesizing services largely depends compatibility of component services and this leads to B2B: Business to Business collaboration. As in business domain collaboration is largely welcome for greater benefits to collaborating business houses as well as users, match making of services is a problem of prime interest.

Service match making studies the composability of component services and one major aspect of studying this includes matching of conversational protocols of these services. Ideally, a conversation is feasible if for a send (of a message) by a service there is a receive at another service. Then we say both the services are in compatible for the conversation. The study of this conversation compatibility in literature is termed as service matchmaking. There has been considerable volume of research on this problem. The various techniques include ontology based [2][3][4], Fuzzy Logic based[5], Rough Set based[6] and Model based matchmaking approaches.

Model based approach is found interesting for its mathematical elegance as well as understanding. Models tried for analysis, orchestration and choreography of web services include EPC [14][15], Pertinets[16][17], and FSM[7][12]. On reviewing recent works on different modeling approaches we observe that FSM is being studied actively for modeling web services. Particularly for choreography to ensure successful communication among collaboration services. In [1] communication protocol of a web service has been extracted as a FSM and each edge is annotated with a message and its sender and receiver names. In order to ensure the correctness of a protocol between two web services, there must be well-defined FSM obtained due to interconnection of two FSMs that model communication of individual web services. A search of a compatible service for a given service essentially looks for a such well-defined FSMs from the searching and searched web services. Well-defined means the correctness in sequence of message transactions between two services i.e. for each send of a message there has to be a receive and the matching of sequence of sends with the sequence of receive of messages. An implementation of the scheme needs storage of FSMs in UDDI a repository of web services.

Because of wide spread usages of Internet there has been increasing trends in offering services, business applications on web giving rise to numerous web services. Maintaining storage of FSMs as graphs is tedious and difficult to scale up. Considering rise in number of web services available on web, managing storage at UDDI and their retrieval have become a challenge that needs an urgent attention. This paper takes up the issue and proposes a relational model to store FSMs and querying a system that matches to a given service and leads to a composed service.

II. SURVEY

FSM [7][12] has been used to model and match conversation protocols. However, we are in an opinion that storing, matchmaking and retrieving FSMs using graph structure is not scalable for representing conversation protocol for thousands of web services that will soon be available on web. In order to cope up with the challenge we propose relational model to store FSMs and a query system to retrieve services that are matching to a given service.

Another work [11] deals with the problem we address here. The authors view composability of services in terms of matching their input and output parameters. Two services match when one's output parameters match with input parameters of the other. While matching, they have used semantic similarity based in WordNet. Service input output parameters are stored in databases. Being soft in semantic matching, we may get several services matching with a given service but at different degrees. The matching result is represented in a directed graph. Where an edge connects two matching services and the label of the edge notes the degree of matching. Then the service composition problem is seen as shortest path finding problem. Other than that QoS is also considered in deciding weight of an edge. Shortest path finding algorithm is modified to operate on relational repository of service data.

Our work is distinct from [11] by taking message transactions into consideration. We argue, for work-ability of a composed service matching of their message communications is important. In-spite of input output matching, a composition will fail to match in case of their mismatch in communication. While matching we consider the sequence as well as importance of messages (defining types like compulsory and optional). Alike [11] we have used relational database to store communication details in order to make the technique scalable. Message communications of each service are recorded in databases and matching of two services is performed by joining of two tables. The join operation is relaxed to provide results of kinds of matching Full Match, Partial Match and No Match. We have also discussed possible implementation of the proposed technique. We view our technique along with the technique proposed in [11] will give precisely workable service composition. In [13] authors proposed a method in which, web services composition are computed in advance and stored in tables. For web services composition searches, they look up the pre-computed tables rather than actual web services.

Rest of the paper is organized as follows. Section 2 illustrates FSM based approach for modeling service communication and matching of compatible services for service composition. Section 3 describes RDBMS schema to store services and its communication protocol. Also this section presents a new matching algorithm and explains it with an example. Section 4 describes the way RDBMS is embedded into the currently available jUDDI architecture and how it is used for matching of services to compose a larger service. Section 5 concludes the paper.

III. SERVICE COMPOSITION AND MATCHMAKING

A service has business logic driven by business communication. In [7] we have shown how both can be modeled by FSM. Here, we are only interested in modeling business communications by FSM as shown in Figure 1. Each node indicates a state and an edge for communication. A label on edge follows a format sender#receiver#message details. Figure 1 specifies two services name Book store and Shipping and the chain of events follow from ordering of book to shipping it are as follows:

Let us consider the example of Ordering a Book process in an electronic book store. The chain of events will be as follows:

1) Customer(c) Requests for a Book in e-Book store (bs).

2) Book store sends availability confirmation to customer if book is available else it says Unavailable to customer.

3) Customer places an order for the book.

4) Book Store responds with make payment message.

5) Customer makes payment.

6) Book Store sends the payment confirmation.

7) bs interacts with shipping service provider (s) with a request message.

8) The shipper replies to bs with availability information if shipping is available to that particular location.

9) Then bs puts a message to Deliver the book.

10) On receiving Delivered message bs chooses any oneof the modes of payment (Credit Card) CCPaymentand (Debit Card) DCPayment.

11) Shipper sends a payment confirmation message onreceiving the payment.

12) Book Store sends Book Delivered message to customer.

The aim of this paper is to identify service partners whose message exchange sequence matches that of the requesting service. From the figure we see that the Book Store service communicates with consumer and shipper. Here shipper is the service collaborator with whom Book Store service collaborates. The states that are darkened involve communication with that of shipper. Now the aim is to find the shipper service whose message flow matches with that of the Book Store service.

Figure 1 shows the AFSM representation of Ordering a Book process. Here bs#c#Avail Conf, c#bs#Order Book, bs# c#MakePayment, bs#c#PayConf and c#bs#BookDelivered and c#bs#Payment are mandatory messages bs#s #CCPayment and bs#s#DCPayment are optional messages.

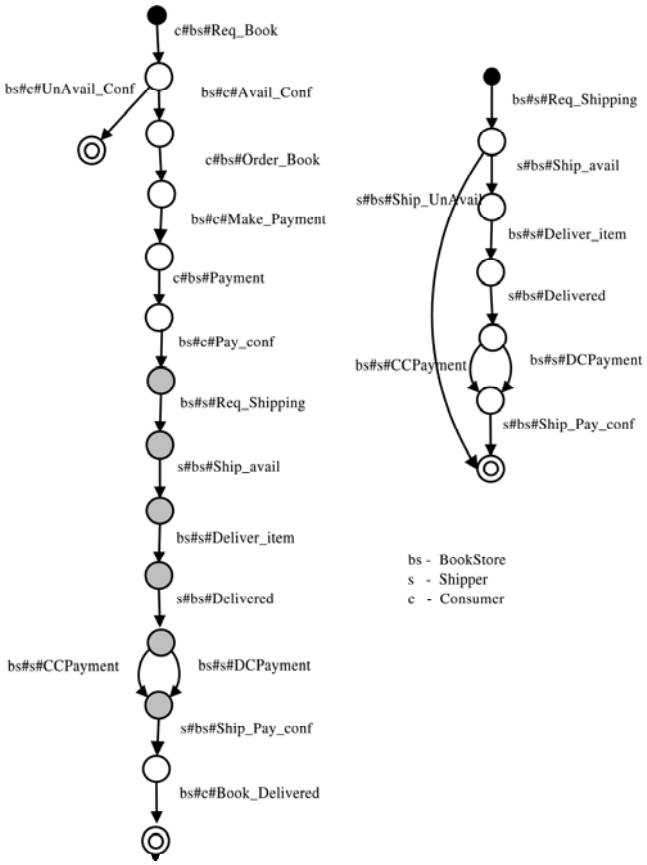


Figure 1. AFSM representation for "Book store" and "Shipping" services

A match between two FSMs is said to have occurred if and only if there exists a non-empty intersection between the two corresponding FSMs. For non-empty intersection, first we will do intersection operation on two FSMs, then we will perform emptiness test on the resulted intersected FSM. If emptiness test fails (i.e., FSM is accepting some string of messages), then it indicates that there is a path from start state to final state in that intersected FSM. So, we say that two FSMs are matched. Drawback of FSM based matchmaking is that storing, matching and retrieving FSMs using graph structure is not scalable when there are huge number of services that are available on the web. Hence, we look for relational repository for this purpose.

IV. A PROPOSED SCALABLE APPROACH

Relational databases have been successful in maintaining and accessing large repository of information. For this problem, we resort to the same technology to store FSMs for each service and to perform matching required for service composition.

A. Service Repository

In the literature, a way of storing web service details and its ontology's using Relational Database was proposed [2] [3]. But it had not taken the message flows into concern. So, we have proposed an RDBMS schema for storing not only messages but also the sequence of communications. Figure 2 shows the E-R diagram of the proposed schema. From the figure, we can observe that a single BusinessEntity can provide many BusinessServices and for every BusinessServices there is a corresponding description about its ServiceMessageFlow. We have also proposed the RequestorMessageFlow table which specifies the format a customer has to send his/her request.

BusinessEntity: A given instance of the BusinessEntity is uniquely identified by a BusinessKey. Simple textual information about the BusinessEntity is given by its name and contacts like e-mail, WebURL.

BusinessServices: A given BusinessServices entity is uniquely identified by its ServiceKey. The BusinessKey attribute uniquely identifies the BusinessEntity which is the provider of the BusinessService. Simple textual information about the BusinessServices is given by its name and short service description and its Category.

ServiceMessageFlow: A given instance of ServiceMessageFlow is uniquely identified by its ServiceKey, msg and followedByMsg. This table stores the message flow of services. Type of message, optional or mandatory, is specified by its type field. snd rcv field specifies whether the message is a sent by customer or received by customer.

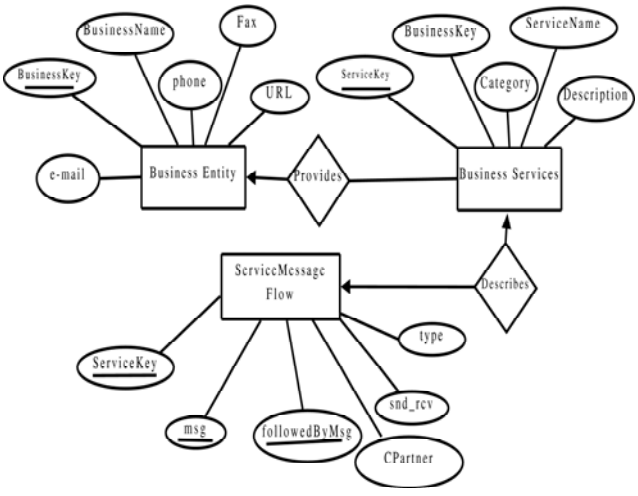


Figure 2. E-R Diagram

RequestorMessageFlow: A given instance of RequestorMessageFlow structure is uniquely identified by its msg and followedByMsg. The table has the message flow of a requested service, which can be given externally or can be generated from the ServiceMessageFlow table. Now, we can

BusinessEntity (BusinessKey, BusinessName, Phone, Fax, e-mail, URL)
Provides (BusinessKey, ServiceKey)
BusinessServices (ServiceKey, BusinessKey, ServiceName, Description, Category)
Describes (ServiceKey, msg, followedByMsg)
ServiceMessageFlow(ServiceKey,msg, followedByMsg, snd rcv, type, CPartner)
RequestorMessageFlow(msg, followedByMsg, snd rcv, type)

Figure 3. Proposed RDBMS Schema Supporting AFSM details

TABLE I. SERVICE MESSAGE FLOW

sk	msg	followedByMsg	snd_rcv	type	Cpart
1	Req_Book	NULL	receive	man	cust
1	Avail_Conf	Req_Book	send	man	cust
1	Order_Book	Avail_Conf	receive	man	cust
1	Make_Payment	Order_Book	send	man	cust
1	Payment	Make_Payment	receive	man	cust
1	Pay_Conf	Payment	send	man	cust
1	Req_Shipping	Pay_Conf	send	man	shipper
1	Ship_Avail	Req_Shipping	receive	man	shipper
1	Deliver_Item	Ship_Avail	send	man	shipper
1	Delivered	Deliver_Item	receive	man	shipper
1	CCPayment	Delivered	send	opt	shipper
1	DCPayment	Delivered	send	opt	shipper
1	Ship_Pay_Conf	DCPayment	receive	opt	shipper
1	Ship_Pay_Conf	CCPayment	receive	opt	shipper
1	Book_Delivered	Ship_Pay_Conf	send	man	cust
2	Req_Shipping	NULL	receive	man	cust
2	Ship_Avail_Req	Req_Shipping	send	man	cust
2	Deliver_Item	Ship_Avail	receive	man	cust
2	DCPayment	Deliver_Item	receive	man	cust
2	Ship_Pay_Conf	DCPayment	send	opt	cust
3	Req_Shipping	NULL	receive	man	cust
3	Ship_Avail	Req_Shipping	send	man	cust
3	Deliver_Item	Ship_Avail	receive	man	cust
3	Delivered	Deliver_Item	send	man	cust
3	DCPayment	Delivered	receive	man	cust
3	Ship_Pay_Conf	DCPayment	send	opt	cust

summarize the tables in the proposed RDBMS schema as described in the Figure3.

Considering the same example *Ordering the book* from section 2. Fig. 1 shows the FSM with annotations and Table I show the schema representation of the corresponding FSM for Book Store with ServiceKey (SK) as 1. We propose 6 column table;Where msg is the message for communication, followedByMsg is the previous message which was received, snd_rcv gives the detail if message is sent or received,there are two types of messages 'mandatory' (man) and 'optional' (op), Communication Partner (CPartner) gives the partner details with whom the service collaborates

Let us take a look at the first row. Here, *followedByMsg* is NULL because there is no preceding message to Book message. And rcv value is *receive* because BookStore(bs) is receiving that message from customer(c) type is set to mandatory. Similarly all the remaining rows in that table are inserted. Number of rows in a table is equal to the number of messages in FSM. Messages *CCPayment* and *DCPayment* are entered into table as optional because, by default if annotated FSM state has two or more messages emerging from it, those messages are treated as optional.

B. Service Matching Algorithm and Analysis

By looking at the example given in section 3, we have observed that a single services message flow spans over multiple rows in a relational table. So, it is not possible for any type of query in RDBMS to perform match operation between two such multiple row spanned message flow sequences. So, we have proposed an algorithm for matchmaking of web services based on above defined tables. This algorithm is also implemented as a stored procedure in MySQL. After matching process is completed, a service falls into one of the three categories namely *Exact match*, *Partial match* and *No match*.

Algorithm IV.1

Input: *servicemessageflow table, ServiceKey(skey) and req_category*
Output: *servicename and type of match*
Algorithm

```
1: match()
2: {
3: Build_RequestMessageFlow Table
4: cnt ← number of tuples in bussinessservices
   where category = req_category;
5: while cnt>0 do
6:    sk ← bussinessservices.servicekey;
      {For each service with unique sk}
7:    create temporary table t1 where service message-
      flow.ServiceKey= sk;
8:    create another temporary table t2 where msg and
      followedByMsg of t1 and RequesterMessageFlow are
      equal which is full outer join table.
9:    full, partial ← 0;
10:   ct ← numberoftuplesint2;
```

```
11:   m1,fm1,sr1,ty1,m2,fm2,sr2,ty2← t2(msg1,
      followedbymsg1,snd_rcv1, type1,msg2,
      followedbymsg2,snd_rcv2, type2);
12:   while ct>0 do
13:     if ((sr1='send' and sr2='receive') or (sr1='receive'
        and sr2='send')) then
14:       if (m1==m2 and( fm1==fm2 or (fm1=NULL or
          fm2=NULL)) then
15:         if (ty1='man' or ty1='op') and (ty2='man' or
            ty2='op') then
16:           full ← 1;
17:         end if
18:       end if
19:     else if (fm1=fm2) then
20:       if (m1 != NULL and m2 == NULL) then
21:         if (ty1='man' and ty2='opt') then
22:           partial ← partial +1
23:         else if (ty1='opt' and ty2='opt') then
24:           full ← 1
25:         end if
26:       end if
27:       if (m1 == NULL and m2 != NULL) then
28:         if (ty1='opt' and ty2='man') then
29:           partial ← partial +1
              full ← 0
30:         else if (ty1='opt' and ty2='opt') then
31:           full ← 1
32:         end if
33:       end if
34:       if (m1!=m2) and m1!= NULL and m2 != NULL)
          then
35:         if (ty1='man' or ty1='opt') and
            (ty2='man or ty2='opt') then
36:           partial ← partial +1
              full ← 0
37:         end if
38:       end if
39:     end if
40:     if ((m1==m2) and (fm1!=fm2)) then
41:       partial ← partial +1
42:       full ← 0
43:     end if
44:     if (m1=NULL and fm1=NULL) and (m2!=NULL
        and fm2!=NULL) and (ty2='man') then
45:       full ← 0
46:     end if
47:     if (m1!=NULL and fm1!=NULL) and (m2=NULL
        and fm2=NULL) and (ty2='opt') then
48:       full ← 0
49:     end if
50:     ct ← ct -1
51:     increment one tuple in t2;
52:   end while
53:   if ( full=1 and partial =0) then
54:     Print ServiceName, "Exact Match";
55:   else if ( full=1 and partial>0) then
56:     Print ServiceName, "Partial Match"
57:   else
58:     Print ServiceName, "No Match";
59:   end if;
60:   cnt ← cnt-1;
61:   increment one tuple in BussinessServices;
62: end while
63: }
```

In case of *exact match*, all the messages of both service provider and requestor message flow sequence matches totally (or) some optional messages of requestor message flow sequence may not be present in message flow sequence of service provider. In case of *Partial match*, some messages of service provider message flow sequence don't have corresponding messages in requestor message flow sequence. In case of *no match*, no messages of service provider and requestor match (or) some mandatory messages of requestor message flow sequence don't have corresponding messages in service provider message flow sequence.

Let us take a look at the steps involved in the proposed matching algorithm. Line numbers at the end of each step refers to the part of the algorithm that is performing that particular step.

- RequestMessageFlow table shown in Table II is created from ServiceMessageFlow table which has the same ServiceKey as given input and whose Cpartner is same as req_Category.
- Temporary table is created for a service from *ServiceMessageFlow* that has matched *Category* with requestor's category. (Line 7,8)
- Another temporary table is created by doing a full outer join operation on the above obtained table and *RequestMessageFlow* table based on their *msg* and *followedByMsg*.
- For each message (row or tuple), if for each send type of message there is receive type of message with the provider service or vice-versa then if the attributes *msg*, *followedByMsg* of both (service provider and requestor) in the above full outer join table are equal and if that is the first message the followed by messages would be null this check is performed and full is set to 1. (Lines 13-15)
- For the messages whose previous messages match different conditions are checked. (Line 19)
- If both message types are equal (or) if service provider has mandatory message corresponding to optional message of requestor then set *full* to 1. (Lines 20-26)
- If there is no message in requestor for the corresponding message in service provider then for

mandatory msg of provider partial is incremented else *full* is set to 1 (Lines 20-25)

- If there is no message in provider for a corresponding message in requestor if the type of requestor is mandatory *partial* is incremented and *full* is set to 0.(Lines 27-29) It is given partial as there is scope for negotiation. If type is optional then *full* is set to 1.
- If there is mismatch in messages then the partial value is incremented and *full* is set to 0.
- Now, if full is 1 and partial is 0 then the services are **'Exactly Matched**' else if *full* is 0 and *partial* >0 then they are **'Partially Matched**' else they are **'Not Matched'**. (Lines 53-59)
- Repeat the above steps for all services that are matched with requestor category. (Lines 5-60)

V. Simulation and Analysis

Let us again consider the same example of Ordering a Book service. The following Table 1 and Table 2 shows two tables: one is ServiceMessageFlow table containing three services (Assuming that only these three services are available and all of them belongs to same category), and other one is RequestorMessageFlow. Let us analyze the matching process of service 2 message flow with requestor message flow. Full outer join table of service 2 and RequestorMessageFlow tables is shown in Table III

TABLE II. Requester Message Flow

msg	followedByMsg	snd_rcv	type
Req_Shipping	Pay_Conf	send	man
Ship_Avail	Req_Shipping	receive	man
Deliver_Item	Ship_Avail	send	man
Delivered	Deliver_Item	receive	man
DCPayment	Delivered	send	man
Ship_Pay_Conf	Payment	receive	man

Initially, full and partial are set to 0. First three messages (rows) of service 2 are exactly matched with that of requester messages. line 14-18 of algorithm handles it. From the join table shown in Table III, There is a mismatch in the fourth message of service 2 and from the join table we can observe that the followedByMsg's are different where as the messages are same. Since there is a skip in the message exchange both the services can negotiate hence lines 41-42 will set the value of full to 0 and increment the partial value. Again there is match in the fifth message hence full is set to 1 but partial value is 2 hence its a Partial Match. Few messages that are specified in the provider service may not be present at the requester's end, and both the requester and provider can negotiate and agree upon the missing message.

In this way our matching algorithm performs matching of each service in *ServiceMessageFlow* table with the *RequesterMessageFlow* table. Similarly, according to the above given algorithm, service 3 matches exactly with requester's one and service 2 has partial match with requester's one. Due to the space constraint we have not represented all the cases. As we told earlier that we have implemented the above algorithm as a stored procedure in MySQL, the way of calling it is as follows: call match();

VI. A New Framework

UDDI [8] uses WSDL [9] to describe interfaces to web services. Some organizations have already implemented UDDI depending on their requirements. One of them is jUDDI[10], which is developed by Apache group and it is open source. We have extended this jUDDI by means of adding *Service communication message flow Database* and *Matching algorithm based on RDBMS* as highlighted in Figure 3. . As we told earlier, Service communication message flow Database stores the message exchange sequences of all services that are registered in UDDI. This database is used by our newly proposed matching algorithm

A. JUDDI Extension

Now, the overall architecture of jUDDI can have the following components.

Publish Service: Service Provider who is interested in providing service will publish their service through this module. They publishes their business information in *Service Database* and service communication message flow details in *Service Communication Message flow Database* as shown in Figure 3

Query Service: Service Requestor who is interested in consuming service sends his/her request through this module. This request is processed by *Matching Algorithm based on RDBMS* component. After the matched services are returned, requestor will contact that particular service provider by getting information from Service Database.

Matching Algorithm based on RDBMS: This component receives request from requestor and performs matching with each service in *Service Database* by means of taking message flow of corresponding service from *Service Communication Message flow Database*.

Service Database: It contains the database schema required for storing information related to business service like Name, email, URL, etc Service Communication

Message flow Database: It contains the database schema required for storing the communication message flow sequence of each service.

Three functionality's are primarily provided by jUDDI namely a) Publish Service b) Query Service and c) Compose service. Service Providers who want to provide their service publishes their details into the Service Database and their message flow sequences into Service Communication

-message flow Database. The organization which requires a service places a query by specifying the type of service it wants. This query is handled by matching algorithm component. The organization which requires service composition sends its required message exchange sequence. This is handled by matching algorithm component and returns the resulted set of matched services by means of interacting with both Service Database and Service communication message flow Database components.

TABLE III. FULL OUTER JOIN TABLE FOR SERVICE 2

Skey	msg	followedByMsg	snd_rcv	type	msg	followedByMsg	snd_rcv	type
2	Req_Shipping	NULL	receive	man	Req_Shipping	Pay_Conf	send	man
2	Ship_Avail	Req_Shipping	send	man	Ship_Avail	Req_Shipping	receive	man
2	Deliver_Item	Ship_Avail	receive	man	Deliver_Item	Ship_Avail	send	man
2	Delivered	Deliver_Item	send	man	Delivered	Deliver_Item	receive	man
2	DCPayment	Delivered	receive	man	DCPayment	Delivered	send	man
2	Ship_Pay_Conf	DCPayment	send	man	Ship_Pay_Conf	DCPayment	receive	opt
2	NULL	NULL	NULL	NULL	CCPayment	Delivered	send	opt
2	NULL	NULL	NULL	NULL	Ship_Pay_Conf	DCPayment	receive	opt

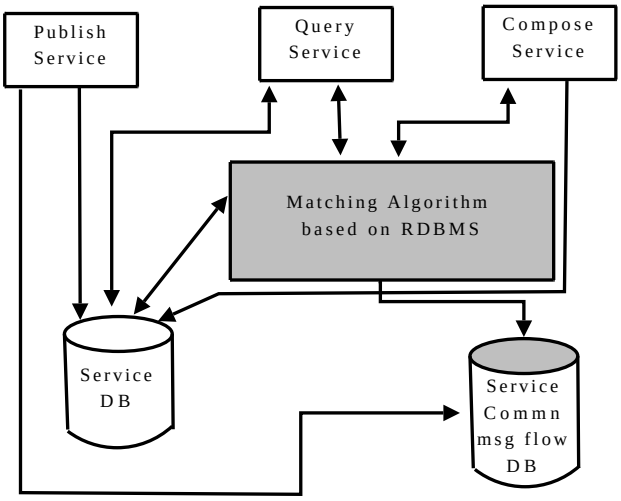


Figure 4. Extended jUDDI Architecture

VII. CONCLUSION

This paper has proposed a scalable approach for matchmaking of web services for composition of services of higher granularity from given atomic services. FSM models for communication for each service is extracted and stored in a relational database. This provides a mechanism that makes the proposed approach scalable to store a large number of web services and retrieve a service as per the requirement of matching to a given service. The proposed idea is implemented in jUDDI open source software. The software is extended with the proposed algorithm and tested to establish the practicality of the proposed technique. This work also proposes usages of established technology in developing web services and making this available on Internet to potential users.

REFERENCES

[1] Andreas Wombacher, Peter Fankhauser, Bendick Mahleko, Erich Neuhold : "Matchmaking for Business Processes based on Choreographies". In Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service, Taipei, Taiwan, pages 359-368,28-31 March (2004)

[2] Ruiqiang Guo, DongDong, Jiajin Le, Discovery for Web Services based on Relationship Model, In Proceedings of the 6th International Conference on Computer and Information Technology, Korea, pages 253-259, September (2006).

[3] Souripriya Das, Eugene Inseok Chong, George Eadon, Jagannathan Srinivasan, Supporting Ontology-based Semantic Matching in RDBMS, In Proceedings of the 30th VLDB Conference, Toronto,Canada, pages 1054-1065, August 29- September 3, 2004.

[4] Hongsuda Tangmunarunkit, Stefan Decker, Carl Kesselman, Ontologybased Resource Matching in the Grid - The Grid meets the Semantic Web, In Proceedings of International SemanticWeb Conference, Sanibel Island, Florida, USA, pages 706-721, 20-23 October 2003

[5] Kuo-Ming Chao, Muhammad Younas, Chi-Chun Lo, Tao-Hsin Tan,Fuzzy Matchmaking for Web Services, In Proceedings of the 19th International Conference on Advanced Information Networking and Applications, Tamkang University, Taipei, Taiwan, pages 721-726, 28-30March 2005.

[6] Maozhen Li, Bin Yu, Chang Hang, Yong-Hua Song Service Matchmaking with Rough Sets, In Proceedings of the 6th International Symposium on Cluster Computing and the Grid, Singapore, Vol-1, pages 123-130, 16-19 May 2006.

[7] Hrushikesha Mohanty, Deepak Chenthati, Supriya Vaddi, and R. K. Shyamsundar. 2008. HUMSAT for State Based Web Service Composition. In Proceedings of the 2008 International Conference on Information Technology (ICIT '08). IEEE Computer Society, Washington, DC, USA, Pages 273-278.

[8] UDDI Version 3.0.2 from http://uddi.org/pubs/uddi v3.htm

[9] WSDL Version 1.1 from http://www.w3.org/TR/wsdl.

[10] JUDDI from http://ws.apache.org/juddi/.

[11] Cheng Zeng, Weijie Ou, Yi Zheng, Dong Han. Efficient Web Service Composition and Intelligent Search Based on Relational Database, In International Conference on Information Science and Applications (ICISA) seoul,Pages 1-8, 21-23 April 2010.

[12] Hu jingjing, Zhao xing, Cao Yuanda, Zhou ruitao. A Service Composition Model with Characteristic of Transaction Based on Finite State Machine. International Conference on Computer and Electrical Engineering, 2008. ICCEE 2008. 20-22 Dec. 2008. Page(s): 450 - 454

[13] Scalable and efficient web services composition based on a relational database, Journal of Systems and Software(JSS- 8735), In Press, Corrected Proof, 7 June 2011.Daewook Lee, Joonho Kwon, Sangjun Lee, Seog Park, Bonghee Hong

[14] August-Wilhelm W. Scheer "Aris-Business Process Frameworks".Springer-Verlag New York, Inc., 2 edition, 1998

[15]] J. Ziemann and J. Mendling. "Epc-based modelling of bpel processes: a pragmatic transformation approach.", In Proceedings of the 7th International Conference Modern Information Technology in the Innovation Processes of the Industrial Enterprises (MITIP 2005), Genova, Italy, 2005.

[16] Hui Kang, Xiuli Yang, Sinmiao Yuan,"Modeling and Verification of Web Services Composition based on CPN,",Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on Network and Parallel Computing Workshops2007 , pages 613 – 617

[17] CPN tools. http://www.daimi.au.dk/CPNtools/

Observation of Human Brainwave Signals Due to Mobile Phone Usage

Zunairah Hj Murat, Ros Shilawani S.Abdul Kadir, Roshakimah Mohd Isa, Aisyah Hartini Jahidin, Mohd Nasir Taib and Norizam Sulaiman

Faculty of Electrical Engineering
Universiti Teknologi MARA
40450 Shah Alam, Selangor Darul Ehsan
zunairahh@yahoo.com

***Abstract* – There is ongoing discussion whether the cellular or mobile phone usage causes any health effects. The aim of this research is to investigate the effects of mobile phone usage on human brainwaves using electroencephalograph (EEG). The brainwave signals were analyzed using Statistical Package for the Social Sciences (SPSS). Thirty samples were interviewed prior to EEG recording. Then, the EEG recordings were performed for three sessions; before, during and after the phone calls. The findings show that alpha is dominant compared to the other frequency bands. In addition, value for the left brainwaves always higher than the right for all frequency bands which means that the samples were left brain dominant. The overall correlation between left and right brainwaves signal for all bands shows decrement for during and after phone calls, thus reducing brainwaves balancing. Therefore, there is evidence that the usage of mobile phones affect the brainwaves.**

Keywords - EEG, Brainwaves, Radiofrequency

I. INTRODUCTION

Mobile or cellular phones are now an integral part of modern telecommunications. In many countries, over half the population use mobile phones and the market is growing rapidly [1]. Given the large number of mobile phone users, it is important to investigate, understand and monitor any potential public health impact [1-2].

Mobile phones communicate by transmitting ultra-high-frequency radio waves through a network of fixed antennas called base stations. Radiated power from an antenna is approximately up to 125mW [3]. Antennas within phones emit the waves while the strength tails off quickly as distance from the antenna increases, a sizable chunk of it is emitted through the brain [4]. With a distance of within 2cm from a user's head, mobile phones can radiate radiofrequency (RF) signals in the range of 450 to 2500 MHz [5]. RF waves are electromagnetic fields, and unlike ionizing radiation such as X-rays or gamma rays, cannot break chemical bonds or cause ionization in the human body [1].

II. LITERATURE REVIEW

A. Brainwaves and EEG

The brainwave is defined as arrhythmic of electric potential between brain cells called neurons and proficiently captured by EEG equipment [6]. Brainwave signals are grouped into four types which are Alpha, Beta, Theta, and Delta. The frequency of alpha wave is from 8 to 12 Hz and significantly present when the person is in a relaxed condition or reflecting with closed eyes [7-8]. During this state, a person is still awake yet resting. Slightly higher from alpha, beta wave's frequency, ranges from 12 to 30 Hz. Beta wave is indicative of active, busy or anxious thinking and active concentration [7]. Thus, related to the alert or working state. Delta wave is the lowest frequency range starting from almost zero and can be only up to 4 Hz. It is higher during sleeping mode, whereas Theta ranges from 4 to 7 Hz. It is dominant when someone is feeling tired and depressive [7-8].

EEG is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a certain period of time [7-8]. EEG test is harmless and painless and can be repeated [9]. Electrodes are placed on specific sites on the scalp to detect and record the electric signal impulses within the brain. EEG electrodes transform ionic current from cerebral tissues into electrical current used in EEG preamplifier [9]. This device will detects and amplifies the electrical signals and record them onto software in the computer. Some applications of EEG are as diagnostics tools for the case of epilepsy, coma and brain death [9-10].

B. Mobile Phone Exposure

During operation, mobile phones emit energy in the form of electromagnetic fields known as radio waves. A

GSM mobile phone can emit waves with a peak power of 2 watts and parts of these waves are absorbed by the user's head [3] as well as the human's body [11-13]. Tissue heating is the principal mechanism of interaction between radiofrequency energy and the human body. At the frequencies used by mobile phones, most of the energy is absorbed by the skin and other superficial tissues, resulting in negligible temperature rise in the brain or any other organs of the body [1].

Usage of mobile phones particularly for a long period of time is known to have some effects on the users [1-5]. Some users complained that the usage of mobile phone indicate certain negative effects on their bodies especially in the heads. Thus, several investigations to observe biological effects of mobile phone exposures have encompassed the investigation of potential connections to cancer, cell division, blood pressure alteration, induction of epilepsy, depression, effects on the eyes, and human cognitive alteration [14-15].

Another study suggested that mobile phone users had a 30% increased risk of brain tumors which occurred close to the ear used for mobile phone listening [2-3]. Furthermore, previous studies had shown that growth of leukemia cells could be increased dramatically after exposure to mobile phone radiation [2]. Therefore, various studies have been going on to investigate the effects of these conditions on the human body [1-5, 14-15]. However; it is presently unclear whether this electromagnetic energy can really cause biological consequences or adverse health effects [11, 15].

III. PROPOSED NEW IDEAS

To date, it is assessed that there are over 1.7 billion mobile phone users world-wide. Thus, if there were to be adverse health consequences due to contacts from mobile phones, the effects could be pervasive amongst huge populations. Therefore, it is essential to examine and to resolve the possibilities of biological effects due to mobile phone radiated fields [15-20]. Thus far, scientific evidence on the effect of mobile phone radiation exposure to the brainwave signals using EEG is not conclusive [11, 15]. Hence, this research will identify and investigate the human brainwave patterns due to the usage of mobile phone using EEG. In addition, the difference between the left hemisphere and the right hemisphere of the brainwaves will be observed and correlated to investigate the brainwave balancing condition [21-22]. Furthermore, interview sessions will be carried out to investigate the usage trend of mobile phone to be more comprehensive in the data collection and data analysis.

IV. METHODOLOGY

Experiments were performed at the Biomedical Research and Development Laboratory for Human Potential, Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. This laboratory has been recognized by the Malaysia Book of Record as the first research and development laboratory for human potential in Malaysia.

30 participants were recruited from undergraduate students of the Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. All participants were in healthy conditions and also not consuming any medicine or drug prior to the test. A combination method consisting of interviews and EEG analysis has been used for this project to be more comprehensive in gathering the information to analyze the effects of mobile phone usage.

A. Interviews

Before recording the EEG measurements, all samples were interviewed concerning their usage of mobile phone. Samples have to answer 11 questions such as number of call that they normally have per day, duration per call and the effects that they experience when using mobile phone for a long period of time. The results were analyzed to observe the trend of mobile phone usage among samples.

B. EEG Experiment

The EEG signals were recorded under three conditions which are before, during and after call using mobile phone. Figure 1 shows the flowchart of the experiment.

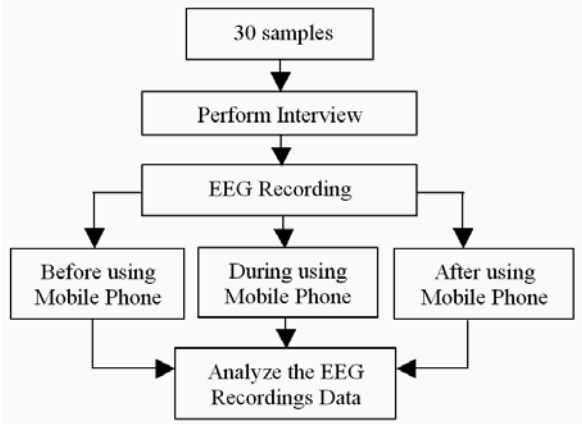


Figure 1: Flow Chart of the Experiment

Table 1: Time Frame Protocol of the Experiment

5-10min	5 min	1min	5 min	1min	5 min
Interview session	Before call Closed eyes Relax EEG recording	Rest	During call Closed eyes Relax EEG recording	Rest	After call Closed eyes Relax EEG recording

Table 1 shows the time frame protocol of the experiment, involving three stages which were before, during and after usage of mobile phone. Initially, samples will undergo the interview sessions to answer 11 item questionnaires related to their usage of hand phones. The interview session is normally conducted between five to ten minutes. The EEG recording duration was five minutes for each stage with one minute rest period in between, giving a total of approximately 22 to 27 minutes for each sample. During EEG recordings, samples were asked to close their eyes, relax (but not sleep), not allowed to talk, thus to minimize artifacts in the EEG signal.

The phone is strapped to the right ear while the phone is in active session. Although the line is open to another party, no conversation took place between them. Observation of brainwave was conducted for both brain hemispheres, though only the right ear is engaged with the phone. In future, the research can be extended by placing the phone to the other ear to observe the effect.

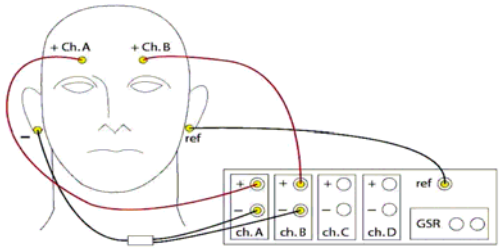


Figure 2: Electrodes and WaveRider Connection

EEG data acquisitions were carried out using EEG equipment (WaveRider Mind Peak model P-0609-5E022) as shown in Figure 2. A bipolar connection is employed using four electrodes, 2 channels and in accordance to International Standard 10-20 electrode placement system. Sampling frequency is 128 Hz with electrode impedance maintained below 5 kΩ. Simultaneously, EEG raw data were transmitted to the processing device. The same processing device was used throughout the signal analysis to maintain consistency.

Electrode from channel A positive was connected to the right forehead and channel B positive to the left forehead. Electrodes from channel A and B negative were connected to the right ear lobe. Finally, electrode from the reference port of the Wave Rider was also connected to the left earlobe.

Figure 3 shows a sample undergoing EEG experiment. In the picture, the sample is in the second stage where the brainwaves are captured during usage of the mobile phone. The phone is activated by making a call to another party; however both parties remain quite for five minutes which means there is no conversation.

Figure 3: A sample is undergoing the EEG measurement

V. RESULTS AND DISCUSSION

The analysis is divided into two parts which are the interview results and the EEG analysis.

A. Analysis of the Interview

Figure 4 shows the percentage of the samples that have dialed or received calls per day. From the interview session, the result shows that 54% of the samples have less than two calls in a single day. On the other hand, only 10% of the samples dialed or received calls 5 to 6 times daily.

The percentage of duration per call is depicted in Figure 5. Only 3% of the samples having the highest period of call which is more than 1 hour per call. 40% of the samples spend less than 5 minutes per call. The majority spend five to twenty minutes per call.

The interview session also gave the percentage of samples that have dizziness after having a conversation for a long period of time as shown in Figure 6. It indicates that 60% of samples having dizziness and depression symptom if they use the phone for more than one hour. While the rest, (12 samples) did not feel dizziness. The findings support previous results, which reported that people exposed to mobile phone showed increased levels of exhaustion and depression [16].

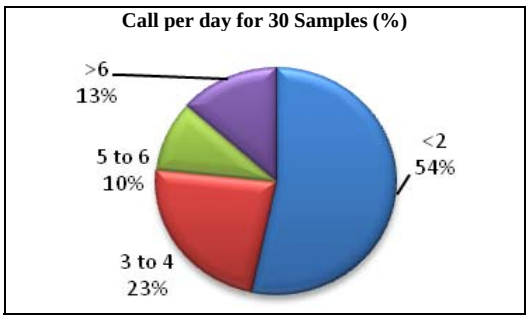


Figure 4: Number of calls per day dialed or received by samples

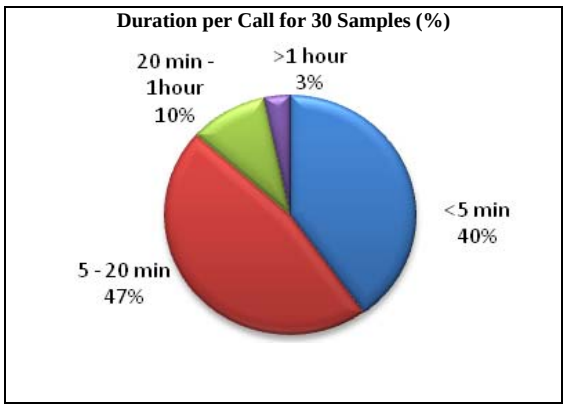


Figure 5: Duration per call

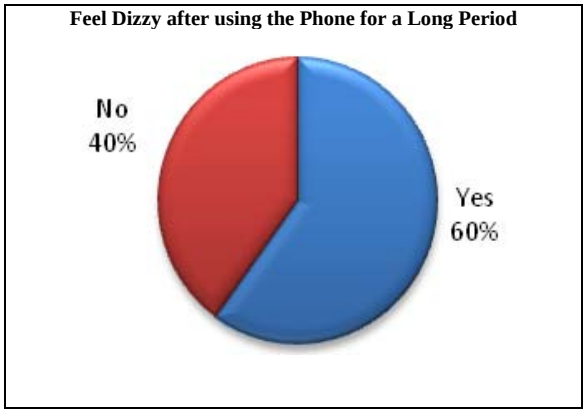


Figure 6: Feel dizzy effects experienced by samples after using the phone for more than 1 hour.

B. Analysis of the EEG Signals

The EEG signals analysis is divided into three stages: before call, during call and after call. The analysis was performed using statistical tools, SPSS version 17.0. The analysis of the data from EEG test focused on the comparison between the three stages and also the correlation between left and right brainwaves for each band individually as well as for overall.

Figure 7 clearly depicts that the alpha level of the right side decreases significantly during call and further decreases within the period of five minutes after call. However, the alpha level of the left side remains consistent throughout the experiment. It also shows that the waves for alpha left almost overlap for all stages.

Figure 8 shows the graph for beta right and beta left brainwaves. For beta right, the mean amplitude levels almost overlap in all stages for before, during and after call with 1.87% increment during mobile phone usage and 0.46% decrement after the exposure. Compared to beta left, the mean amplitude levels increases from 11.94% to 13.65% in between the activities.

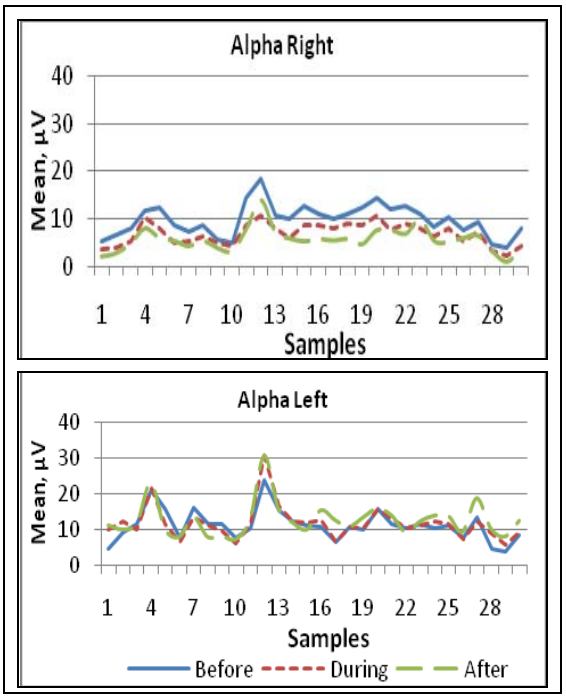


Figure 7: Comparison of alpha right and alpha left for before, during and after call

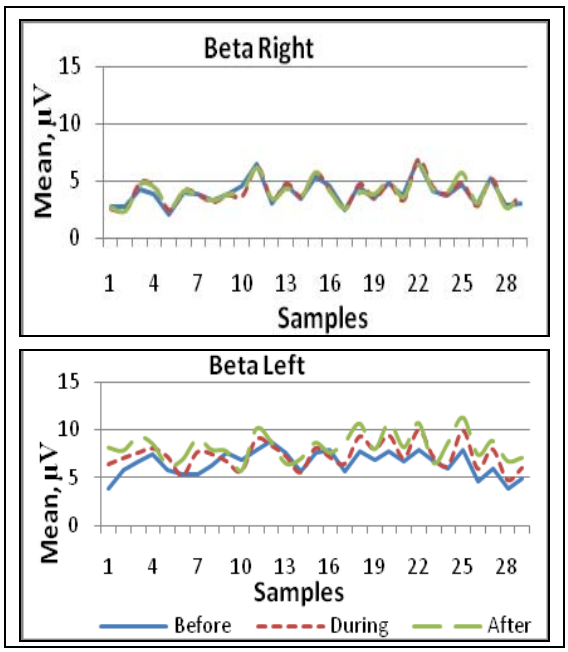


Figure 8: Comparison of beta right and beta left for before, during and after call

Figure 9 depicts the delta level of all stages for the laboratory experiment. There are no significant changes of the brainwaves due to the effects of exposure for both sides of the right and the left brain. The brainwaves pattern is also similar with theta waves as shown in Figure 10. Overall, delta and theta bands show slightly higher levels during and after calls for both left and right brainwaves. This observation agree with previous results which found that after 20-40 seconds exposure to mobile phone, a slow wave activity (2.5-6 Hz) appeared in the contra lateral frontal and temporal areas and suggested that mobile phones may affect the human brain [23]. It is concluded that mobile phones may reversibly influence the human brain, inducing abnormal slow waves in EEG of an awaken person [23].

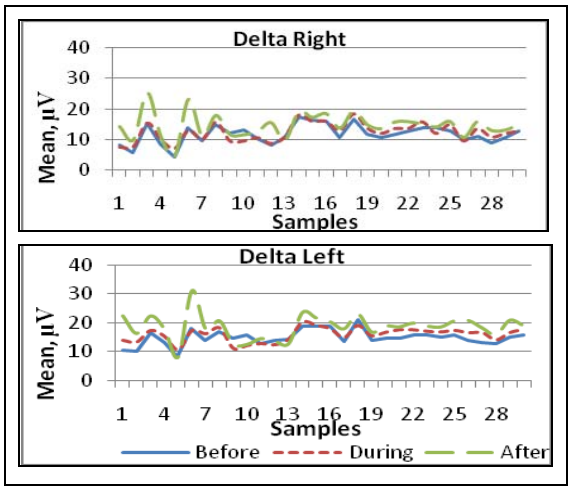


Figure 9: Comparison of delta right and delta left for before, during and after call

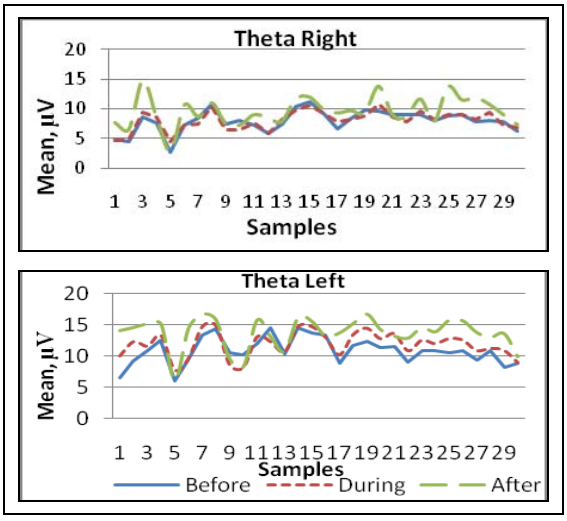


Figure 10: Comparison of theta right and theta left for before, during and after call

Table 2: Brainwaves mean value between stages

Band	Right Brainwaves (µV)			Left Brainwaves (µV)		
	Before	During	After	Before	During	After
Alpha	9.67	6.81	5.76	11.18	11.71	12.4
Beta	4.28	4.36	4.34	6.87	7.69	8.74
Theta	7.96	8.05	9.62	10.92	11.93	13.7
Delta	11.77	12.39	14.58	14.92	16.11	18.58

Table 2 shows the mean values of left and right brainwaves. It is known that alpha wave increase during relax and closed eyes. The left alpha brainwave increases slightly from 11.18 micro volts (before call) to 11.71 micro volts (during call) and further increases to 12.35 micro volts (after call). Thus, the increment of alpha left in the result shows samples are more relaxed [17].

In contrast, significant result shows that mean values for alpha right decreases during call, from 9.67 micro volts to 6.81 micro volts and further decreases to 5.76 micro volts after call. This decrement could be due to the effect of radiation from the mobile phone attached to the right ear during call. This observation is in line with findings from [18] which concluded that in the case of a person using a mobile phone, most of the heating effect will occur in the surface of the head, causing its temperature to increase by a fraction of a degree. Once the temperature is back to normal, there will be no radiation effects. Scientist have shown that this radiation might cause human biological damage through heating effects since human body is made up of approximately 65-70% water, electrolytes and ions [19].

Further analysis using paired T-test was carried out to compare between right and left brainwaves for each stage to deduce the correlation and brainwave balancing as shown in Figure 11. Correlation values for all frequency bands decreases during calls and further decreases five minutes after calls. The most significant results occur in alpha wave. The results agree with another research which found that EEG spectral power was influenced in some bins of the alpha band [20]. This effect was greater when the Electromagnetics Field (EMF) was on during the EEG recording session than before it [20]. It follows that the correlation between the left and the right alpha brainwaves signal decreases significantly from 0.983 (before call) to 0.824 (during call) and further decreases to 0.741 within the period of five minutes after call as shown in Figure 11. Initially, the alpha brainwaves were highly balanced before call. However, due to usage of mobile phone on the right side, the brainwaves became left dominant.

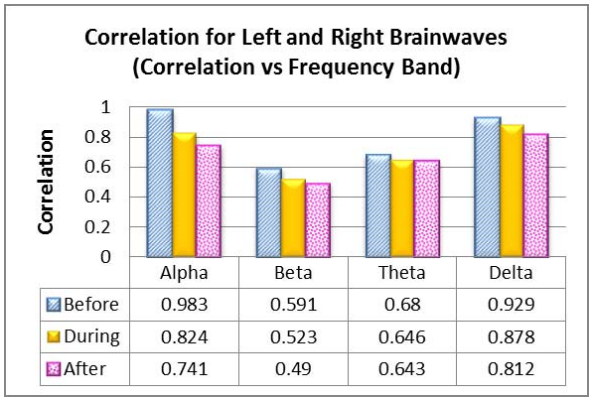


Figure 11: Correlation for left and right brainwaves

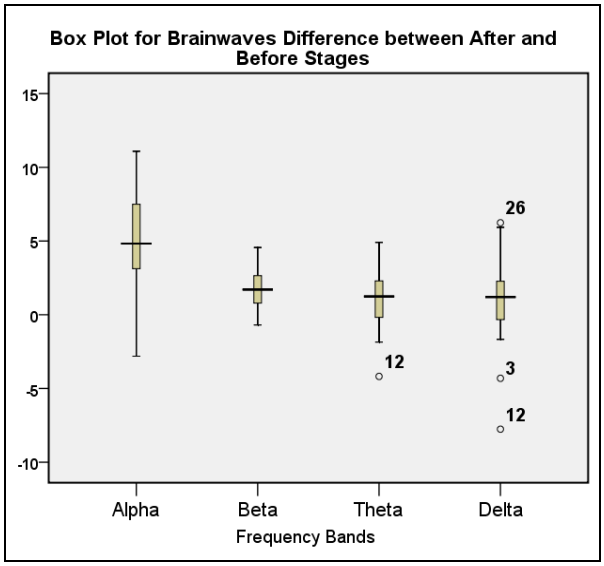


Figure 12: Box plot for brainwaves difference between after and before stages

Figure 12 shows the box plot for brainwaves difference between after and before call stages taking the left brainwaves and subtracting the right brainwaves. Results are consistent for alpha, beta, delta and theta in terms of domain data. The contributions of the outliers were insignificant; therefore, we decided not to remove the outliers from the analysis. The means of each box is above zero indicating that left brainwaves is dominant compared to the right for all frequency bands.

VI. CONCLUSION

By conducting this research, the effects of mobile phone usage on human brainwaves were studied using EEG. It was found that the alpha level of the right side (where the phone is attached) decreases significantly during the calls and further decreases within the period of five minutes after the calls were ended. However, the alpha level of the left side remains consistent throughout the experiment. The result from this research somewhat agrees with findings from [10-11], that is alpha wave decreases while exposure to radiation. Whereas, other frequency bands of the brain signals increased for both left and right side. It follows that the overall correlation between left and right brainwaves signal for all frequency bands decreases significantly during the calls and further decreases within the period of five minutes after calls. Therefore it reduces brainwaves balance. There is evidence that the usage of mobile phones affect the brainwaves.

In future, this research could be expanded to observe whether usage of mobile phones has some effects on other physiological variables such as heart beat, heart condition and lung condition. In addition, the observation can also be extended by attaching the phone to the left ears. Another interesting experiment is to compare the effect on the brainwaves due to different types of mobile phone (2G, 3G, iPhones, Samsung, Nokia etc.).

ACKNOWLEDGMENT

The writers would like to express gratefulness to the staff at Biomedical Research and Development Laboratory for Human Potential, Faculty of Electrical Engineering, UiTM. Our gratitude also goes to the Research Management Institute UiTM and the government of Malaysia.

REFERENCES

[1] http://www.who.int/mediacentre/factsheets/fs193/en/index.html, Fact sheet No.193, “Electromagnetic fields and public health: mobile phones”, May 2010.

[2] S. Kovach, “Report: The Hidden Dangers of Cell Phone Radiation”, Life Extension Magazine, August 2007.

[3] D.L. Hamblin and A.W. Wood, “Effects of Mobile Phone Emissions on Human Brain Activity and Sleep Variables”. International Journal of Radiation Biology, vol. 78, no. 2, pp. 659-669, 2002.

[4] D. Mosher, “Cellphone Radiation Increases Brain Activity”, Wired Magazine, New York, February 22, 2011.

[5] C. Sage and D.O. Carpenter, “Public Health Implications of Wireless Technologies”, Journal of Pathophysiology, vol. 16, no. 2, pp. 233-246, August 2009.

[6] E. Niedermeyer and F.H. Lopes da Silva. “Electroencephalography: Basic Principles, Clinical Applications and Related Fields”, 3rd edition, Lippincott, Williams & Wilkins, Philadelphia, 1993.

[7] A. S. Gevins and A. Remond, “Methods of Analysis of Brain Electrical and Magnetic Signals”, Elsevier, Amsterdam. 1987.

[8] M. Teplan, “Fundamentals of EEG Measurement”, Measurement of Science Review, vol. 2, no. 2, 2002.

[9] S. Sanei and J.A Chambers, “EEG Signal Processing”, Wiley, England, 2007.

[10] M. Putten, "Essentials of Neurophysiology: Basic Concepts and Clinical Applications for Scientists and Engineers”, Book. Series: Series in Biomedical Engineering, Springer Publishing Company Incorporated, United States, 2009.

[11] A. Ahlbom, A. Green, L. Kheifets, D. Savitz and A. Swerdlow, “Epidemiology of Health Effects on Radiofrequency Exposure”,

Environmental Health Perspectives, vol. 112, no. 17, pp. 1741–1754, 2004.

[12] D. G. Brunet and G. B. Young, "Co-Chair Electroencephalography Task Force," Guidelines for Clinical Practice and Facility Standards Electroencephalography. The College of Physicians and Surgeons of Ontario, Canada, 2000.

[13] K. Mann and J. Röschke, "Sleep under Exposure to High-Frequency Electromagnetic Fields," Sleep Medicine Reviews, vol. 8, pp. 95, 2004.

[14] H. D'Costa, "Influence of mobile phone electromagnetic field exposures on nervous function in the human brain and heart." School of Electrical &Computer Engineering Portfolio, RMIT, December 2008.

[15] M. M. R. Moussa, "Review on health effects related to mobile phones. Part II: results and conclusions," The Journal Of The Egyptian Public Health Association, vol. 86, p. 79.

[16] A. Johansson, S. Nordin, M. Heiden and M. Sandstrom, "Symptoms, Personality Traits and Stress in People with Mobile Phone-related Symptoms and Electromagnetic Hypersensitivity", Journal of Psychosomatic Research, vol. 68, no. 1, pp. 37-45, Sweden, 2010.

[17] J. L. Cantero, M. Atienza and R. M. Salas, "Human Alpha Oscillations in Wakefulness, Drowsiness Period, and Rem Sleep: Different Electroencephalographic Phenomena within the Alpha Band", Neurophysiologie Clinique/Clinical Neurophysiology, vol. 32, no. 1, pp. 54-71, 2002.

[18] P. Wainwright, "Thermal Effects of Radiation from Cellular Telephones", Physics in Medicine and Biology, vol. 45, no. 8, pp. 2363, 2000.

[19] D.A.A. Mat, F. Kho, A. Joseph, K. Kipli, S. Sahrani, K. Lias and A.S.W. Marzuki, "The Effect of Headset and Earphone on Reducing Electromagnetic Radiation from Mobile Phone Towards Human Head", Information and Telecommunication Technologies (APSITT), 2010 8th Asia-Pacific Symposium on, pp.1-6, 15-18 June 2010.

[20] G. Curcio, M. Ferrara, F. Moroni, G. D'Inzeo, M. Bertini and L. De Gennaro, "Is the Brain Influenced by a Phone Call?: An EEG Study of Resting Wakefulness", Neuroscience Research, vol. 53, no. 3, pp. 265-270, November 2005.

[21] M. N. T. Zunairah Haji Murat, Zodie Mohamed Hanafiah, Ros Shilawani S. Abdul Kadir and Husna Abdul Rahman " "Comparison of Brainwave Signals Between Electrical Engineering Students and Sport Science Students of Universiti Teknologi MARA Using EEG"," in 4th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, Malaysia., March 7-9, 2008.

[22] Z. H. Murat, M. N. Taib, S. Lias, R. S. S. A. Kadir, N. Sulaiman, and M. Mustafa, "The conformity between brainwave balancing index (BBI) using EEG and psychoanalysis test," neurophysiology, vol. 3, p. 6.

[23] A.V. Kramarenko, and U. Tan, "Effects of High-Frequency Electromagnetic Field on Human EEG: A Brain Mapping Study", International Journal Neuroscience, vol. 113, no. 7, pp. 1007-1019, 2003.

Classification of EEG Spectrogram Image with ANN approach for Brainwave Balancing Application

Mahfuzah Mustafa[1,2]
[1]Faculty of Electrical & Electronics
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia
mahfuzah@ump.edu.my

Mohd Nasir Taib[2], Zunairah Hj Murat[2],
Norizam Sulaiman[1,2], Siti Armiza Mohd Aris[2]
[2]Faculty of Electrical Engineering
Universiti Teknologi MARA
40450, Shah Alam, Selangor, Malaysia
dr.nasir@ieee.org

***Abstract*—In this paper, an Artificial Neural Network (ANN) algorithm for classifying the EEG spectrogram images in brainwave is presented. Gray Level Co-occurrence Matrix (GLCM) texture feature from the EEG spectrogram images have been used as input to the system. The GLCM texture feature produced large dimension of feature, therefore the Principal Component Analysis (PCA) is used to reduce the feature dimension. The result shows that the proposed model is able to classify EEG spectrogram images with 77% to 84% accuracy for three classes of brainwave balancing application with an optimized ANN model in training by varying the neurons in the hidden layer, epoch, momentum rate and learning rate.**

Keywords – EEG, spectrogram image, GLCM, PCA, ANN

I. INTRODUCTION

Artificial neural network (ANN) is inspired from the human brain by mimicking the action of neurons in the brain. ANN is a popular and powerful algorithm in data mining. The best performance can be achieved by varying the weights during training process. The feed forward training algorithm is normally used for the ANN in Electroencephalogram (EEG) analysis and has been proven by many researchers to be a great tool for classification, recognition and prediction in the EEG application [1-3]. According to these research findings, it denotes a promising result in the biomedical field. Nevertheless, the use of ANN as a classifier in balancing the brainwave never has been reported via literature. However, this paper will introduce an application of ANN for the balance brainwave application. Brainwaves are grouped into four bands identified as Delta, Theta, Alpha and Beta frequency bands[4]. Delta is the lowest frequency band with the highest amplitude while Beta is the highest frequency band with the lowest amplitude. Human brain is divided into two main regions which are the right and left hemisphere. The right hemisphere is superior in thinking, remembering, perceiving, understanding and emoting whereas the left hemisphere is dominant in activities involving analysis, arithmetic, language and speech [5, 6]. Balance brainwave is using both the right and left hemisphere of brain simultaneously. Balanced thinking that simultaneously uses both right and left will lead to a balanced life and lead to better health [6, 7].

II. RELATED WORK

EEG is an example of biosignal and other biosignal are electrocardiogram (ECG), electromyogram (EMG) and magnetoencephalogram (MEG). The EEG signals are characterized by the amplitude (voltage) and frequency. The frequency varies in each band, the Delta ranges within 0.5 to 4 Hz, Theta ranges from 4 to 8 Hz, Alpha ranges from 8 to 13 Hz and Beta ranges from 13 to 32 Hz [8]. However, the raw EEG signals need to be analysed in order to extract useful information for specific research.

Generally, the EEG signals are processed using the signal processing approach which extracted based on the time and frequency. The EEG signal is collected in time based and to transform this signal into frequency based, the Fourier Transform (FT) will be employed in this signal. The EEG signal also can be processed using image processing approach via the time-frequency based. The Short Time Fourier Transform (STFT) is one of the popular techniques to process signal through time-frequency based. The STFT is to perform an FT on the signal, then mapping the signal into a two-dimensional function of frequency and time.

There are a few researches using image processing technique in biosignal. However, there is an example of using image processing technique in analysing the ECG signal. The spectrogram image was produced using STFT in order to recognize heart abnormalities in the ECG waveform [9]. Next, the spectrogram images need to be further analysed, for example by using texture analysis. Gray Level Co-occurrence Matrix (GLCM) is a popular technique in various applications such as wood, satellite and ultrasound. There is a study that uses the GLCM as a texture analysis to

detect sleep disorder breathing in the ECG signal [10]. Usually, the Principal Component Analysis (PCA) is used for data reduction, classification and regression and it has been reported elsewhere. There is a study that uses the PCA for data reduction [11]. The PCA chooses three components out of the eight components from the GLCM texture feature. The result demonstrates that the three components give better accuracy than the eight components.

ANN history begins in the year 1940's and was initiated by McCulloch and Pitts'. However, it was popular in the 1980s [12]. ANN is actually a mathematical model to solve a variety of problems in control, prediction, pattern recognition, and optimization. There are several issues in the ANN design, including the number of training samples, activation function, learning parameters, and network model and size. Nevertheless, there is no general guideline to choose the best ANN architecture for a particular size of the training. Training unconstrained networks using standard performance measures such as the mean squared error may produce an unsatisfying result [13]. ANN is highly suited to process feature rich data [12-16]. There are studies using the extracted EEG signal features to be fed into the ANN in various applications. For example, the ANN is employed to analyzed the epileptic seizure [14], Parkinson disease [15] and brain-computer interface [16].

III. EEG Spectrogram Image and GLCM Analysis

An inspiration using time-frequency based is based on research in acoustic signal [17], sound [18], heart rate from ECG [19]. There is a study using time-frequency based approach in analyzing EEG signal in Brain Computer Interface (BCI) application [16]. Based on this research, time-frequency based have the same meaning as time-frequency representation, spectrogram image and lofargram but in this paper uses the term EEG spectrogram image. In this paper, uses a STFT to generate EEG spectrogram image for balanced brain application. After produced an EEG spectrogram image, GLCM is used to extract features. GLCM is a second order texture analysis in image processing. The GLCM is used comprehensively in analyzing images texture in applications such as satellite [20], acoustic signal [17], ultrasound [21], and as well as wood recognition [22]. This paper is improvement from the previous paper [23]. The previous paper emphasizes training process in ANN, meanwhile this paper employ ANN for training and testing process. In addition, number of training to testing ratio is evaluated at 70 to 30 and 80 to 20 to find the best model. In this paper, the EEG signal is processed by using image processing technique and confirmed by the brain dominance questionnaire. Next section will be described implementing the proposed method in the experiment.

IV. Methods

A. Subjects

The data collections were performed at Biomedical Research and Development Laboratory for Human Potential, Faculty of Electrical Engineering, Universiti Teknologi MARA Malaysia. The EEG signals were collected from 51 volunteers. The volunteers comprised of 28 males and 23 females with the mean age of 21.7. All volunteers were in healthy condition and did not consume any medication prior to the test. This study was approved by the ethics committee from Universiti Teknologi MARA.

B. EEG Measurement

The EEG data were collected with 2-channel electrodes Fp1 and Fp2 and reference to earlobes A1, A2 and Fz .The electrodes using gold disc with 256Hz sampling rate and the connections are in accordance to 10-20 International system. The EEG signal was recorded for five minutes using the g.MOBIlab, with wireless EEG equipment. The setup of EEG measurement is shown in Fig. 1. The impedance was maintained below 5kΩ using Z-checker equipment. Prior to the EEG recording, volunteers have to answer the eleven items Brain Dominance Questionaire [24]. Once the questionaire is completed, the score is calculated to determine the index of each sample. This index is produced from the previous experiment [25]. Table I shows the sample per index. Index 3 is for moderately balanced brain, Index 4 is for balanced brain and Index 5 is for highly balanced brain. Data for Index 1 and Index 2, corresponding to the unbalanced brain and less balanced brain, respectively are not available.The data was collected and processed by using the MATLAB program.

Figure 1. EEG measurement set up.

TABLE I. Data sampel per index.

Index	Description	Samples
Index 3	Moderately balanced brain	9
Index 4	Balanced brain	37
Index 5	Highly balanced brain	5

C. EEG Signal Pre-processing

Fig. 2 denotes the flow diagram for the EEG signal analysis from EEG signal collection up to process classification using the ANN. EEG signal pre-processing includes the

artifact removal and band pass filter. Artifacts occur when the volunteers blink his or her eyes. This artifacts were removed by the means of a program designed using MATLAB tools by setting a threshold value. The threshold was set to eliminate data when the values are less than -100μV and more than 100μV. The band pass filter was set for the frequency from 0.5Hz to 30Hz using Hamming window with 50% overlapping.

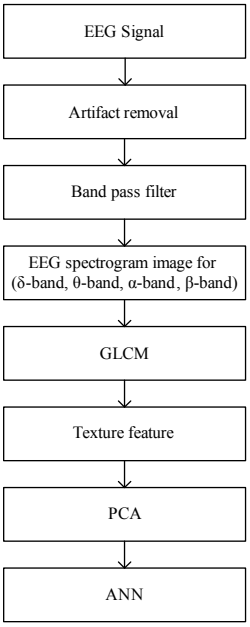


Figure 2. Flow diagram for EEG signal analysis.

D. Short Time Fourier Transform (STFT)

The STFT has produced spectrogram images for both Fp1 and Fp2 channels with image size 436 x 342. In STFT programming, each frequency band is set. The Beta band is set from 13Hz to 30Hz, Alpha band is set from 8Hz to 13Hz, Theta band is set from 4Hz to 8Hz and Delta band is set from 0.5Hz to 4Hz. The STFT is done by multiplying the Fourier Transform (FT) of the EEG signal by window function.

E. Gray Level Co-occurrence Matrix (GLCM)

There are parameters need to be set in GLCM, including the grey level, orientation and displacement. In [26] proposed grey level less than 64 and greater than 24 because grey level greater than 64 will produce an expensive computational cost whereas grey level below 24 will produce low accuracy. Most researchers employ all the four orientations (0^0, 45^0, 90^0 and 135^0) in their experiments [26, 27].The displacement, d=1 chosen by many researchers [27]. In this experiment, the grey level is set with 32, all four orientations (0^0, 45^0, 90^0 and 135^0) and displacement, d=1. Subsequently, the texture features were extracted for each GLCM. In this research, the texture feature is the combination of Haralick [27], Soh [28] and Clausi [26] technique. The 20 texture features are the Inverse difference normalized, Inverse difference moment normalized, Information of correlation 1, Information of correlation 2, Different variance, Different entropy, Sum average, Sum variance, Sum entropy, Maximum probability, Variance, Entropy, Homogeneity, Dissimilarity, Energy, Cluster prominence, Cluster shade, Autocorrelation, Contrast and Correlation.

F. Principal Component Analysis (PCA)

Output from the GLCM texture feature generates big matrices. In order to reduce big matrices, the PCA was employed to find optimum features. The optimum features will reduce the execution time for the classification process. The first principal components contain most of the useful information, and the last principal components contain mostly noise. Therefore, these last principal components can be removed without significantly affecting the information content of the GLCM texture feature.

G. Artificial Neural Network (ANN)

A feed-forward ANN was used to analysis the EEG spectrogram image and was trained using Levenberg-Marquardt algorithm [2, 14]. The system has an 8 inputs and 1 output. The best ANN model can be obtained by optimizing four parameters, namely the number of neurons in the hidden layer, epoch, momentum rate and learning rate [14] . The optimum parameters can be achieved by finding the highest accuracy and the lowest mean square error (MSE) [29]. Many studies refer to MSE as the error goal [29, 30]. In this experiment, the sigmoid was selected for the ANN activation function. The parameters to be optimized vary while the three parameters were fixed. Next, accuracy and MSE were observed and collected. Finally, the best model for the experiment was selected for the final application. This experiment uses two sets of data. The first set uses 70% of the data for training ANN and 30% of data for testing the ANN model. The second set using the ratio 80:20 for training and testing the ANN model.

V. RESULT AND DISCUSSION

Spectrogram images produced using the STFT are as shown in Figs. 3 (a)-(h). These figures illustrate the Delta band, Theta band, Alpha band and Beta band for both the Fp1 and Fp2 channels. Based on these figures, the spectrogram is texture shaped and each frequency band produces different texture. Each EEG sample will produce eight images for both channels Fp1 and Fp2. The number of spectrogram generated is shown in Table II.

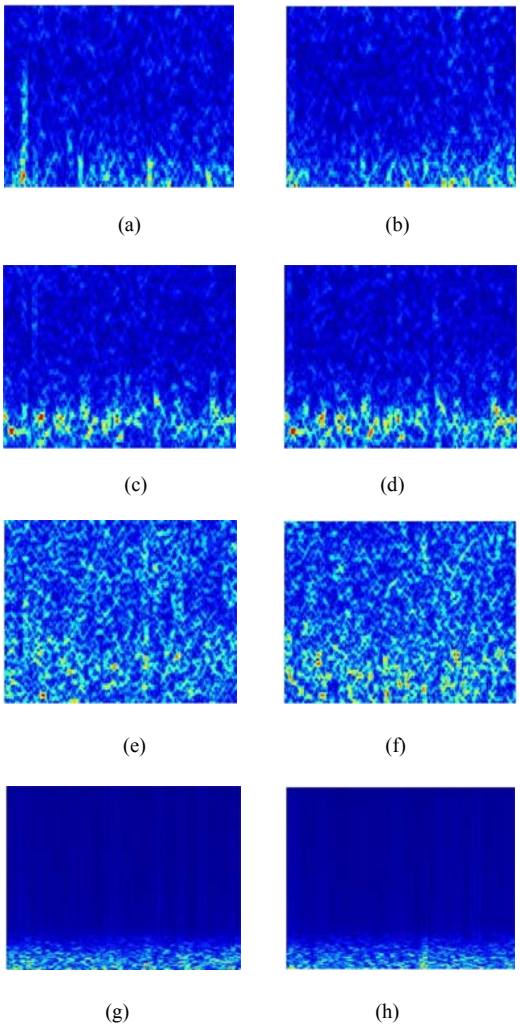


Figure 3. Spectrogram images for (a) Delta band from Fp1 channel (b) Delta band from Fp2 channel (c) Theta band from Fp1 channel (d) Theta band from Fp2 channel (e) Alpha band from Fp1 channel (f) Alpha band from Fp2 channel (g) Beta band from Fp1 channel (h) Beta band from Fp2 channel

TABLE II. NUMBER OF EEG SPECTROGRAM IMAGE.

Index	Samples	EEG spectrogram image
Index 3	9	72
Index 4	37	296
Index 5	5	40
TOTAL	51	408

The GLCM was generated for grey level=32, matrix orientations for 0^0, 45^0, 90^0 and 135^0, with displacement=1 for each spectrogram image and then texture feature from the combination of Haralick , Soh and Clausi were extracted. Eighty GLCM texture features were extracted and PCA is used to reduce this data dimension. Fig. 4 shows the percentage of eigenvalue produced by 80 principal components. The graph shows that the percentage gradually decreases until the last components. The first components show the highest percent, with 70% eigenvalue of covariance from original data. Some components have been chosen for the purpose of classification based on the results of PCA, and 8 principal components were selected because they produced a high percentage of eigenvalue.

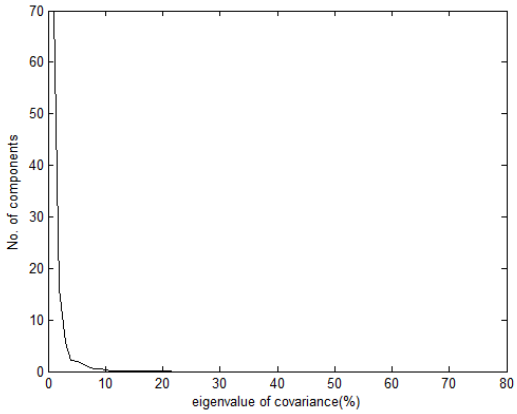


Figure 4. Graph of eigenvalue in percent

Performance of optimization of the ANN is presented in Figs. 5 to 8 for data ratio 70:30. In the figures, legend 'solid' line and 'dot' line represents mean squared error and accuracy percentage. Fig. 5 illustrates the result for optimizing the number of neurons in the hidden layer size. In the figure, the 'solid' line shows a decreasing trend with respect to the number of neurons, while the 'dot' line shows an increasing line with respect to the number of neurons. It was found that the hidden layer 24, 22, 20, 18, 13, and 10 may produce good prediction outcome. In this experiment, the network with hidden layer size 10 with accuracy rate 88.5% with MSE 0.0598 was selected.

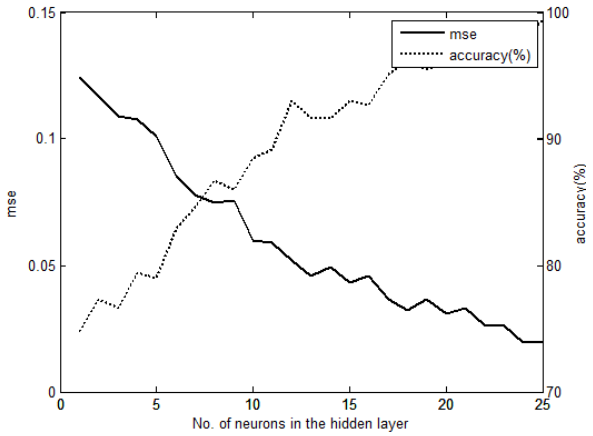


Figure 5. Training performance and prediction accuracy with varying hidden layer size

Fig. 6 shows the result of the finding of the optimum epoch. From this figure, it was found that the epoch value of 10000 and 50000 may produced good outcome. The epoch 10000 was found to be optimum with an accuracy of 88.81% with MSE 0.0937.

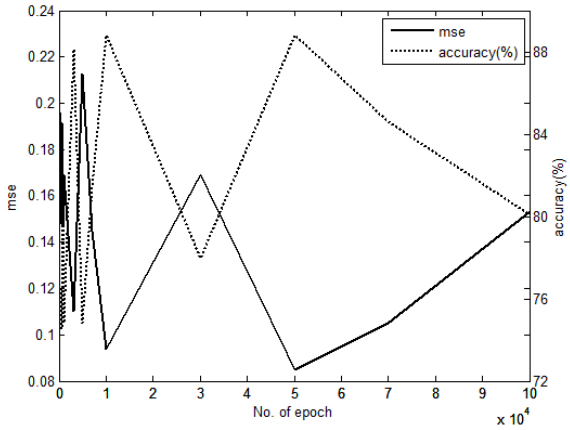


Figure 6. Training performance and prediction accuracy with varying epoch.

Fig. 7 illustrates the result of the finding of the momentum rate. From the figure, 'solid' line shows a decreasing trend until it reaches 0.3 momentum rate, at this point the trend started to increase gradually. The 'dot' line gradually decreases until it reaches 0.9 momentum rate. The figure shows a learning rate of 0.2 and 1 may produce a good prediction outcome. The momentum rate of 0.2 was found to be the optimum accuracy 89.5% with MSE 0.0586.

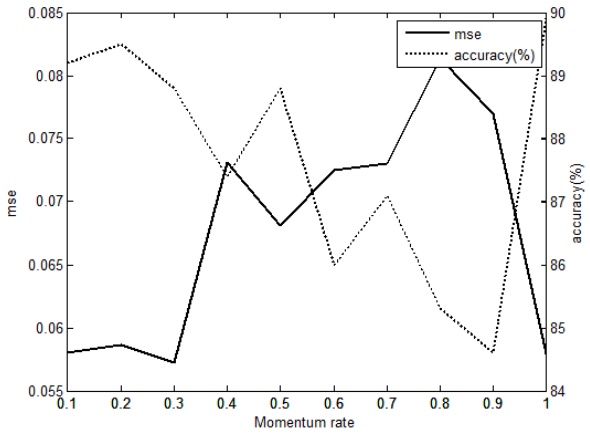


Figure 7. Training performance and prediction accuracy with varying momentum rate

Fig. 8 shows the result of the finding of the optimum learning rate. From this figure, it was found that the learning rate values of 0.2 and 0.6 may produce a good outcome with a lower point of MSE. The learning rate of 0.8 was found to be the optimum accuracy 89.2% with MSE 0.0606. Finally, the best network defined by the 10 hidden neurons, 10000 epoch, 0.2 momentum rate and learning rate of 0.6.

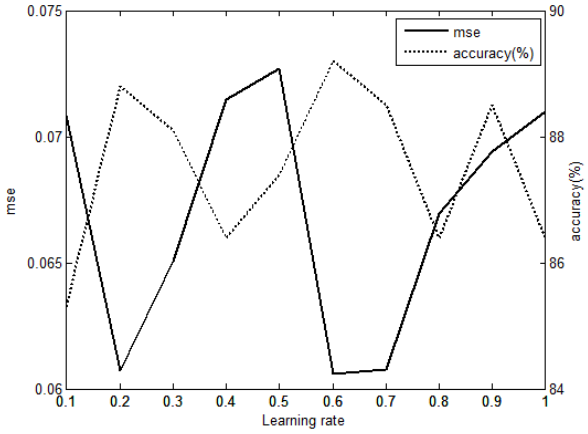


Figure 8. Training performance and prediction accuracy with varying learning rate

Table III illustrates the confusion matrix for the EEG spectrogram classification after testing using the ANN with optimized parameters. From the figure, legend I3, I4 and I5 represent Index 3, Index 4 and Index 5. From this table, accuracy for the EEG spectrogram according to the Index 3 to Index 5 is 77%.

TABLE III. CONFUSION MATRIX FOR ANN TESTING RESULT FOR DATA RATIO 70:30.

	Index 3	Index 4	Index 5
Index 3	17	6	1
Index 4	9	68	5
Index 5	0	7	9

Performance of optimization of the ANN is presented in Figs. 9 to 12 for data ratio 80:20. Again, the legend 'solid' line and 'dot' line represent the mean squared error and accuracy percentage. Fig. 9 illustrates the result for the optimizing number of neurons in the hidden layer. In the figure, it was found that the hidden layer 15, 22, 24 and 25 may produce a good prediction outcome. In the experiment, the network with hidden layer 22 with an accuracy rate of 94.21% with 0.0421 MSE was selected.

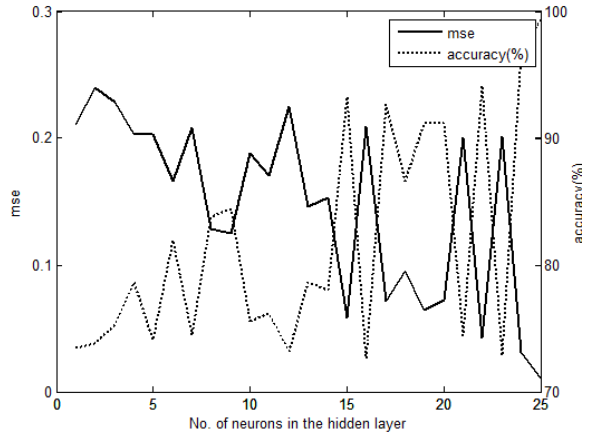


Figure 9. Training performance and prediction accuracy with varying hidden layer size

Fig. 10 shows the result for the finding optimum epoch. From this figure, it was found that epoch value of 3000, 30000 and 100000 may produce a good prediction outcome. The epoch of 3000 was found to be optimum with an accuracy of 78.5% with MSE 0.1379.

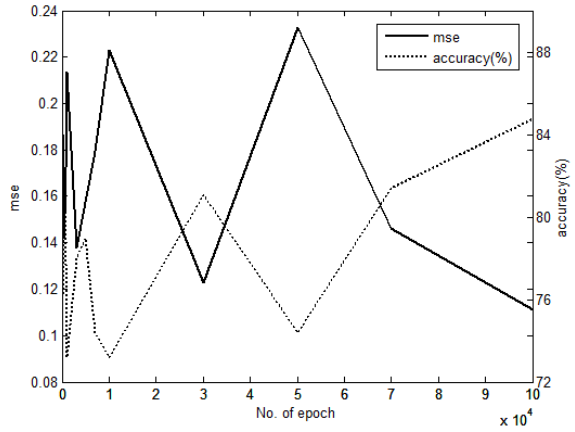


Figure 10. Training performance and prediction accuracy with varying epoch

Fig. 11 illustrates the result for the finding optimum momentum. From the figure, 'solid' line gradually decreases until reaches 0.5 momentum rates, at this point the trend started to increase. The 'dot' line reaches highest peak at point 0.5 momentum rate. In the figure shows momentum rate of 0.3 and 0.5 may produce a good prediction outcome. The momentum rate of 0.5 was found to be optimum accuracy 85.06% with MSE 0.1109.

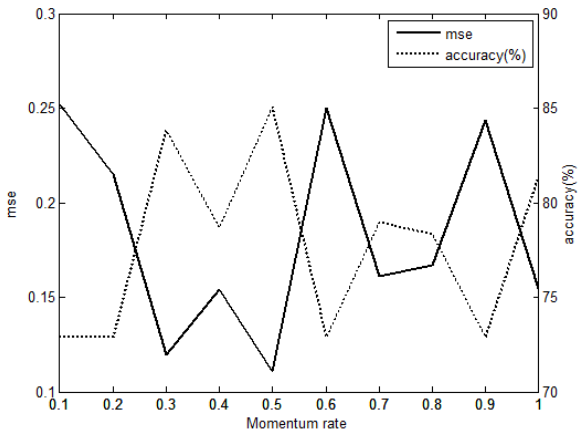


Figure 11. Training performance and prediction accuracy with varying momentum rate

Fig. 12 presents the result of finding the optimum learning rate. From the figure, it shows that the learning rates 0.6 and 0.9 may produce a good prediction outcome. The learning rate of 0.9 was found to be optimum accuracy 84.15% with MSE 0.1185. Eventually, the best network defined by 22 hidden neurons, 3000 epoch , 0.5 momentum rate and learning rate of 0.9.

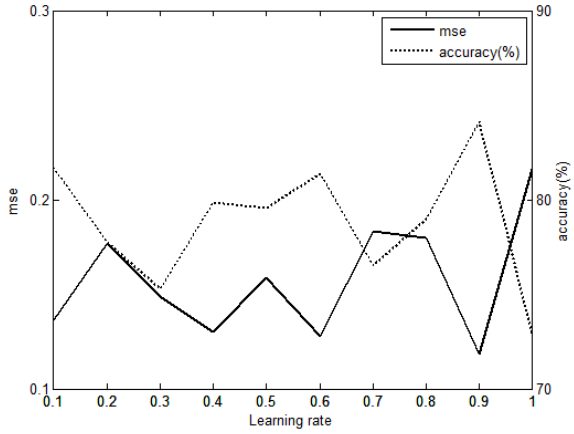


Figure 12. Training performance and prediction accuracy with varying learning rate

Table IV illustrates the confusion matrix for the EEG spectrogram classification after testing using the ANN with optimized parameters. Again, the legend I3, I4 and I5 represent Index 3, Index 4 and Index 5. From this table, the accuracy for the EEG spectrogram according to the Index 3 to Index 5 is 84%. Based on Table 2 and Table 3, the 80:20 ratio data give a higher percentage of accuracy than the

70:30 ratio data. It is therefore accepted that the training set should be larger than the testing set to obtain a higher percentage of accuracy.

TABLE IV. CONFUSION MATRIX FOR ANN TESTING RESULT FOR DATA RATIO 80:20.

	Index 3	Index 4	Index 5
Index 3	16	0	0
Index 4	4	44	8
Index 5	0	1	7

VI. CONCLUSION

In this paper, the classification using the ANN algorithm is presented with the aim to classify the EEG spectrogram as a moderate balanced brain, balanced brain and high balance brain. In order to achieve good result, the ANN model were optimized in training phase by varying the neurons in the hidden layer, epoch, momentum rate and learning rate. The accuracy rate is between 77% to 84%. The experimental result also shows that the PCA is able to reduce the original GLCM texture feature data.

ACKNOWLEDGMENT

The author would like to acknowledge the members of Biomedical Research Laboratory for Human Potential, FKE, UiTM for their support and assistance and UMP for the studentship of Mahfuzah Mustafa

REFERENCES

[1] A. Vuckovic, V. Radivojevic, A. C. N. Chen, and D. Popovic, "Automatic recognition of alertness and drowsiness from EEG by an artificial neural network," *Medical Engineering & Physics,* vol. 24, pp. 349-360, 2002.

[2] A. Subasi, A. Alkan, E. Koklukaya, and M. K. Kiymik, "Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing," *Neural Networks,* vol. 18, pp. 985-997, 2005.

[3] M. K. Kiymik, M. Akin, and A. Subasi, "Automatic recognition of alertness level by using wavelet transform and artificial neural network," *Journal of Neuroscience Methods,* vol. 139, pp. 231-240, 2004.

[4] M.Teplan, "Fundamental of EEG measurement," *Measurement Science Review,* vol. 2, pp. 1-11, 2002.

[5] R. Sperry, "Left-brain, right-brain," *Saturday Review,* vol. 2, pp. 30-3, 1975.

[6] R. W. Sperry. (1981, December 8) Some Effects of Disconnecting The Cerebral Hemispheres. *Division of Biology, California Institute of Technology, Pasadena*. 1-9.

[7] P. J. Sorgi, *The 7 systems of balance: a natural prescription for healthy living in a hectic world*. Deerfield Beach, Florida: Health Communicatios, Inc., 2001.

[8] D. Cvetkovic, "Electromagnetic and audio-visual stimulation of the human brain at extremely low frequencies," PhD Thesis, RMIT University, 2005.

[9] M. Saad, M. Nor, F. Bustami, and R. Ngadiran, "Classification of Heart Abnormalities Using Artificial Neural Network," *Journal of Applied Sciences,* vol. 7, pp. 820-825, 2007.

[10] A.-A. Mohammad, B. Khosrow, J. R. Burk, E. A. Lucas, and M. Manry, "A New Method to Detect Obstructive Sleep Apnea Using Fuzzy Classification of Time-Frequency Plots of the Heart Rate Variability," in *Proceedings of the 28th IEEE EMBS Annual International Conference*, 2006, pp. 6493-6496.

[11] H. Murray, A. Lucieer, and R. Williams, "Texture-based classification of sub-Antarctic vegetation communities on Heard Island," *International Journal of Applied Earth Observation and Geoinformation,* vol. 12, pp. 138-149.

[12] A. K. Jain, M. Jianchang, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer,* vol. 29, pp. 31-44, 1996.

[13] M. Egmont-Petersen, D. de Ridder, and H. Handels, "Image processing with neural networks-a review," *Pattern Recognition,* vol. 35, pp. 2279-2301, 2002.

[14] K. P. Nayak, T. K. Padmashree, S. N. Rao, and N. U. Cholayya, "Artificial Neural Network for the Analysis of Electroencephalogram," in *Proceedings of the Fourth ICISIP Fourth International Conference*, 2006, pp. 170-173.

[15] R. Rodrigues, P. Miguel, T. Teixeira, and J. Paulo, "Classification of Electroencephalogram signals using Artificial Neural Networks," in *Proceedings of the 3rd BMEI International Conference*, pp. 808-812.

[16] H. Dongmei, Z. Hongwei, and Y. Naigong, "High Resolution Time-Frequency Analysis for Event-Related Electroencephalogram," in *Proceedings of the Sixth WCICA World Congress*, 2006, pp. 9473-9476.

[17] K. S. Thyagarajan, T. Nguyen, and C. E. Persons, "An image processing approach to underwater acoustic signal classification," in *Proceedings of the IEEE Computational Cybernetics and Simulation International Conference*, 1997, pp. 4198-4203 vol.5.

[18] Y. Guoshen and J. J. Slotine, "Audio classification from time-frequency texture," in *Proceedings of the IEEE ICASSP International Conference*, 2009, pp. 1677-1680.

[19] O. H. Colak, "Preprocessing effects in time-frequency distributions and spectral analysis of heart rate variability," *Digital Signal Processing,* vol. 19, pp. 731-739, 2009.

[20] C. I. Christodoulou, S. C. Michaelides, and C. S. Pattichis, "Multifeature texture analysis for the classification of clouds in satellite imagery," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 41, pp. 2662-2668, 2003.

[21] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *Medical Imaging, IEEE Transactions on,* vol. 22, pp. 902-912, 2003.

[22] R. Bremananth, B. Nithya, and R. Saipriya, "Wood Species Recognition Using GLCM and Correlation," in *Proceedings of the ARTCOM International Conference*, 2009, pp. 615-619.

[23] M. Mustafa, M. N. Taib, Z. H. Murat, N. Sulaiman, and S. A. M. Aris, "The Analysis of EEG Spectrogram Image for Brainwave Balancing Application Using ANN," in *Proceedings of the 13th UKSim International Conference*, 2011, pp. 64-68.

[24] L. Mariani. (1996, 1 May 2010). *Brain-dominance questionaire*. Available: http://www.learningpaths.org/questionnaires/lrquest/lrquest.htm

[25] Z. H. Murat, M. N. Taib, S. Lias, R. S. S. A. Kadir, N. Sulaiman, and M. Mustafa, "The Conformity Between Brainwave Balancing Index (BBI) Using EEG and Psychoanalysis Test," *International Journal of Simulation Systems, Science & Technology,* vol. 11, pp. 86-92, 2010.

[26] D. A. Clausi and M. E. Jernigan, "A fast method to determine co-occurrence texture features," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 36, pp. 298-300, 1998.

[27] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 3, pp. 610-621, 1973.

[28] L. K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 37, pp. 780-795, 1999.

[29] N. F. Güler, E. D. Übeyli, and I. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert Systems with Applications,* vol. 29, pp. 506-514, 2005.

[30] T. Ah Chung and A. D. Back, "Locally recurrent globally feedforward networks: a critical review of architectures," *Neural Networks, IEEE Transactions on,* vol. 5, pp. 229-239, 1994.

Improvement of Bond Graph Model Based Diagnosis with Bayesian Networks Approach

Abdelaziz Zaidi, Moncef Tagina
Ecole Nationale des Ingénieurs de Tunis
Department of Electrical Engineering
BP 37, 1002 Belvédère TUNISIA
e-mail: Abdelaziz.Zaidi@isetso.rnu.tn,
Moncef.Tagina@ensi.rnu.tn.

Belkacem Ould Bouamama
Ecole Polytechnique de Lille
Cité scientifique
Dép. IMA F59655 Villeneuve d'Ascq
cedex, FRANCE
e-mail: Belkacem.Ouldbouamama@polytech-lille.fr

***Abstract* — The method of Bond Graph based Analytical Redundancy Relations in Fault Detection and Isolation is explicitly associated with components faults, this is due to architectural and functional aspect of the Bond Graph tool. This allows using the reliability of each component to improve the decision-making step. The purpose of this paper is the improvement of the classical binary method of decision-making, so that it can treat unknown and identical signatures of failures. This approach consists of associating the measured residuals and the components reliability data to build a Hybrid Bayesian Network. This network is used to determine the posterior probabilities of the failures. The developed methodology is applied to a real steam generator pilot process.**

Keywords – component, Bond Graph, diagnosis, Bayesian Networks, decision-making, reliability

I. INTRODUCTION

In a Model Based Diagnosis (MBD) approach, the methods of Fault Detection and Isolation (FDI) are based explicitly or implicitly on the generation of Analytic Redundancy Relations (ARR). The problem of FDI using ARR received a growing attention during the last years due to the persistent development of the power of computers. The generation of ARRs is based on two main approaches. The first one is direct; it consists in the elimination of all unknown variables keeping input-output relations involving only observable variables. Among these methods, one will find those of observers [31] and parity space [14]. The second approach is indirect; it estimates the states, outputs or parameters, in order to generate signals as difference between the variables and their estimates [17].

Graphical methods are based essentially on structural models, where the nodes of the graph are the system variables and system behavior equations, and links connect variable nodes to the equation nodes in which they appear, are well-suited for qualitative approaches to the diagnosis task. Typically these graph structures are independent of the numerical values of the system parameters. Furthermore, the graphical model structure is general, and accommodates relations that can be linear, non linear, or even expressed in table or rule format. The properties of the system model graph can be used to establish monitorability (i.e., which part of the system can be monitored) by studying the graph connectedness.

The main kinds of graphical tools can be cited: digraphs, bipartite graphs, signed directed graphs (SDG) and bond graphs. Comparing with other graphical methods, bond graph is also a graph $G(N,A)$ but the nodes N consists of generic physical elements and junctions and A is the interchanged power between them.

The Bond Graph (BG) tool invented in 1961 by Paynter [32], is a graph of structured bonds that facilitates the access to the modeling, the analysis and the simulation of physical systems. It is known as a multidisciplinary graphical language that permits the representation of the power transfers within a system [27]. From 1990, the graphical aspect of the bond graph has been initially exploited for control analysis (structural controllability and observability) [10]. Thereafter, it is widely used for the design of fault detection and isolation procedures using qualitative and causal analysis approaches [6] and quantitative approach to generate ARRs [36]. Specific software was developed for the generation of failure signature matrix (FSM) [30].

The step of ARRs generation is followed by the evaluation of the residuals and decision-making for robust fault detection and isolation. The decision rule may be based on a geometric method such as a simple threshold test on the instantaneous residual values or moving averages of the residuals, adaptive thresholds [37], interval models [4], or on cumulative sums [5] of residuals. Some decision rules are based on statistical methods, e.g. generalized likelihood ratio test or sequential probability ratio test [39].

The end result of analysis by the classic decision-making from the FSM is often binary (component is faulty or healthy). When the signature is unknown due to measurement noises and uncertainty of the model, the decision may not be feasible. Recently, in [12] the authors

applied robust FDI with respect to parameter uncertainties of the BG model. This last allows representing explicitly parameter uncertainties under multiplicative form for each bond graph element. But in real industrial process components can be degraded and this is a situation between the two states which can be associated to a continuous value in the interval [0, 1]. This value can be only the posterior value of the component reliability.

ARRs generated from bond graph models are explicitly associated with components faults. This is due to the architectural and functional aspect of the BG tool. When designing a supervision strategy, this allows an easy matching with the reliability of each component as an additional data for the diagnosis model. With the development of FDI algorithms, the decision of the diagnosis module should be more significant than a boolean one. When it becomes continuous in the interval [0, 1] (extreme values of a binary decision), the supervision module can treat some problems such as unknown signatures, or residuals corresponding to the signature of more than one fault. The efficiency of the FDI decision module is then ameliorated without increasing the number of sensors.

In this field, several papers have been published. The use of reliability data in FDI is introduced by [40] who proposed the improvement of decision making in ARR based approaches by using reliability data and Bayesian Networks (BN) [33]. The authors proposed a Dynamic Bayesian Network (DBN) with two kind of nodes; ones associated to the residuals and others to the failure of the components which have exponential probability distribution functions (PDF). By such method and for large systems, one will have a fastidious representation of the model. The approach supposes that ARRs are already generated, and it is not proposed for a specific generation method.

The DBN approach is also used for health monitoring in [18]-[35]. The structure of the BN is deduced from the Temporal Causal Graph (TCG) [25], which is a representation deduced from the BG model. Also by the same TCG representation, it is possible to perform qualitative reasoning for ARRs [6]. In cited papers the qualitative approaches did not take into account the uncertainties and did not reflect the real degradation of the component and cannot incorporate statistics and historical data because of its kind of model.

The innovative interest of presented paper consists of developing a methodology that extends the ARR BG model based approach to support reliability data to build an intelligent supervision strategy. The first FDI step (alarm generation) is performed by the BG model because of its causal and structural properties and the second step (decision procedure) is improved by introducing the reliability of each component to be monitored (associated with a BG element). The improvement of the decision-making step for the diagnosis module is realized through a Hybrid Bayesian Network (HBN) model that permits to calculate, by a hybrid inference procedure, the posterior probabilities of the components faults. This network is used in two distinct inference procedures: one for the continuous part and the other for the discrete part. The continuous nodes of the network are the prior probabilities of the components failures, which are used by the inference procedure on the discrete part to compute the posterior probabilities of the failures.

The paper is organized as follows: first, an overview on ARR based FDI approaches is given. The second section is devoted to the developed methodology where is presented briefly the BG approach and the Bayesian formulation for the decision-making. The fourth section is dedicated to an application on a steam generator pilot process. Fifth part concludes the paper.

II. State of the Art

A. Bond Graph methodology for FDI design

The key of bond graph modeling is the representation (by a bond) of power as the product of efforts (intensive variable) and flows (derivative of extensive variable) with elements acting between these variables and junction structures to put the system together. As shown on the Fig. 1, the bond graph symbol gives us four informations: the existence of physical link between two systems by the bond, the type of power (electric, mechanical...) by the power variables, the power direction by the half arrow and the causality by the stroke.

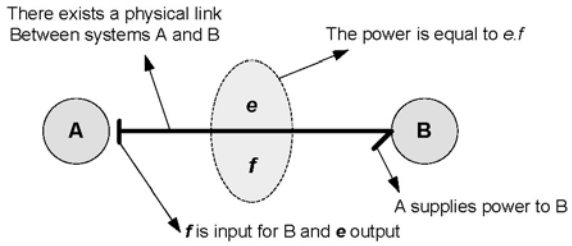


Fig. 1. Bond Graph representation

In bond graph methodology, physical phenomena and components are modeled by graphical symbols in a unified way for all the physical domains. *R* elements are used for passive energy dissipation phenomena, and *C* and *I* elements for passive energy storage ones. The junction elements 0, 1, TF, GY are used for connecting the passive elements; they compose the model structure and are power conservative. Sources of effort (*Se*) and sources of flow (*Sf*)

represent sources of energy. Sensors are represented by effort (*De*) and flow detectors (*Df*). The passive elements are described by generic constitutive equations: dissipative *R*-elements (electrical resistor, hydraulic friction …) are described by algebraic relationship $F_R(e,f)=0$, potential storage energy *C*-element (capacitor, tank, spring) are modeled by an integral equation linking effort and integral of flow $F_C(e,\int f(t)dt)$ and kinetic storage energy and *I*-element (mechanical inertia, electric coil…) is quantified by integral equation between integral of effort and flow $F_I(f,\int e(t)dt)$.

Although the method based on ARRs widely used, is one of the most important methods in model based FDI, this approach can inherit some problems, especially in the phase of conception. However, one can have identical signatures for different failures and it would be difficult or expensive to place a supplementary sensor to improve the isolation performance. Besides, it would also reduce the global reliability of the system. To overcome the problem of monitorability (ability to detect and isolate faults) of the sensors and the sources of control in BG based ARR approach, some methods are proposed in literature. The first method is based on material redundancy; it is based on the evaluation of the parameters and therefore it requires two sensors (one for flow and another for effort) for every component to supervise. The second is based on the notion of bicausal BG that permits to use the rest of the model to determine the values of effort and flow, without using the characteristic of the component [13]. Reference [12] proposed a FDI BG model for generation of robust ARRs. The method is based on unknown variables elimination using covering causal paths through the graph. However, the decision procedure is based on structural residuals in the Boolean Fault Signature Matrix (FSM).

In [8] an algorithm is presented to derive automatically temporal information in FSM for the set of possible conflicts to improve the isolation capabilities of BG based ARR approach. The approach uses Temporal Causal Graph (TCG) as an intermediate structure to generate the set of possible conflicts.

B. Bayesian Networks, Bond Graph and FDI

In the last decade, there is a growing common area between BN and FDI. A BN is a pair N=<(*V*,*E*),*P*>, where (*V*,*E*) are the nodes (vertices) and arcs (edges) of a directed acyclic graph (DAG) and *P* is a probability distribution on *V* [33]. Each node contains a random variable, and the directed edges between them define conditional dependence scales. Finally, in the last layer, BNs are used to describe the conditional dependence between faulty domains and fault signatures.

or independence among the random variables. In [24] a method was proposed for sensor fault detection and identification. It consists of using multi-stage BN to detect different sensor fault types (bias, drift and noise). This paper presents a method that reduces the size of required conditional probability data. Improving decision making in ARR based approaches using BN and reliability data is treated in [40]. The authors proposed a DBN (BN with two time slices; (*t-1*) and (*t*)) incorporating nodes with exponential failure distributions for the components to facilitate the expression of passing from slice (*t-1*) to slice (*t*). The approach supposes that ARRs are already generated, and it is not proposed for a specific generation method. The given approach is applied only for components whose distribution of failure is exponential. The structure of the network becomes more complex if the number of components increases since we need two time slices for every component.

Reference [18] have elicited the Dynamic Bayesians Networks for monitoring dynamic systems. It is pointed out that Hidden Markov Model (HMM) processes and Kalman filters are particular cases of DBN. The structure of the BN is deduced from the Temporal Causal Graph (TCG), which is a representation deduced from the BG model. Reference [2] studied the comparison between different filtering algorithms with DBN and noted the interest of particles filtering approach with a proposal distribution generated by an Unscented Kalman Filter (UKF) for networks with large size. In [35] a Bayesian approach is proposed for the monitoring of model parameters deviations. The elicited FDI architecture is an observer based on a DBN modeling the nominal operation of the system. The structure of the network is also deduced from the BG model. The inference algorithm is the Extended Kalman Filter (EKF) to treat the non linearities of the system. The authors used a qualitative reasoning from the TCG to generate the possible hypotheses of the failure. To achieve the isolation, a DBN incorporating discrete nodes is used to indicate the possible failures of the continuous parameters. Reference [41] addresses FDI in complex plants by using a hierarchical strategy involving different modeling approaches. The BG tool is proposed as a first physical domain layer. Thereafter, the principle component analysis (PCA) is used to reduce the data dimension and a discrete wavelet transform (DWT) is applied to abstract the dynamics of the plant at different

III. Bond Graph and Bayesian Networks for Reliable Methodology

A. Introduction

The growing interest to model based methods in FDI is essentially due to the fact that this kind of approaches does not require learning the model contrary to non-model based ones. Furthermore, because of its graphical, structural and causal properties BG tool is used for modeling and fault indicators generation based on covering causal path for unknown variables elimination (for more detail see [36]). To improve the efficiency of decision-making step in Bond Graph ARR based FDI approach, the measured residuals are associated to a Bayesian model that incorporates data on the reliability of the components. Associating reliability data to the diagnosis scheme will not only improve decision-making step but also some other tasks related to the intelligent supervision strategy:

- Programming preventive maintenance,
- Analysis of the failure cost by using utility nodes,
- Risk based reconfiguration of the faulty system by controlling its global or partial reliability (prognosis tasks).

B. Formulation of the bond graph based FDI system

1) Structural analysis

A system, *S*; may be described by a set of constraints, *F* (which represents the system model); a set of variables, *Z*; and a set of parameters *Θ*. Each variable may be known, or unknown: *S*=*S*(*F,Z, Θ*). Let *s* be a binary relation between *F* and *Z*; $s(f_i, z_j)$ =1 means that constraints $f_i \in Z$ (*s*=0 otherwise). The structure leads to a bipartite graph [7] whose binary incidence matrix represents the links between the known and the unknown variables, and the constraints. Reference [11] has shown that only over-constrained sub-systems can be monitorable and can provide ARRs. This subsystem contains more constraints *F* than unknown variables *X* and it is the only one to exhibit some redundancy which can be expressed as an ARR. Thus, an ARR is a relationship between a set of known variables of the form $f(K)=0$, where *K* is the set of known variables. In a bond graph based approach, the known variables are the sources (*Se and Sf*), the modulated sources (*MSe and MSf*), the measurements from sensors (*De and Df*), the model parameters (θ) and the controller outputs (*u*). An ARR is then written as

$$ARR: f(De, Df, Se, Sf, MSe, MSf, u, \theta) = 0. \quad (1)$$

The bond graph model of the monitored process is generated by using preferred derivative causality. The integral causality is recommended for engineering simulation in order to avoid the numerical problems arising out of differentiation. However, the derivative causality is more suitable in ARR expression to avoid influence of the initial conditions. As initial conditions are unknown in real processes, these relations are directly generated from BG model in derivative causality. The ARR generation algorithm is a recursive elimination technique [36]. The main idea is to eliminate all unknown variables of this equation using a covering causal path from each unknown variable to known one [38]. This leads to an oriented graph. This algorithm has been developed and implemented by the coauthor in dedicated software [30].

The ARR generation is the first step in a global diagnosis system design. The second step consists in alarm evaluation to avoid false alarm and non detection.

2) Classical approach for decision making

The procedure of decision-making is based on the evaluation of residuals. A residual, *r*, is the evaluation of an ARR when faults occur in the process, in the controllers or in the sensors or actuators:

$$r = Eval[f(K)]. \quad (2)$$

The residuals will be coherent with the model of the system. The coherence of each residual is tested. The procedure of test can vary from a residual to another. The elements, $c_i (i = 1..n)$ of the binary coherence vector $\mathscr{C} = [c_1, c_2, ..., c_n]$ are determined from one or more decision procedures, ϑ_i. These procedures generate the alarm conditions. Hence, $\mathscr{C} = [\vartheta_1(r_1), \vartheta_2(r_2), ..., \vartheta_n(r_n)]$. A simple test procedure consists of comparing the residual r_i with a threshold ε_i fixed a priori. Therefore, each component c_i of $\mathscr{C}$ is obtained using the following rule

$$c_i = \vartheta(r_i) = \begin{cases} 1, if \mid r_i \mid > \varepsilon_i; \\ 0, otherwise. \end{cases} \quad (3)$$

For modeling uncertainties, process and measurement noises, adaptive thresholds can be used [37]-[12].The final step in decision making is to compare the coherence vector to the Fault Signature Matrix (FSM) to find the

corresponding fault signature. The FSM noted as matrix $\mathscr{S}$ describes the structural sensitivity of each residual to various faults in physical devices, sensors, actuators and controllers. The elements of matrix $\mathscr{S}$ are determined from the following analysis:

$$S_{ji} = \begin{cases} 1, \text{if the j}^{\text{th}} \text{ residual is sensitive to fault in the i}^{\text{th}} \text{ component;} \\ 0, \text{otherwise.} \end{cases} \quad (4)$$

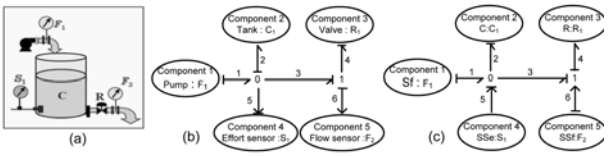


Fig. 2. Hydraulic system (a), BG model in integral causality (b), BG model in derivative causality after dualizing sensors (c)

3) Illustration example

Consider the simple hydraulic system (Fig. 2 (a)) with two sensors: an effort sensor (S_1) permitting to measure the pressure (linked to the mass stored in the tank C_1) and a flow sensor (F_2) measuring the flow in the valve R_1. The source F_1 represents the flow delivered by the pump. The model in integral causality is used for simulation (Fig. 2 (b)), the second model in derivative causality (Fig. 2 (c)) provides ARRs. This last is made by dualizing effort (or flow) sensors into a signal source *SSe=De* (or *SSf=Df*) modulated by the measured value.

In laminar regime, the BG model is linear. The equations deduced from junctions are:

$$0 - Junction : f_1 - f_2 - f_3 = 0 \Rightarrow r_1 = F_1 - \frac{1}{R_1}.S_1 - C_1.\frac{dS_1}{dt}, \quad (5)$$

$$1 - Junction : e_3 - e_4 = 0 \Rightarrow r_2 = R_1.F_2 - S_1. \quad (6)$$

The residuals r_1 and r_2 ((5) and (6)) are determined by eliminating the unknown variables using causal covering paths (from unknown to known variables). This leads to the well known oriented graphs. The FSM can be then deduced (Table I.). The row *M* indicates the detectability index (M_i =1 if it exists at least one residual sensible to the i^{th} component fault). The row *I* indicates the isolability index (I_i =1 if the boolean signature vector of i^{th} component fault is different from others). Note that F_1 and C_1 have identical failures signatures [1,0], as well as S_1 and R_1 [1,1]. Therefore, there is a problem of isolability of failures. To overcome this problem, we can insert additional sensors, what will need also the monitoring and isolation of the new sensors faults. As can be observed in Table I, a false alarm or a non detection can cause the same binary coherence vector for most of the components.

TABLE I
FAULT SIGNATURE MATRIX OF THE HYDRAULIC EXAMPLE

	F_1	S_1	C_1	R_1	F_2
r_1	1	1	1	1	0
r_2	0	1	0	1	1
M	1	1	1	1	1
I	0	0	0	0	1

C. Introducing reliability with Bayesian thinking

The equations of junctions deduced from a BG model are based on conservative laws. Suppose a leakage in the tank (Fig. 2(a)), this fault can be modeled by a flow source with a negative value connected to the *0*-junction (Fig. 2(b)). The first candidate ARR (5), which is generated from the conservative mass law at this junction, will be no more conserved if any fault may affect the component. Note that this is one of the advantages to use a BG model for monitoring compared to classical approaches (parity space, observer…) [1]. In Bayesian thinking, the leakage is a cause to not satisfying the conservative law and consequently to modify the corresponding residual value. The event of leakage itself is related to the reliability of the tank in the normal operating scheme. In the same example, sensor S_1 and valve R_1 have the same signature fault. If one knows that R_1 was repaired lastly or one have statistical data informing that this valve is unreliable compared to the effort sensor, it will be thought that the most probable cause of the fault signature [1,1], could be the valve failure.

In conclusion, the introduction of a Bayesian model associating the reliability of components and the measured residuals in the supervision module can improve the efficiency of the decision-making in diagnosis.

D. Hierarchical Bayesian modeling

1) Introduction

Hierarchical Bayesian modeling is another aspect of DAG describing the influence of the parameters to the global function of them. Let us suppose that one has *n* i.i.d. samples representing the data set $D=(x_1,...,x_n)$ from a density f_θ , with the unknown vector of parameters $\theta=(\theta_1,\theta_2,...,\theta_k)$, the associated likelihood function is

$$L(\theta \mid D) = \prod_{i=1}^{n} f_\theta(x_i). \tag{7}$$

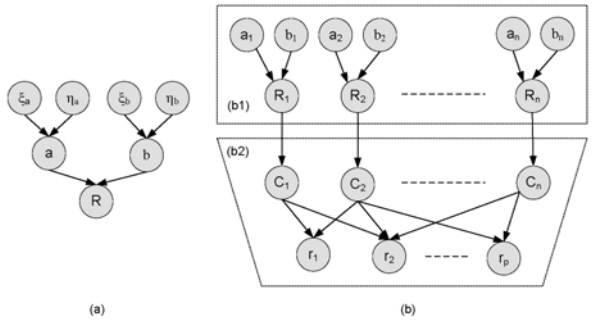


Fig. 3. (a) Hierarchical model of reliability with uncertain Weibull parameters, (b) The proposed Bayesian decision-making model (continuous part (b1), discrete part (b2))

This quantity represents the fundamental entity for the analysis of observation data about θ through D and the Bayesian inference will be based on this function. The posterior distribution of the parameter θ is given by the relation

$$p(\theta \mid D) = \frac{L(\theta \mid D)\pi(\theta)}{\int L(\theta \mid D)\pi(\theta)d\theta} \propto L(\theta \mid D)\pi(\theta), \tag{8}$$

$\pi(\theta)$ is the prior distribution of the parameter θ. The denominator is a normalizing constant. Generally, this integral does not have a close form, and therefore it is necessary to use approximate inference such as Markov Chain Monte Carlo (MCMC) algorithms [3]. The use of a two-level hierarchical model is the most current in the literature, but a model with higher number of levels is possible to construct.

2) Hierarchical Bayesian model of the Weibull distribution

The Weibull distribution of the failure, with its two parameters (shape and scale) permits the modeling of different regions of the bathtub curve in the lifecycle of a great number of components. The probability distribution function (PDF) of the Weibull distribution is defined by

$$f(t \mid a,b) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} \exp\left[-\left(\frac{t}{b}\right)^{a}\right], t \geq 0, \tag{9}$$

a is the parameter of shape, b is the parameter of scale and t is time. When these parameters are uncertain and we have a set of Data failures times or tests of the component, the hierarchical model of the Fig. 3(a) permits the determination of the component's reliability. Let $(t_1,...,t_l)$, the time failures of l identical components so that

TABLE II
FALSE ALARM AND NON DETECTION PROBABILITIES

	r_j	
C_i	D (*Detected*)	ND (*Not Detected*)
F (*Faulty*)	$1\text{-}P_{ndij}$	P_{ndij}
S (*Safe*)	P_{faij}	$1\text{-}P_{faij}$

$$t_i \sim Weibull(a,b),\ i = 0,...,l. \tag{10}$$

The likelihood function is the product of the Weibull distributions for every failure time t_i

$$L(a,b \mid t) = \prod_{i=1}^{l} \frac{a}{b}\left(\frac{t_i}{b}\right)^{a-1} \exp\left[-\left(\frac{t_i}{b}\right)^{a}\right]. \tag{11}$$

For the inference of this hierarchical model, it is necessary to sample from the prior distributions of (*a, b*) then the *Weibull*(*a, b*) distribution. Since (*a, b*) are positive, it is common to use Gamma prior distribution as conjugated of the Weibull one [16]. The Gamma Distribution is defined by:

$$f(x \mid \alpha,\gamma) = x^{\alpha-1}\gamma^{\alpha}\frac{e^{-\gamma x}}{\Gamma(\alpha)}, x > 0. \tag{12}$$

The two parameters are sampled separately:

$$a \sim Gamma(\xi_a, \eta_a),$$
$$b \sim Gamma(\xi_b, \eta_b),$$

with ξ_a and η_a the shape hyperparameters and ξ_b and η_b the scale hyperparameters. The inference on the global hierarchical model can be performed by using adaptive rejection sampling [15] and Gibbs sampling [9].

E. The Decision-making method

1) The decision module

Suppose our system composed of n components C = $\{C_i; 1 \leq i \leq n\}$ with Weibull distributions of failures. The Bayesian model of decision contains random variables associated to the residuals r = $\{r_j; 1 \leq j \leq p\}$, to the components as well as the Bayesian reliability model of these components. The proposed Bayesian decision-making model is displayed in Fig. 3(b). An arc that joins node C_i to node r_j (we really join associated random variables) indicates that r_j is sensitive to the failure of the component C_i. For a residual r_j there are two states {D(*Detected*), ND(*Not Detected*)}, we

have also two states {F(*Faulty*), S(*safe*)}for a component C_i. Every component C_i is associated with its reliability R_i.

As can be observed, this structure is hybrid; there are discrete and continuous nodes. A hybrid BN represents a probability distribution over a set of random variables where some are discrete and others are continuous. In literature, the most widely used subclass of hybrid BNs is the conditional linear Gaussian (CLG) model [21]. This model is discrete parents and continuous leaves model. Many kinds of inference algorithms are stated in literature; exact inference [22], approximate inference [19], dynamic discretisation [20], mixtures of truncated exponential [26]. In [23], a new inference algorithm has been provided for the filtering in HBN in order to supervise and diagnose hybrid dynamic systems.

This network can be treated as being an association of a Discrete BN and a Continuous BN (CBN). The CBN permits to prepare the prior information on the failure of the component. So when a residual is detected at instant t, the component C_i has the prior probabilities; $P(C_i=Faulty)=F_i(t)=1-R_i(t)$ (The function F_i designates the cumulative distribution function (CDF)).

The discrete part possesses a structure that depends on the failures signatures; when a residual r_j is not sensitive to the failure of a component C_i no arc is pulled from node C_i toward node r_j. The inference of the two parts can be performed separately. After the detection of residuals, the posterior probabilities of the failures $p(C_i|r_1,...,r_p)$ can be determined by inference on the discrete part of the network.

2) Inference on the continuous part

At this stage, we have to estimate the reliability of each component using the posterior density of parameters. The expected value for a specified operating time T is determined by the formula

$$E[R(T \mid Data)] = \int R(T)\, p(\theta \mid Data) d\theta. \tag{13}$$

With MCMC simulations, one can easily assess characteristics such as mean, median and quantiles. The credible limit (*CL*) is defined for the two sided reliability interval $[R_l,R_u]$. Generally, there are two choices for the value of *CL*. For example, for the ball-bearing industry, the tradition is to specify (*CL=0.9*) [34]. Another choice is possible which is the value that corresponds to the median (*CL=0.5*). This value is more stable than the mean one. Therefore, the prior probabilities of failures can be written as follows

$$P(C_i=Faulty)=1-R_{i(0.5)}(\mathrm{T}), \tag{14}$$
$$P(C_i=Safe)=R_{i(0.5)}(\mathrm{T}). \tag{15}$$

3) Prior probabilities of false alarm and non detection

Before starting the inference on the discrete part, it is clear we need to determine the prior probabilities of false alarm and non detection. In the case of a residual r_j sensitive to failure of C_i and the probabilities of false alarm P_{fa} and non detection of the residual P_{nd}, the conditional probability table (*CPT*) is defined according to Table II. In the absence of prior knowledge on these probabilities, the method using statistics and tests [40] can be applied.

The conditional probabilities $p(r_j|C_1,...,C_n)$ are determined according to the Bayes rule :

$$p(r_j \mid C_1, C_2, ..., C_n) = \frac{p(C_1, C_2, ..., C_n \mid r_j)\, p(r_j)}{p(C_1, C_2, ..., C_n)}. \tag{16}$$

We suppose the events joined to the different failures are independent. When the marginal distributions $p(r_j)$ of the residuals are unknown, one can take the prior conditional probabilities $p(r_j|C_1,...,C_n)$ as being the product of the conditional priors

$$\begin{aligned} &p(r_j = ND \mid C_1, C_2, ..., C_n) \\ &= p(C_1 \mid r_j = ND)\, p(C_2 \mid r_j = ND) ... p(C_n \mid r_j = ND), \end{aligned} \tag{17}$$

and

$$p(r_j = D \mid C_1, C_2, ..., C_n) = 1 - p(r_j = ND \mid C_1, C_2, ..., C_n). \tag{18}$$

For example, for a residual r_j sensitive to the failures of two components C_1 and C_2, we have

$$p(r_j = ND \mid C_1 = F, C_2 = S) = P_{nd1j}(1 - P_{fa2j}),$$
$$p(r_j = D \mid C_1 = F, C_2 = S) = 1 - P_{nd1j}(1 - P_{fa2j}).$$

4) Inference on the discrete part with observations

For inference on discrete BNs, one can use the exact method or the approximate (or stochastic) one. Indeed, the choice of the method depends on the size of the network; for small networks one can perform exact inference. The most important methods are variable elimination and junction tree [33]. On the other hand, if the size of the network is important and the exact inference is not tractable, one can use Markov Chain Monte Carlo (MCMC) algorithms. In the BN formalism, the joint probability of the network is the product of the conditional probabilities

$$p(C_1, ..., C_n, r_1, ..., r_p) = \prod_{j=1}^{p} p(r_j \mid Par(r_j)) \prod_{i=1}^{n} p(C_i). \tag{19}$$

After the observation of the residuals r_j, the inference is achieved and these observations are considered as evidence in the BN theory. The algorithm of inference permits to calculate the probability of the failure of component C_i

conditionally to these observations $p(C_i|r_1,...,r_p)$.

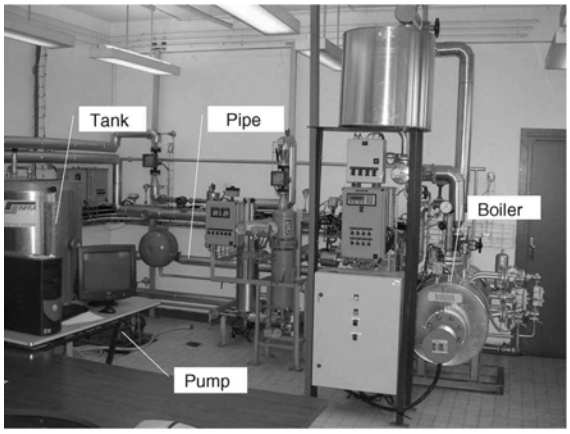


Fig. 4. Overview of the steam generator installation

IV. APPLICATION

A. System to be monitored

The application system is a steam generator pilot process installation (Fig. 4). This plant represents a reduced scale of a power station. The whole installation is constituted of four principle subsystems: a receiver with the feed water supply system, a boiler heated by a 55 kW thermal resistor, a steam flow system, and a complex condenser coupled with a heat exchanger. As can be seen in Fig. 5(b), the heated boiler is fed by water via a tank, a redundant pump and a pipe. To simplify the size of the graphical Bayesian decision model, our study is focused on only these latest parts (Fig. 5(b)). The steam generator is a thermo-fluid process involving both convection and conduction heat transfer. For sufficiently low velocities, the kinetic energy is negligible and the convected energy $\dot{H}$ is calculated from the mass flow $\dot{m}$ and the specific thermal capacity c_p, as follows:

$$\dot{H} = \dot{m}h = \dot{m}c_p T, \tag{20}$$

h is the specific enthalpy and T is the temperature. Thus the pseudo-bond graph vector power variables (e and f) for thermo-fluid systems are chosen as

$$e = \begin{bmatrix} e_h & e_t \end{bmatrix} = \begin{bmatrix} P & T \end{bmatrix}, f = \begin{bmatrix} f_h & f_t \end{bmatrix} = \begin{bmatrix} \dot{m} & \dot{H} \end{bmatrix},$$

where P is the pressure. The word BG of the monitored plant is represented in Fig. 5(a). There are five principle components (tank, pump, pipe, boiler, heater) associated to some sensors to perform control and diagnosis of the application.

The remainder of the paragraph is organized as follows; first we introduce all the necessary physical knowledge about the plant, also the used hypothesis. Thereafter, it is required to present the failure rates of the components. Finally the developed theory is applied to the process to be monitored.

B. Bond graph model of the process

1) Introduction

Before starting to explain the functionality of each component let us see the BG model in Fig. 6. One of the most important properties of the BG language is that every element of the representation graph is associated with a physical component of the process. Such a property is interesting when we aim to associate to the BG model the reliability of each component. Our innovative interest is to combine BG modeling with a Bayesian reliability model to improve the decision making task in FDI. The BG model of the process (Fig. 6) is given in derivative causality because the initial conditions are unknown and the model will be used for diagnosis. We must note here that all effort (or flow) sensors are dualized into a signal source *SSe=De* (or *SSf=Df*) and when it is not possible there is a physical redundant component.

2) BG model of the tank

The tank in the steam generator is considered as a coupled

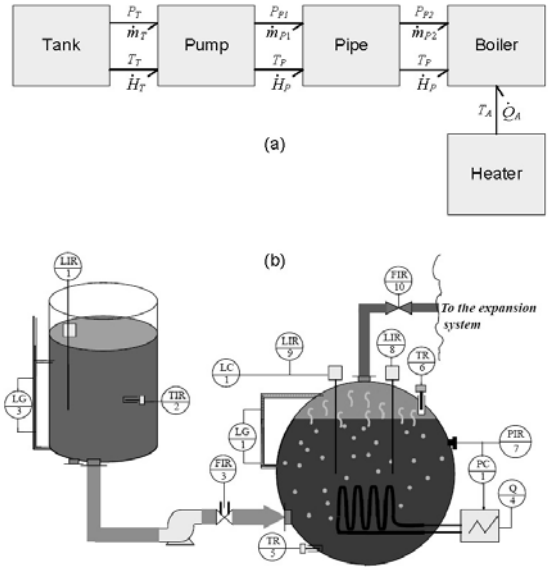


Fig. 5. Application system (b) and its word BG (a)

thermo-fluid storage device. The coupled and stored thermo-fluid energy in the tank is modeled by the two port C-element (C_h: hydraulic, C_t: thermal) and the two derived state variables correspond to the stored mass and total enthalpy.

When thermodynamic regime in the tank is saturated, the thermal element *C:* C_t is modulated by hydraulic effort power variable, where the internal energy depends on the stored mass this is because the two state variables (thermal and hydraulic) are coupled. The tank is assumed to be initially full and the input volumetric flow *Sf :* $\dot{m}_{in}$ is assumed equal to zero. The following equation is deduced from junction 0_{h1}

$$\dot{m}_{out} = -C_h.\frac{dP_1}{dt}, \tag{21}$$

where $\dot{m}_{out} = \dot{m}_3 = \dot{m}_T$ is the outlet volumetric flow from the tank, expressed in (m^3/s), *C:*C_h represents the hydraulic capacity of the tank and *De:* P_1=P_T is the measured fluid pressure inside the tank. By considering that the studied tank is cylindrical, the hydraulic capacity C_h can be expressed as follows:

$$C_h = A_T.(\rho_T.g)^{-1}, \tag{22}$$

where A_T describes the section of the tank, ρ_T is the fluid density and *g* is the gravity acceleration.

The enthalpy flow at the output of the *C:*C_t element is given by the following equation:

$$\dot{H}_5 = -T_2.c_p.\dot{m}_{out}, \tag{23}$$

where c_p is the fluid specific heat capacity at constant pressure and T_2 is the sensor measurement of the fluid temperature inside the tank.

3) BG model of the pump

The pump is a redundant component. The mass flow rate from tank to the boiler is a function of the pressure head across the pump. From bond graph point of view, the pump is a non-linear resistance *R:*R_p modulated by the expression (24), which describes the relation between the pressure $\Delta P = P_{14} - P_3$ and the volumetric flow $\dot{m}_{14}$ generated by the pump.

$$\dot{m}_{14} = (k_1.\Delta P + k_2).mO_2, \tag{24}$$

where k_1 and k_2 are the characteristics of the pump and mO_2 is a binary signal from the output of the controller (boiler level controller).

4) BG model of the pipe

The parameter *R:*R_z depends on the tubing characteristics and the supply valve; it is calculated with the relation:

$$R_Z = \frac{8\rho_l L_p}{\pi r_p^4}, \tag{25}$$

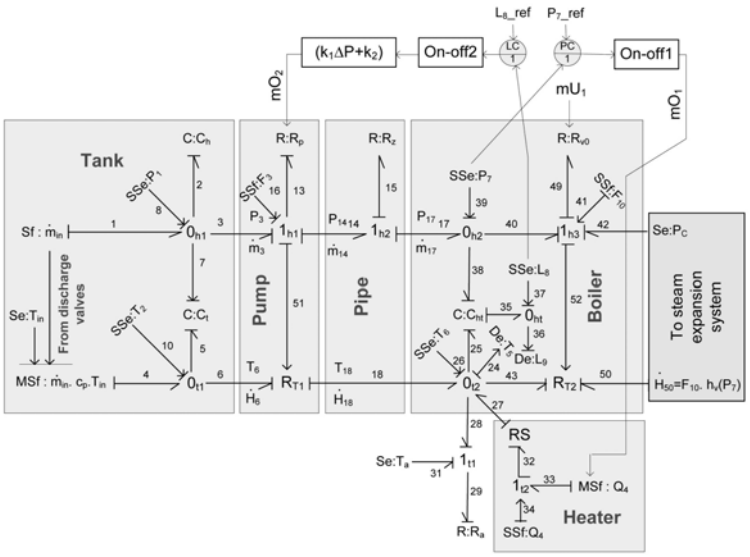


Fig. 6. BG model of the system to be monitored in derivative causality

TABLE III
FAILURE SIGNATURE MATRIX OF THE APPLICATION SYSTEM

	L_1	T_2	F_3	L_8	P_7	Q_4	T_6	F_{10}	$On1$	$On2$	Tnk	Pmp	Ppe	Blr	Htr	V_0
r_1	1		1								1					
r_2	1		1		1							1	1			
r_3			1	1	1			1						1		
r_4	1	1	1								1					
r_5		1	1	1	1	1	1	1						1		
r_6						1									1	
r_7					1				1							
r_8				1						1						
r_9					1		1									
r_{10}					1			1								1
M	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1

with L_p being the pipe length and r_p its radius.

The volumetric flow $\dot{m}_{17}$ is calculated using *Bernoulli* law as follows:

$$\dot{m}_{17} = \frac{1}{R_z}\sqrt{|P_{14} - P_{17}|}.sign(P_{14} - P_{17}).mO_2. \tag{26}$$

5) BG model of the boiler

The storage of hydraulic and thermal energies is modeled by the two-port graph *C*-element *C:* C_{ht}. During boiling, it is assumed that the water and the steam are saturated and are in thermal equilibrium. The studied boiler system is instrumented with two redundant sensors of temperature (*De:* T_5 and *De:* T_6), two redundant volume sensors (*De:* L_8 and *De:* L_9), a pressure sensor (*De:* P_7), and a volumetric flow sensor at the output of the boiler (*Df:* F_{10}).

The volumetric flow stored by the boiler depends on the variation of the steam-liquid mass, and is expressed as follows:

$$\begin{cases} \dot{m}_{C_{ht}} = \dfrac{d}{dt}(\rho_l.\mathrm{V}_l + \rho_v.\mathrm{V}_v), \\ \dot{H}_{C_{ht}} = \dfrac{d}{dt}(\rho_l.\mathrm{V}_l.\mathrm{h}_l + \rho_v.\mathrm{V}_v.\mathrm{h}_v - \mathrm{P}_B.V_B), \end{cases} \tag{27}$$

where ρ_l, h_l, V_l and ρ_v, h_v, V_v are respectively, the density, the specific enthalpy and the volume of the water and the steam inside the boiler. P_B is the measured boiler pressure given by the detector *De:* P_7 and V_B is the volume of the boiler. All the variables ρ_l, h_l, ρ_v and h_v are functions of the pressure *De:* P_7 and can be identified or measured as follows:

- Water volume V_l is given by the volume detector *De:* L_8.
- Steam volume V_v=V_B-V_l is equal to the difference between the total volume of the accumulator V_B and the water volume V_l.
- ρ_l, h_l, ρ_v and h_v are calculated using a polynomial interpolation algorithm.

The outlet enthalpy flow from the boiler to the expansion system can be calculated as follows:

$$\dot{H}_{43} = T_{25}.c_v.\dot{m}_{40}, \tag{28}$$

where c_v is the specific heat capacity at constant volume, T_{25} and $\dot{m}_{40}$ are taken from the temperature detector *De:* T_6 and the volumetric flow sensor *Df:* F_{10}.

Consequently, the outlet enthalpy flow $\dot{H}_{50} = \dot{H}_{43}$ depends on the measurement values of F_{10} and P_7 via the thermodynamic function h_v:

$$\dot{H}_{50} = F_{10}.h_v(P_7). \tag{29}$$

6) BG model of the heater

The heating process is a thermal resistor modeled by *R:RS* element. The power provided from this resistor is measured with a flow sensor *Df:*Q_4. The heating energy is controlled by *On-Off1* according to P_7*_ref*.

The dissipation of the heat flow $\dot{Q}$ via the boiler wall (30), which we neglected the correspondent *C*-element, can be determined using the thermal conductivity λ, the thickness e_B, the temperature difference (T_B-T_a) (T_a is the ambient temperature) between the wall sides and the section A_B of the boiler wall:

$$\dot{Q} = \lambda \frac{A_B}{e_B}(T_B - T_a). \tag{30}$$

The heat transfer from boiler to the environment is described by $R : Ra = \lambda \dfrac{A_B}{e_B}$.

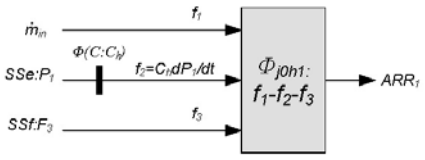


Fig. 7. Oriented graph for deduction of ARR_1

C. ARRs generation

The first candidate ARR is generated from the junction 0_{h1}:

$$\Phi_{j0h1} : f_1 - f_2 - f_3 = 0, \quad (31)$$

f_1, f_2 and f_3 are unknown variables; they will be eliminated using covering causal paths from unknown to known ones.

$f_1 = Sf : \dot{m}_{in} = 0;$

f_2 will be eliminated from the following path :

$f_2 \to \Phi(C : C_h) \to e_2 \to SSe : P_1,$

where $\Phi(C : C_h)$ is the constitution equation of C-element,

$f_2 = C_h \frac{dP_1}{dt}$.

f_3 is calculated from the causal path: $f_3 \to f_{10} \to SSf : F_3$, thus, f_3=F_3. Finally the first ARR is deduced by substituting the unknown variables in candidate ARR, this yields to:

$$ARR_1 : -C_h \frac{dP_1}{dt} - F_3 + \dot{m}_{in} = 0. \quad (32)$$

The cited covering causal paths can be summarized in an oriented graph (Fig. 7).

ARR_2 comes from the junction 1_{h2} connected to the flow sensor F_3:

$$\Phi_{j1h2} : e_{14} - e_{15} - e_{17} = 0. \quad (33)$$

The expression of the outlet volumetric flow f_{14}:

$$f_{14} = f_3 = -\frac{A_T}{\rho_T . g}\left(\frac{de_3}{dt}\right), \quad (34)$$

This is also the flow through the pump; it has the following transfer expression:

$$f_{14} = f_{13} = (k_1(e_{14} - e_3) + k_2).mO_2, \quad (35)$$

Then e_{14} is determined using the equality of (34) and (35) with the condition that mO_2=1 (the dynamic of the system is hybrid):

$$e_{14} = -\frac{A_T}{\rho_T . g}\left(\frac{de_3}{dt}\right) - \frac{k_2}{k_1} + e_3. \quad (36)$$

Using the same methodology for ARR_1 and knowing that:

TABLE IV
FAILURE RATES OF THE APPLICATION COMPONENTS

Component	Symbol	Failure rate (*Fail./10⁶ op. Hrs*)	MTTF (*Hrs*)
Tank	*Tnk*	0.01	10^8
Pump	*Pmp*	17.56	56947
Pipe	*Ppe*	0.56	1785714
Boiler	*Blr*	0.05	2x 10^7
Heater	*Htr*	0.02	5x10^7
Valve	V_0	1.25	8x10^5
Sensors	Q_4	0.3	3,33x 10^6
	F_3	133	7518
	F_{10}	186	5376
	L_1	77	12987
	L_8	108	9260
	P_7	39	25641
	T_2	6.6	1.5x10^5
	T_6	9.3	1.07x10^5
Controllers	*On1, On2*	10	10^5

$e_3 = P_1; e_{15} = R_z F_3; e_{17} = P_7,$

ARR_2 can be written as:

$$ARR_2 : -R_z F_3 - \frac{A_T}{k_1 \rho_T g}\left(\frac{dP_1}{dt}\right) - \frac{k_2}{k_1} + P_1 - P_7 = 0. \quad (37)$$

Writing the equation around 0_{h2} leads to ARR_3.

$$\Phi_{j0h2} : f_{17} - f_{38} - f_{40} = 0, \quad (38)$$

$f_{17} = F_3; f_{40} = F_{10}; f_{38} = \dot{m}_{C_{ht}} = \frac{d}{dt}(\rho_l V_l + \rho_v V_v)$ (see (27)).

$$ARR_3 : F_3 - \frac{d}{dt}(\rho_l V_l + \rho_v V_v) - F_{10} = 0. \quad (39)$$

ARR_4 can be expressed from equation of junction 0_{t1}:

$$\Phi_{j0t1} : f_4 - f_5 - f_6 = 0. \quad (40)$$

$f_4 = \dot{m}_{in} c_p T_{in} = 0; f_6 = T_2 c_p F_3; f_5 = \frac{d}{dt}(C_t e_5)$, with

$C_t = m.c_p = C_h.e_2.c_p$ (m is the mass of liquid) ;

$e_2 = \rho_T g L_1; e_5 = T_2,$ thus $f_5 = A_T \rho_T \left[T_2 \frac{dL_1}{dt} + L_1 \frac{dT_2}{dt}\right].$

By substituting the unknown variables in Φ_{j0t1}, this yields to:

$$ARR_4 : -T_2 c_p F_3 - A_T \rho_T \left[T_2 \frac{dL_1}{dt} + L_1 \frac{dT_2}{dt}\right] = 0. \quad (41)$$

ARR_5 can be expressed from equation of junction 0_{t2}:

$$\Phi_{j0t2} : f_{18} + f_{27} - f_{25} - f_{28} - f_{43} = 0. \quad (42)$$

The expressions of flows are:

$f_{18} = F_3 c_p T_2; f_{27} = RS.Q_4; f_{28} = Ra(T_6 - T_a);$

$f_{25} = \dot{H}_{C_{ht}} = \frac{d}{dt}(\rho_l.V_l.h_l + \rho_v.V_v.h_v - P_7.V_B)$ (see (27));

$f_{43} = \dot{H}_{43} = F_{10} c_v T_6.$

ARR_5 can be deduced:

$$ARR_5 : F_3 c_p T_2 + RS.Q_4 - \frac{d}{dt}(\rho_l.V_l.h_l + \rho_v.V_v.h_v - P_7.V_B) - Ra(T_6 - Ta) - F_{10} c_v T_6 = 0. \quad (43)$$

The equation of heating control yields to

$$ARR_6 : Q_4 - W_p.mO_1 = 0, \quad (44)$$

W_p is the power of the heater.

ARR_7 and ARR_8 are deduced from the equation of *OnOff* controllers:

$$ARR_7 : mO_1 - OnOff1(P_{7\_ref}, P_7) = 0. \quad (45)$$

$$ARR_8 : mO_2 - OnOff2(L_{8\_ref}, L_8) = 0. \quad (46)$$

Using the thermodynamic function *Ps2Ts*(.) [29] to calculate saturated steam temperature from known pressure yields to ARR_9:

$$ARR_9 : T_6 - Ps2Ts(P_7) = 0. \quad (47)$$

ARR_{10} is deduced by writing the equation of junction 1_{h3}:

$$\Phi_{j1h3} : e_{40} - e_{49} + e_{42} = 0. \quad (48)$$

$e_{40} = P_7; e_{42} = P_c; e_{49} = \sqrt{|P_7 - P_c|}.sign(P_7 - P_c).mU_1,$

P_c is the pressure at the exit of the exhaust valve and mU_1 is a manual operating control. The constraint related to the component *R:* R_{v0} (valve V_0) permits to deduce ARR_{10}:

$$ARR_{10} : F_{10} - \Phi^{-1}_{Rv0}(P_c - P_7) = 0. \quad (49)$$

Finally, ARR_{10} can be written as:

$$ARR_{10} : F_{10} - V_{0\_cd}\sqrt{|P_7 - P_c|}.sign(P_7 - P_c).mU_1 = 0, \quad (50)$$

with V_{0_cd} the discharge coefficient of valve V_0.

The theoretical FSM is presented at Table III. The symbols used are described in Table IV. As the application is well instrumented, all faults are isolable only faults arising from the pump and the pipe. A fault in both of these components has a direct effect on the residual r_2 correspondent to ARR_2.

D. Reliability data for the components

After establishing the FSM and observing the problem of isolation of the pump and the pipe failures, now we aim to apply the incorporation of reliability data to improve the decision task in diagnosis. In the absence of historical reliability data of the plant, we used a reliability Handbook to estimate the failure rates.

1) Reliability failure rates of the pump and the pipe

As the pump used in the steam generator is centrifugal, its failure rate can be estimated using (51) [28]:

$$\lambda_p = \lambda_{SE} + \lambda_{SH} + \lambda_{BE} + \lambda_{CA} + (\lambda_{FD}.C_{TLF}.C_{PS}.C_C) \quad (51)$$

where

λ_{SE} : Total failure rate for all pump seals (Failures/million operating hours),

λ_{SH} : Total failure rate for the pump shaft,

λ_{BE} : Total failure rate for all pump bearings,

λ_{CA} : Total failure rate for all pump casing,

λ_{FD} : Total failure rate for all pump fluid driver,

C_{TLF}: Thrust load multiplying factor,

C_{PS} : Operating speed multiplying factor,

C_C : Contaminant multiplying factor.

Using the basic value of the failure rate (when missing informations), we estimated these failure rates and multiplying factors to:

λ_{SE}=2.4; λ_{SH}=5; λ_{BE}=10; λ_{CA}=0.001; λ_{FD}=0.2; C_{TLF}=1; C_{PS}=0.74; C_C=1.1.

The global failure rate of the pump is λ_p=17.56 Fail./ 10^6 op. hours.

The pipe is a part of fluid conductors in the plant. It is important to note that most failures of fluid conductor systems occur at or within the interconnection points such as fittings and flanges. Since the failure rate of a piping assembly usually depends primarily on the connection joints, the basic failure rate of a piping assembly can be estimated at 0.47 Fail./10^6 op hours per connection and the failure rate of the pipe assembly can be estimated with the following equation [28]:

$$\lambda_{Ppe} = \lambda_{P,B}.C_E \quad (52)$$

$\lambda_{P,B}$: Base failure rate of pipe assembly estimated to 0.47 (Fail./10^6 op. hours), C_E : Environmental factor. Taking C_E =1.2 yields to λ_{Ppe} =0.56 Fail./10^6 op. hours.

2) Reliability of the rest of components

The failure rates of the application components are given in Table IV (Estimated according to the same handbook [28]). We assume that all failure distributions are exponential. Note that this can be considered as prior information about reliability, and this data can be refined to Weibull or any other PDF models of reliability when it is

learned with new experimental and historical failures data (see equation (8)). So we do not discuss, in the analyses presented here, the uncertainty of the failure rates.

E. Application of the proposed methodology for diagnosis

To build the Bayesian decision model, we supposed the parameters associated to false alarm and non detection P_{faij} (i is the index of the component and j the index of the residual) and P_{ndij} identical for all components. These parameters are deduced from tests on the plant (P_{fa}=0.04, P_{nd}=0.02).

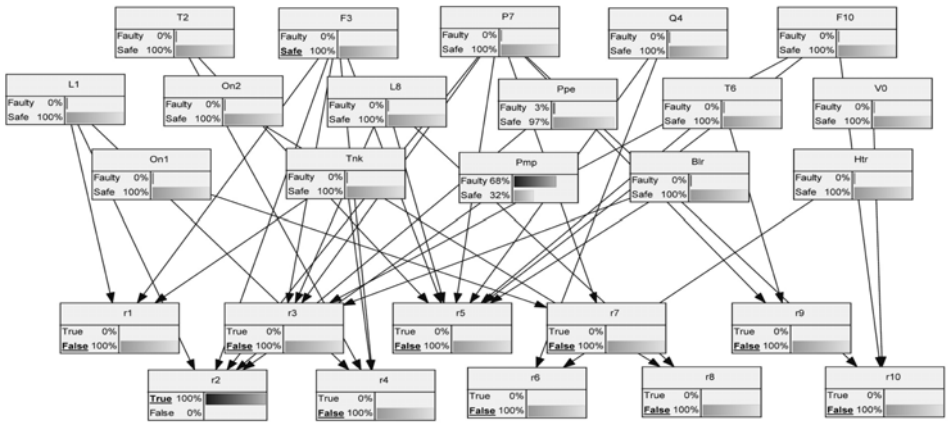
For the inference of the discrete part of the decision module, we used the free software GeNie 2.0 (http://genie.sis.pitt.edu) after introducing the prior probabilities which are calculated using (17) and (18). Since we assume certain failure rates, the prior probabilities of failures deduced from the continuous part of the model are calculated by the CDF:

$$F_i(T) = 1 - R_i(T) = 1 - \exp(-\lambda_i T), \tag{53}$$

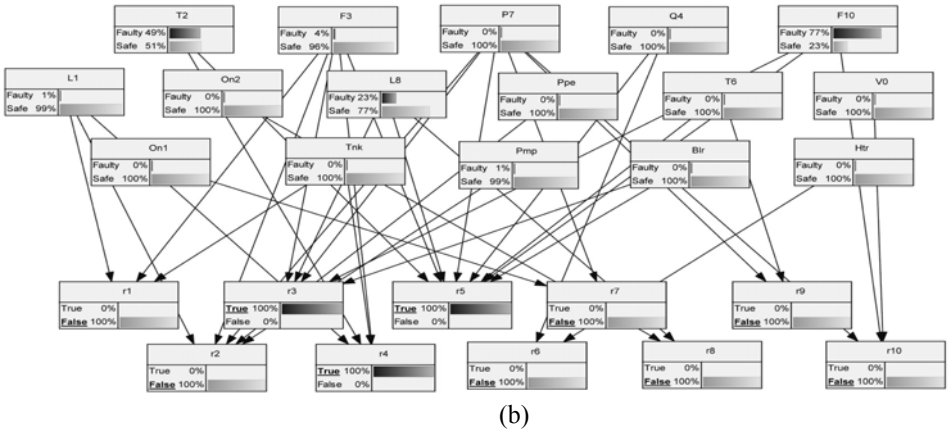
with λ_i the failure rate of the component (Failure/10^6 operating hours). To test the decision model, we will simulate three scenarios.

1) Scenario 1

After an operating time of 20000 hours (*Hrs*), we detected the coherence vector $\mathscr{C} = [\vartheta(r_1), \vartheta(r_2), ..., \vartheta(r_{10})]$ = [0,1,0…0] that corresponds to the failure of both the pump (*Pmp*) and the pipe (*Ppe*) (see Table III). Fig. 8(a)



(a)



(b)

Fig. 8. Results of analysis : scenario 1 (a), scenario 2 (b)

summarizes the result of analysis; the cause of the failure is 68% the pump whereas it is 3% the pipe. By the classic method of diagnosis, which gives the same chance for both of failures as can be seen in FSM (Table III), the decision module cannot make a final decision. Given this result, the supervision module can deduce that the pump is the most probable faulty component in this situation.

2) Scenario 2

In this case, we will suppose to have unknown residuals, which is a frequent situation in FDI by ARR approach caused by noise and uncertainty of the parameters of the bond graph model. Let us assume that after 20000 *Hrs* we have detected three active residuals (r_3,r_4,r_5). As can be observed in Table III, the failure signature [0,0,1,1,1,0,0,0,0,0] is not matched to any component, but there are some close signatures associated to the components :T_2, L_8, F_{10} and *Blr*. Also by the classic method it is not possible to decide the origin of failure. The inference shows (Fig. 8(b)) the posterior probabilities of failures: 77% for F_{10}, 49% for T_2, 23% for L_8, 4% for F_3 and 0% for *Blr*.

Given the Mean Time To Failure (MTTF) (Table IV) of temperature sensor T_2 and flow sensor F_{10} (respectively $1.5x10^5$ and 5376), one can deduce that the component F_{10} is probably defective for this analysis.

3) Scenario 3

For this scenario, we will suppose that before the process arrives to an operating time of 20000 Hrs, even that no residuals are detected we checked the prior probabilities of failures. Consider, for instance, the case in which the analysis is made after an operating time of 10000 Hrs with no residuals detected. Fig. 9 resumes prior and posterior probabilities of failures for each component. Clearly, most of sensors begin to be in a critical situation. Even though L_8 is a redundant component, F_3 and F_{10} need certainly some preventive maintenance actions.

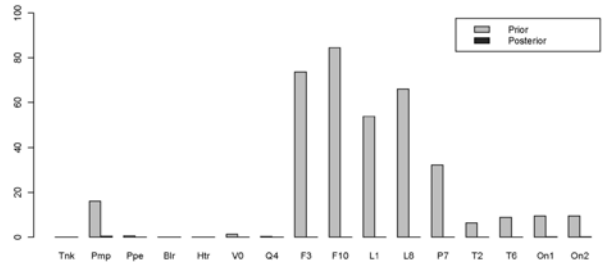


Fig. 9. Prior and posterior probabilities of failures with no residuals detected and after an operating time of 10000

As stated before, the improvement of decision making aims to not only take a decision in the case of non isolable failures or unknown signatures but also to be a part of a prognosis module to prevent undesired outcomes. Here we raise the issue about the need of such module in the intelligent supervision strategy which can be classified in risk based supervision.

V. Conclusions

In this paper it is shown how the bond graph as an integrated tool design and the Bayesian networks can be used as an intelligent framework to decision-making in model based diagnosis. We presented an issue to the problems revealed by the classical binary decision-making step in ARR model based FDI. We focused on the BG method because of its functional aspect by associating a physical component to each graphical element. The proposed methodology provides continuous decision variables in the form of posterior probabilities of failures so that the model permits to represent the degradation of the components. These variables can be used for further intelligent supervision tasks; programming preventive maintenance, analysis of the failure cost by using utility nodes, risk based reconfiguration of the faulty system by controlling its global or partial reliability (prognosis tasks).
The proposed method can be applied to large systems with components having all types of failures distributions. The response time of the decision model depends on the efficiency of the inference algorithms. The precision of results is influenced by the reliability Data. Although we used certain exponential parameters in the given application example, we have highlighted how to deal with uncertain parameters and the use of Weibull distributions. The results of application on a steam generator pilot process are satisfactory.

Acknowledgment

This research work is supported by laboratories LAGIS (Laboratoire d'Automatique, Génie Informatique et Signal, Lille, France) and LACS (Laboratoire d'Analyse et Commande des Systèmes, Tunis, Tunisia).

References

[1] A. Aitouche, and B. Ould-bouamama," Sensor location with respect to fault tolerance Properties," *International Journal of Automation and Control* Inderscience Publishers , Vol.4, Issue 3. pp. 298 - 316, 2010.

[2] M. Anderson, R. Anderson & K. Wheeler," Filtering in hybrid dynamic Bayesian networks," *Interntional Conference on Acoustics, Speech and Signal Processing*, vol 5 , pp. 773 -776, 2003.

[3] C. Andrieu, N. de Freitas, A. Doucet & M. I. Jordan, "An introduction to MCMC for Machine Learning," Kluwer Academic Publishers vol 50, pp. 5-43, 2001.

[4] J. Armengol, J. Vehi, M. A. Sainz, and P. Herrero," Fault detection in a pilot plant using interval models and multiple sliding windows," In

N. Eva Wu, editor, Safeprocess 2003, pages 729-734, Washington DC, USA, IFAC, 2003.

[5] M. Basseville, and I. V. Nikiforov," Detection of Abrupt Changes: Theory and Application," Prentice Hall, ISBN 0-13-126780-9, 1993.

[6] G. Biswas, Simon, G., Mahadevan, N., Narasimhan, S., Raminez, J. & G. Karsai,"A robust method for hybrid diagnosis of complex systems," In 5th IFAC symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS, 1125-1131, 2003.

[7] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki,"Diagnosis and Fault Tolerant Control," Springer-Verlag, 2003.

[8] A. Bregon, B. Belarmino, G. Biswas, X. Koutsoukos,"Generating possible conflicts from bond graphs using temporal causal graphs," 23rd European Conference on Modelling and Simulation ECMS'09, June 9-12, Madrid, Spain 675-682, 2009.

[9] G. Casella, and E. I. George,"Explaining the Gibbs Sampler," American Statistician, Vol. 46, No 3.pp. 167-174, August, 1992.

[10] G. Dauphin-Tanguy, A. Rahmani, and C. Sueur," Formal determination of controllability/observability matrices for multivariables systems modelled by bond graph," In International IMACS/SILE Symposium on Robotics, Mechatronics and Manufacturing Systems'92, pages 573—578, Kobe, Japon, 1992.

[11] P. Declerck, "Analyse structurelle et fonctionnelle des grands systèmes : Application à une centrale PWR 900 MW, " Ph. D. thesis, Université des Sciences et Technologies de Lille, France. (1991)

[12] M. A. Djeziri, B. Ould Bouamama and R. Merzouki, "Modelling and robust FDI of steam generator using uncertain bond graph model," *Journal of Process Control*, Vol. 19, Issue 1, January, pp. 149-162. 2009.

[13] P.J. Gawthrop,"Bicausal bond graphs," In ICBGM'95, pp. 83-88, Las Vegas, USA, 1995.

[14] J. Gertler, "Fault detection and isolation using parity relations," Control Engineering Practice, 5 No. 5:653-661. 1997

[15] W. R. Gilks, P. Wild,"Adaptive Rejection Sampling for Gibbs Sampling," Appl. Statist. 41, No. 2, pp.337-348, 1992.

[16] M.S. Hamada, A. G. Wilson, C. S. Reese, H. F. Martz," Bayesian Reliability," Springer series in statistics, ISBN 978-0-387-77948-5, 2008.

[17] R. Isermann,"Process fault detection based on modelling and estimation methods : A survey," Automatica, 20:387-404, 1994.

[18] D. Koller, U. Lerner," Sampling in Factored Dynamic Systems," In Sequential Monte Carlo Methods in practice: pp. 445-464, Springer-Verlag, New York, 2000.

[19] D. Koller, U. Lerner, D. Angelov,"A General Algorithm for Approximate Inference and its Application to Hybrid Bayes Nets," In Proceed. Of the Fifteenth Annual Conf. on Uncertainty in Artificial Intelligence UAI-99 pp. 324-333, Stockholm, Sweden, August. 1999

[20] A. V. Kozolov, and D. Koller," Nonuniform dynamic discretisation in hybrid networks," In Proceed. Of the Thirteenth Annual Conf. on Uncertainty in Artificial Intelligence UAI-97 pp. 314-325, August 1-3, 1997.

[21] S. Lauritzen, and F. Jensen," Stable local computation with conditional Gaussian distributions," Technical Report R-99-2014, Dept. Math. Sciences, Aalborg Univ. , 1999.

[22] U. Lerner, E. Segal, D. Koller," Exact inference in Networks with Discrete Children of Continuous Parents," Uncertainty in Artificial Intelligence, vol. 17, Morgan Kaufmann, San Francisco, CA, pp. 319-328, 2001.

[23] U. Lerner, R. Parr, D. Koller & G. Biswas," Bayesian fault detection and diagnosis in dynamic systems," Proc. Of the 17th National Conference on Artificial Intelligence (AAAI) pp. 531-537, 2000.

[24] N. Mehranbod, M. Soroush, C. Panjapornpon," A method of sensor fault detection and identification," *Journal of Process Control* 15, pp. 321-339, 2005.

[25] P. Mosterman and G. Biswas," Diagnosis of continuous valued systems in transient operating regions," *IEEE Trans. on Systems, Man and Cybernetics,* 29(6):554—565, 1999.

[26] S. Moral, R. Rum, and A. Salmeron,"Mixtures of truncated exponentials in hybrid Bayesian networks," In Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Volume 2143 of Lecture Notes in Artificial Intelligence, pp. 145-167. Berlin, Germany: Springer-Verlag, 2001.

[27] A. Mukerjee and A.K. Samantaray,"System modeling through bond graph objects on symbols 2000," International Conference on bond graph Modeling and Simulation (ICBGM'01), Vol. 33, pp. 164-170, Simulation series, 2001.

[28] Naval SurfaceWarfare Center Carderock Division," Handbook of Reliability Prediction Procedures for Mechanical equipment", CARDEROCKDIV, NSWC-07, September 28, 2007.

[29] B. Ould-bouamama, K. Medjaher , A.K. Samantary et M. Staroswiecki ,"Supervision of an industrial steam generator Part I and II," *Control Engineering Practice* 14 (pp. 71-83), 2006.

[30] B. Ould-bouamama, M. Staroswiecki and A.K. Samantaray," Software for Supervision System Design In Process Engineering Industry,". 6th IFAC, SAFEPROCESS, pp. 691-695.Beijing, China, 2006.

[31] R. J. Patton and J. Chen," Observer-based fault detection and isolation : Robustness and applications," *Control Engineering Practice*, 5(5):671-682, 1997.

[32] H. Paynter," Analysis and Design of Enginnering Systems," MIT press. 1961.

[33] J. Pearl," Probabilistic reasoning in intelligent systems: Networks of plausible inference," Morgan Kaufman Publishers, Inc., San Mateo,CA, 2nd edition, 1988.

[34] H. Rinne," The Weibull Distribution A Handbook," Justus-Liebig-University Giessen, Germany. ISBN 978-1-4200-8743-7, CRC press, Taylor & Francis Group, 2009.

[35] I. Roychoudhury, G. Biswas & X. Koutsoukos. *A Bayesian Approach to efficient Diagnosis of Incipient Faults.* 17th International Workshop on Principles of Diagnosis (DX 06), pp. 243-250, Spain. 2006

[36] A. K. Samantaray and B. Ould-bouamama," Model-based Process Supervision: A Bond Graph Approach," Springer-Verlag ISBN 978-1-84800-158-9, 2008.

[37] Z. Shi, F. Gu, B. Lennox, and A.D. Ball," The development of an adaptive threshold for model-based fault detection of a nonlinear electro-hydraulic system," *Control Engineering Practice*, 13:1357-1367, 2005.

[38] M. Tagina, "Application de la Modèlisation Bond Graph À la Surveillance Des Systèmes Complexes, " Thèse de doctorat, Université des Sciences et Technologies de Lille, France, 1995.

[39] R. Wang," Statistical theory," Xian Jiaotong University Press China, 2003.

[40] P. Weber, D. Theilliol, C. Aubrun, A. Evsukoff," Increasing effectiveness of model-based fault diagnosis: A dynamic bayesian network design for decision making," 6th IFAC SAFEPROCESS, Beijing, RP China, August, 2006.

[41] X. Zhang, K. A. Hoo, "Effective fault detection and isolation using bond graph-based domain decomposition," Computers and Chemical Engineering 35, pp. 132-148, 2011

Modeling Variation of Performance Metric of Distributed Memory Heterogeneous Parallel Computer System Using Analytic and Recursive Models

Osondu E. Oguike[1], Monica N. Agu[2] and Stephenson C.Echezona[3]

Department of Computer Science
University of Nigeria
Nsukka, Enugu State
Nigeria

e-mail:
[1] osondu.oguike@unn.edu.ng
[2]monica.agu@unn.edu.ng
[3]stephenson.echezona@unn.edu.ng

***Abstract*— In a heterogeneous parallel computer system, the computational power of each of the processors differs from one another. Furthermore, with distributed memory, the capacity of the memory, which is distributed to each of the processors, differs from one another. Using queuing system to describe a distributed memory heterogeneous parallel computer system, each of the heterogeneous processors will have its own heterogeneous queue. The variation of a performance metric of heterogeneous parallel computer system with distributed memory needs to be modeled because it will help designers of parallel computer system to determine the extent of variation of a performance metric. It will also help users to know when to realize minimum variation of a performance metric. This paper models the variation of a performance metric of distributed memory heterogeneous parallel computer system using analytic and recursive models.**

Keywords - heterogeneous parallel computer, distributed memory, parallel computer system, queuing network, variation, recursive model, analytic models

I. INTRODUCTION

A heterogeneous parallel computer system is one in which the computational power of each of the processors differs from one another. With distributed memory, it means that each of the heterogeneous processors has its own memory. Describing the system using queuing network, each of the processors has its own queue. With a round robin scheduling algorithm, processes can be scheduled to the various parallel processors, whenever a process needs to perform an I/O operation, it joins the appropriate I/O queue. Therefore, the queuing network of a heterogeneous parallel computer system consists of parallel processors, parallel processor queues, I/O processors and I/O queues. Suppose there are n different parallel processor queuing systems and k different I/O queuing systems. A queuing system in this context is defined as a processor, together with its own queue. We also assume that the various queues are finite [1, 2, 3, 4] i.e. there is a limit to the number of jobs that can be admitted into the queues, and negligible communication overhead. Suppose X_1, X_2, X_3, … , X_n , X_{n+1}, X_{n+2}, X_{n+3}, …, X_{n+K} are the maximum number of processes that can be admitted into the respective queues. We assume that processes arrive at the various queues according to Poisson distribution, and they are serviced according to exponential distribution [5, 6]. Figure 1 illustrates a model of the queuing network of a heterogeneous parallel computer system with distributed memory.

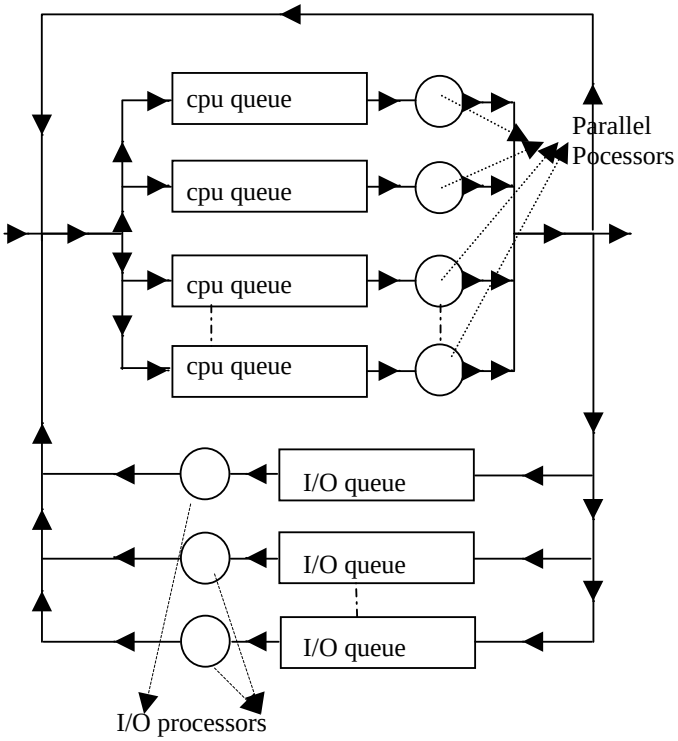


Figure 1: Queuing network of a heterogeneous parallel computer system with distributed memory.

II. Statement of the Problem

Though analytic queuing method can be used to model the performance of a heterogeneous parallel computer system with distributed memory, however, it may not be possible to model some performance metrics of heterogeneous parallel computer, like variation of a performance metric. The reason is because the analytic method cannot determine the exact convergence of some mathematical series that are used in modeling the variation of a performance metric of a heterogeneous parallel computer system. Therefore, there is the need for the use of another model, rather than analytic model. The use of efficient linear recursive model [9] can efficiently model the variation of a performance metric of a distributed memory heterogeneous parallel computer system, because recursive models can be used to determine the exact convergence of any series used in modeling the variation of a performance metric of distributed memory parallel computer system.

III. LITERATURE REVIEW

Queuing approach has been used extensively in the literature to model the performance of computer systems. However, this has been done in different ways and for different models of computer systems. In [20], the authors used a recursive computation approach to solve the steady state equations, thereby leading to the modeling of the various performance metrics of a multi-terminal system that is subject to breakdown. Furthermore, the author in [24] used a rigorous approach to model the performance of heterogeneous parallel computer system without introducing any constraint on the kind of interconnection between the heterogeneous nodes. Furthermore, in [24], systems with the same interconnection speed were considered when modeling the performance of heterogeneous parallel computer system. The authors in [25] looked at alternative ways of measuring the performance of heterogeneous parallel computer system, by modeling linear speed and linear efficiency using simulation-modeling techniques. In [26], the author showed that Little's formulae could be universally applicable, if properly interpreted to take account of state-varying entrance rates, batch arrivals, and multiple customer classes. In [27], the author confirmed that Little's formula could be applied to very general queuing systems (not just M/M/1), even whole networks! The authors in [28] considered a new performance metric, variation of the computing power as a unique performance metric that is ideal for a heterogeneous network of workstations, though an approach different from queuing approach was used to do this. In [29], analytic models were used to model the performance of computer intensive applications of parallel computers, while [30] used recursive models only to evaluate the performance of compute intensive application of a parallel computer system. In [31], recursive models were used to evaluate the performance of heterogeneous parallel computer system with distributed memory.

IV. Methodology

This paper models the variation of a performance metric of distributed memory heterogeneous parallel computer system. A queuing approach, with finite queues has been used to achieve the above aim, with parallel processors depicting parallel servers. The statistical method of probability density function and other probability theory concepts have used [15, 23]. A novel method of deriving the recursive model that determines the xth terms and the convergence of important mathematical series has been used to develop the recursive models. The simulation of the models on the computer has been done using Java programming language and the statistical regression/trend line analysis has been used to analyze the results of the simulation [11].

V. Developing the Models

As a result of the use of the above methodologies, the following models have been developed for one queuing system and for all the queuing systems of the queuing network.

A. Models Based on a Queuing System

The following models have been developed for one queuing system

- Probability Density Function of the Number of Processes in a Queue.

Let X_i denotes the maximum number of processes that can be in the ith finite queuing system at any time [12, 13, 14]. Suppose the arrival rate, λ_{xi} when x_i processes are in the ith queuing system of the queuing network be described as:

$$\lambda_{x_i} = \begin{cases} \lambda_i, & x_i = 0,1,2,,3,...X_i - 1 \\ 0, & otherwise \end{cases} \quad (1)$$

Since the various processors are heterogeneous, therefore, it implies that the departure rate will vary, which can be described as:

$$\mu_{x_i} = \begin{cases} \mu_i, & x = 1,2,3,4,..., X_i \\ 0, & otherwise \end{cases} \quad (2)$$

Using the steady state probability as stated in [7, 16] the probability that x_i processes will be in the ith queuing system is

$$P_{x_i} = \begin{cases} \rho_i{}^x P_{0i}, & x \le X_i \\ 0, & otherwise \end{cases} \quad (3)$$

The utilization factor for the ith queuing system, ρ_i is defined as:

$\frac{\lambda_i}{\mu_i}$. To obtain the value of P_{0i} in equation (3), we sum all the probabilities for the ith queuing system and equate it to 1. This implies that:

$$\sum_{x_i=0}^{X_i} P_{xi} = 1. \quad (4)$$

From equations (3) and (4), it implies that:

$P_{0i}+\rho_i P_{0i}+\rho_i^{\,2}P_{0i} + \rho_i^{\,3}P_{0i} + \rho_i^{\,4}P_{0i} + \ldots + \rho_i^{\,X_i}P_{0i} = 1.$ (5)

Factorizing equation (5), it implies that

$P_{0i}\,(1 + \rho_i + \rho_i^{\,2} + \rho_i^{\,3} + \ldots + \rho_i^{\,X_i}) = 1.$ (6)

Analytically, the series in equation (6) converges to: $\frac{1-\rho_i^{\,X_i+1}}{1-\rho_i}$, provided that $\rho_i \neq 1$, otherwise it converges to $X_i + 1$.

It implies that $P_{0i}\left(\frac{1-\rho_i^{\,X_i+1}}{1-\rho_i}\right) = 1$, when $\rho_i \neq 1$ and $P_{0i}(X_i + 1) = 1$, when $\rho_i = 1$. Solving for P_{0i}, we have that:

$$P_{oi} = \begin{cases} \frac{1-\rho_i}{1-\rho_i^{\,X_i+1}} & \rho_i \neq 1 \\ \frac{1}{X_i+1}, & \rho_i = 1 \end{cases} \quad (7)$$

Using equation (7) in equation (3), we have the following:

$$P_{xi} = \begin{cases} \frac{\rho_i^{x_i}(1-\rho_i)}{1-\rho_i^{X_i+1}}, & \rho_i \neq 1, \quad x = 0,1,2,3..X_i \\ \frac{1}{X_i+1}, & \rho_i = 1 \end{cases} \quad (8)$$

Equation (8) is the probability density function that models the probability that x_i processes will be admitted in the ith queuing system.

- Average Number of Processes in One Queuing System.

Furthermore, the average number of processes in the ith queuing system (i.e the queue and the processor) can be described statistically as expectation of x_i, where x_i is the random variable that denotes the number of processes in the ith queuing system. This can be written as

$$E(x_i) = \sum_{x_i=0}^{X_i} x_i P_{i\,x_i}. \quad (9)$$

Using equation (8) in equation (9), we obtain the following:

$$E(x_i) = \sum_{x_i=1}^{X_i}\left(\frac{x_i\rho_i^{x_i}(1-\rho_i)}{1-\rho_i^{X_i+1}}\right), \quad \rho_i \neq 1, \quad x_i = 0,1,2,3..X_i \quad (10)$$

Equation (10) can be simplified as:

$$E(x_i) = \left(\frac{1-\rho_i}{1-\rho_i^{X_i+1}}\right)\rho_i\left(1+2\rho_i+3\rho_i^{2}+4\rho_i^{3}+\ldots+X_i\rho_i^{X_i-1}\right) \quad (11)$$

A recursive model has been used in [30,31] to determine the convergence of the series in equation (11). The recursive model is called $\text{Sum2}_i(X_i, \rho_i)$, and it is given as:

$$\begin{cases} 1, & X_i = 1 \\ \text{Term2}_i(X_i)*\text{term1}_i(X_i-1, \rho_i) + \text{Sum2}_i(X_i-1, \rho_i), & X_i \neq 1 \end{cases} \quad (12)$$

$\text{Term2}_i(X_i)$ and are given as:

$$\text{Term2}_i(X_i) = \begin{cases} 1, & X_i = 1 \\ 1 + \text{term2}_i(X_i-1), & X_i \neq 1 \end{cases} \quad (13)$$

and $\text{term1}_i(X_i, \rho_i)$ is defined below as:

$$\text{Term1}_i(X_i, \rho_i) = \begin{cases} 1, & X_i = 0 \\ \rho_i * \text{Term1}_i(X_i-1, \rho_i), & X_i \neq 0 \end{cases} \quad (14)$$

Therefore, using equation (12) in equation (11), we obtain:

$$E(x_i) = \left(\frac{1-\rho_i}{1-\rho_i^{X_i+1}}\right)\rho_i\ \text{Sum2}_i(X_i, \rho_i) \quad (15)$$

B. Models Based on The Whole Queuing Network.

Having developed the models for the performance metrics of one queuing system, these models can be extended to the whole queuing systems of the queuing network of a heterogeneous parallel computer system. It is necessary to define δ_i as the probability that a process will join the ith queue after each cpu burst, and δ_0 as the probability that the execution of a process has been completed. Arrival of processes into the various parallel processor queues can come from the outside world or from the various I/O queues or from the particular parallel processor, at the expiration of the time quantum for that process. Let λ_i be the rate of arrival of processes into the ith queuing system, and λ, the rate of arrival of processes from the outside world.. Under the steady state, when we

consider the queuing network, the overall utilization factor has been defined in [31] as:

$$\rho_i = \begin{cases} \dfrac{\lambda}{\delta_0 \mu_i}, & i = 0 \\ \dfrac{\lambda \delta_i}{\delta_0 \mu_i}, & i = 1,2,3,\ldots,n+k \end{cases} \quad (16)$$

- Variation of Average Number of Processes in all the Queuing Systems of the Queuing Network

Suppose x_i is the random variable that denotes the number of processes in the ith queuing system. Therefore, another random variable, Y, can denote the average number of processes in the queuing systems of the queuing network, as:

$$Y = \frac{\sum_{i=1}^{n+k} x_i}{n+k} \quad (25)$$

Therefore, the variation of the average number of processes in all the queuing systems of the queuing network can be defined statistically as:

$$VAR\,(Y) = VAR\left(\frac{\sum_{i=1}^{n+k} x_i}{n+k}\right) \quad (26)$$

Using one of probability theory laws in [23], we obtain:

$$VAR(Y) = \frac{1}{(n+k)^2}\sum_{i=1}^{n+k} VAR(x_i) \quad (27)$$

From [23], the variance can be defined statistically as:

$$VAR(x_i) = E(x_i^{\,2}) - (E(x_i))^2 \quad (28)$$

Simplifying equation (28) further, we obtain:

$$E(x_i^{\,2}) = \sum x_i^{\,2} P_{i\,x_i} \quad (29)$$

Using equation (8) in equation (29), we obtain:

$$E(x_i^{\,2}) = \sum_{x_i=1}^{X_i} \left(\frac{x_i^{\,2}\rho_i^{\,x_i}(1-\rho_i)}{1-\rho_i^{\,X_i+1}}\right) \quad (30)$$

Simplifying equation (30), we obtain:

$$E(x_i^2) = \left(\frac{1-\rho_i}{1-\rho_i^{X_i+1}}\right)\left(1^2\rho_i + 2^2\rho_i^2 + 3^2\rho_i^3 + 4^2\rho_i^4 + \ldots + X_i^2\rho_i^{X_i}\right) \quad (31)$$

Simplifying equation (31) further, we obtain:

$$E(x_i^2) = \left(\frac{1-\rho_i}{1-\rho_i^{X_i+1}}\right)\left(1\rho_i + 4\rho_i^2 + 9\rho_i^3 + 16\rho_i^4 + \ldots + X_i^2\rho_i^{X_i}\right) \quad (32)$$

Factorising equation (32), we obtain:

$$E(x_i^{\,2}) = \left(\frac{1-\rho_i}{1-\rho_i^{\,X_i+1}}\right)\rho_i\left(1+4\rho_i+9\rho_i^{\,2}+16\rho_i^{\,3}+\ldots+X^2{}_i\,\rho_i^{\,X_i-1}\right) \quad (33)$$

The convergence of the series may be difficult or impossible to obtain analytically; therefore we seek for its convergence using recursive models. The same approach used earlier can be used to determine the convergence of the series, $\left(1+4\rho_i+9\rho_i^{\,2}+16\rho_i^{\,3}+\ldots+X^2{}_i\,\rho_i^{\,X_i-1}\right)$. The series can be considered as two sequences, which are: sequence1 = 1, 4, 9, 16, …, X^2, while the other sequence is: sequence2 = $\left(1, \rho_i, \rho_i^{\,2}, \rho_i^{\,3}, \ldots, \rho_i^{\,X_i-1}\right)$. The recursive model that can be used to determine the xth terms of sequence1 can be obtained by adding 2X-1, which is the common difference between the xth term and the (x-1)th term to the (x-1)th term of the sequence. The recursive model can be represented as shown below in equation (34), as:

$$\mathrm{Term3_i}(X_i) = \begin{cases} 1, & X_i = 1 \\ (2*X_i\text{-}1) + \mathrm{Term3_i}(X_i\text{-}1), & X_i \neq 1 \end{cases} \quad (34)$$

The recursive model that determines the xth terms of sequence2 has been developed in equation (14).Therefore, combining equation (30) and equation (14), the series in equation (33) converges to this recursive model, called $\mathrm{Sum3_i}(X_i, \rho_i)$, which is shown below as:

$$\begin{cases} 1, & X_i = 1 \\ \mathrm{Term1_i}(X_i\text{-}1, \rho_i)*\mathrm{Term3_i}(X_i) + \mathrm{Sum3_i}(X_i\text{-}1, \rho_i), & X_i \neq 1 \end{cases} \quad (35)$$

Therefore, using equation (35) in equation (33), we obtain:

$$E(x_i^{\,2}) = \left(\frac{1-\rho_i}{1-\rho_i^{\,X_i+1}}\right)\rho_i Sum3(X_i, \rho_i) \quad (36)$$

Using equations (15) and (36) in equation (28), we obtain:

$$VAR(x_i) = \left(\frac{\rho_i(1-\rho_i)(Sum3(X_i,\rho_i))}{(1-\rho_i^{\,X_i+1})}\right) - \left(\frac{\rho_i(1-\rho_i)(Sum2(X_i,\rho_i))}{(1-\rho_i^{\,X_i+1})}\right)^2 \quad (37)$$

Therefore, using equation (37) in equation (27), we obtain:

$$VAR(Y) = \frac{1}{(n+k)^2}\left(\sum_{i=1}^{n+k}\left(\left(\frac{\rho_i(1-\rho_i)(Sum3(X_i,\rho_i))}{(1-\rho_i^{\,X_i+1})}\right) - \left(\frac{\rho_i(1-\rho_i)(Sum2(X_i,\rho_i))}{(1-\rho_i^{\,X_i+1})}\right)^2\right)\right) \quad (38)$$

Equation (38) models the variation of the average number of processes in all the queuing systems of the queuing network by combining analytic and recursive models.

VI. RESULTS OF THE SIMULATION

The result of the simulation was analyzed to determine how variation of the performance metric under consideration changes as a particular parameter varies, while other parameters remain constant [10]. Figure 2 shows the result of the simulation, if the probability of a process leaving the system is known to be 0.2 and the probabilities that a process will join the first and second queues are 0.775 and 0.025, respectively. Suppose the first processor is a high-speed processor with high departure rate of 30, while the second processor is a low speed processor with a low departure rate of 10. Suppose the arrival rate in the system from the outside world is 4 and maximum number of processes to be allowed into first queue is 20, while maximum number of processes to be allowed into the second queue is 5. The experimental trials were carried out several times, in each trial, the arrival rate was changed, and the corresponding variation was obtained as the result of the simulation.

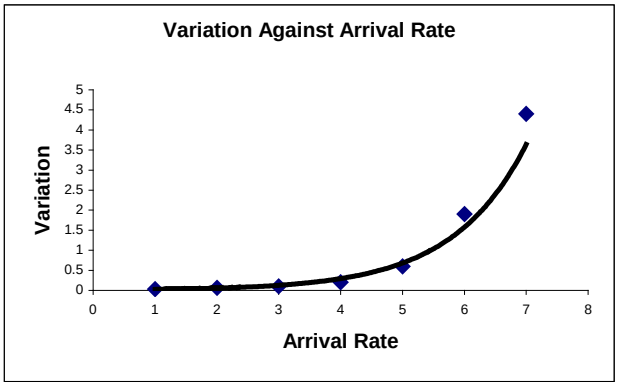


Figure 2: Variation Against Arrival Rate

The result shows that for non-compute intensive application, where the overall utilization factor of each of the processors is less than 1, increasing the departure rate will lead to a corresponding increase in the variation of the performance metric of the heterogeneous parallel computer.

Furthermore, figure 3 shows the simulation result when as we keep the following input parameters constant, the probability that a process will leave the network is 0.2, the probabilities that a process will join queue 1 and 2 are 0.775 and 0.025, respectively. The departure rates for processor 1 and 2 are 30 and 10, respectively, while the maximum number of processes in queue 1 and 2 (degree of multiprogramming for the two queues) are 20 and 5, respectively, and the arrival rate from the outside world is 4. By varying the degree of multiprogramming (maximum number of processes in the system) for the two queues of a two-processor parallel computer system, we obtain the corresponding variation shown in figure 3.

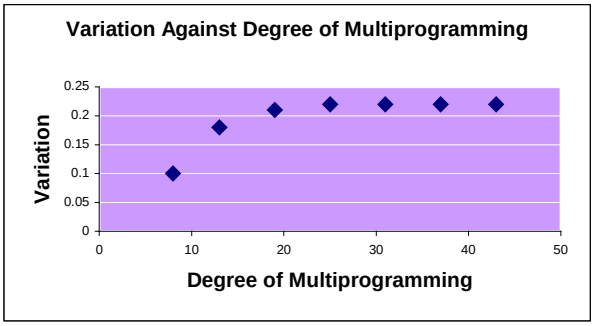


Figure 3: Variation Against the Degree of Multiprogramming

The interpretation of the result is that for non-compute intensive applications, where the overall utilization factor for each of the queues is less than 1, increasing the maximum number of processes that can be in each of the queues (the degree of multiprogramming) will lead to corresponding increase in the variation of the performance metric.

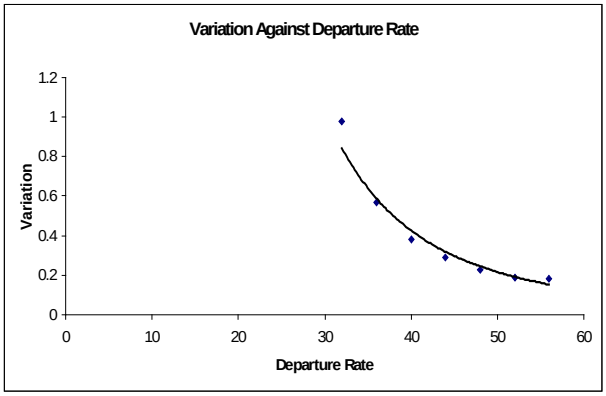


Figure 4 Variation Against Total Departure Rate

In a similar manner, as we keep the following input parameters constant, probability of a process leaving the network is 0.2, while the probability of a process going to queue 1 and 2 is 0.4. The maximum number of processes that can be in queue 1 and 2 are 15 and 14, respectively. By changing the departure rates of the two processors, we obtain the corresponding variation of the performance metric, as shown in figure 4. The result shows that for non-compute intensive application, increasing the speed of the various processors will lead to a corresponding decrease in the variation of the performance metric under consideration.

VII. SUMMARY AND CONCLUSION

This paper has been able to model the variation of a performance metric of heterogeneous parallel computer system with distributed memory by combining analytic and recursive models, using queuing approach. The models have been simulated on the computer and the results of the simulation have been analyzed in order to determine when to realize minimum variation.

REFERENCES

[1] Henry H. Liu and Pat V. Crain, An Analytic Model for Predicting the Performance of SOA-Based Enterprise Software Applications, Proc. International Conference of Computer Measurement Group, (2004).

[2] S. Balsamo et al, A Review of Queueing Network Models with Finite Capacity Queues for Software Architecture Performance Prediction, (2002).

[3] Catalina M. Liado et al, A Performance Model Web Service, Proc. International Conference of Computer Measurement Group, (2005).

[4] Rosselio, J et al, A Web Service for Solving Queueing Network Models Using PMIF. www.perfeng.com/paperndx.htm, (2005).

[5] Cathy H. Xia, Zhen Liu., Queueing systems with long-range dependent input process and subexponential service time. Proc. ACM SIGMETRICS international conference on Measurement and modeling of computer systems,(2003).

[6] Shanti Subramanyam, Performance Modelling of a J2EE Application to meet Service Level s, Agreement, Proc. International Conference of Computer Measurement Group, (2005)

[7] Hamdy A. T.,. Operation Research: An Introduction, Prentice-Hall of India, (1999).

[9] Ivan Stojmenovic; Recursive Algorithms in Computer Science Courses : Fibonacci Numbers and Binomial Coefficients; IEEE Transactions on Education; Vol. 48, No. 3

[10] Arjan J.C. van Gemund; Performance Modelling of Parallel Systems: An Introduction.

[11] Justyna Berlinska, The Statistical models of parallel applications, Annales UMCS Informatica, (2005).

[12] Arranchenkov, K.E., Vilchersky, N.O., Shevlyakor, G.L Priority queueing with finite buffer size and randomized push-out; mechanism. Proc. of ACM SIGMETRICS international conference on measurement and modeling of computer systems.; (2003).

[13] Abunday, B.D., and Khorram, E. The finite source queueing model for multiprogrammed computer systems with different CPU times and different I/O times. Acta Cybern. 8, 4 , (1998)

[14] J. Sztrik; Finite-Source Queueing Systems and their Applications: A Biliography;

[15] Trivedi K. Shridharbhai, Probability and Statistics with Reliability, Queuing and Computer Science Applications, John Wiley & Sons Inc., (2002).

[16] Per Brinch Hansen. Operating System Principles. Prentice-Hall of India Private Limited, (1990).

[20] J. Sztrik[a] and T. Gál A recursive solution of a queueing model for a multi-terminal system subject to breakdowns; Performance Evaluation Volume 11, Issue 1, Published by Elsevier, (1990).

[23] Robert V. Hogg and Allen T. Craig; Introduction to Mathematical Statistics; Macmillan Publishing Co. Inc.; (1978).

[24] Andrea Clemantis, Angelo Corana; Modelling Performance of Heterogeneous Parallel Computer System; Journal of Parallel Computing, Volume 12, Issue 9, Elsevier; pages 1131-1145; (1999).

[25] E. Post, H.E. Goosen; Evaluating the Parallel Performance of a Heterogeneous System

[26] Beutler, F; Mean sojourn times in markov queuing network: Little's formula revisited; IEEE Transaction on Information Theory; Volume 29, Issue 2, page 233-241; (2003).

[27] Ken Vastola; http://networks.ecse.rpi.edu/~vastola/pslinks/perf/node46.html

[28] Xiaodong Zhang, Yong Yan; Modeling and Characterizing Parallel Computing Performance on Heterogeneous Network of workstations; Proceedings of the 7th IEEE Symposium on Parallel and Distributeed Processing (SPDP '95) 1063-6374/95 $10.00 © 1995 IEEE

[29] O.E. Oguike et al; Modelling the Performance of Computer Intensive Applications of Parallel Computer System; Proc. Of IEEE 2nd International Conference on Computational Intelligence, Modeling and Simulation; (2010).

[30] O.E. Oguike et al; Evaluating the Performance of Parallel Computer System Using Recursive Models; Proc. Of IEEE 4th UKSim European Modeling Symposium; (2010).

[31] O.E. Oguike et al; Evaluating the Performance of Heterogeneous Distributed Memory Parallel Computer System Using Recursive Models; 2nd IEEE International Conference on Intelligent Systems, Modeling and Simulation; (2011).