FU JEN STUDIES

SCIENCE AND ENGINEERING

NO. 33, DEC. 1999

CONTENTS

輔仁學誌—理工類

中華民國八十八年十二月　　第三十三期

目　錄

Detection of DNA-damaging Agents in the Environment by Using Bacteria Carrying a *recA-luxAB* or a *lexA-luxAB* Fusion

Mei-Kwei Yang, Pei-I Wu and Tien-Chun Yang

Department of Biology

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

The DNA damage inducible *recA* and *lexA* genes of *Xanthomonas campestris* pathovar citri has been fused to the *luxAB* genes from *Vibrio fischeri* that encode the enzyme luciferase. Cells transformed with this transcriptional fusions responded to DNA-damaging agents with an increase in luciferase activity. A marked bioluminescence response was observed when cells were exposed to samples from industrial waste water. These results demonstrate that these two *lux* biosensors may be used as tools for detecting DNA-damaging agents in the environmental samples

Key Words: *Xanthomonas campestris* pathovar citri; *recA*; DNA damage; gene promoter; transcriptional fusion

INTRODUCTION

Direct determination of specific chemicals and biological toxicity assays are two common approaches to detection of environmental contaminants. There is good evidence that biochemical and microbial assays have gained more recognition than bioassays based on animal models or tissue culture techniques. In response to environmental stress, organisms encounter significant metabolic alternation or cellular

adaptive response via the induction of specific gene expression (Walker, 1984). The detection of transcriptional activation is achieved by fusion of promoter elements to reporter genes. A particularly useful reporter system is made up of the structural genes, *luxAB*, of the bioluminescence operon derived from the marine bacterium *Vibrio fischeri* (Meighen, 1991; Stewart and Williams, 1992). *Escherichia coli* strains containing the promoters of heat shock or other stress-induced genes fused to the *lux* reporter increase bioluminescence in response to many chemicals (Van Dyk et al., 1994).

Many industrial wastes and manufacturing sites contain genotoxic agents that cause DNA aberration directly or damage DNA indirectly. Thus, the monitoring of the presence of mutagens in the environment is very important. A more rapid, simple and inexpensive screening method where mutagenicity is assayed by light emission is therefore developed (Justus and Thomas, 1998; Min et at., 1999; Vollmer et al., 1997). In constrast to Ames test, a bacterial mutagenicity test, advances in molecular biology allowed genetically manipulated bacteria to lead the improvemnet in this field. When the cells' DNA has been damaged, the repair system is turn on to ensure cell survival. In *E. coli*, the genes of the SOS global regulatory response which are involved in error-prone DNA repair following either chemical or physical DNA damage are activated (Walker, 1984; Miller and Kokjohn, 1990). Reporter genes fused to SOS genes as the basis for generating new mutagenicity tests are therefore developed (Quillardet et al., 1982; Ptitsyn, 1996; Manen et al., 1997). Examples include *recA*, *uvrA*, *alkA*, and *umuC'* fused to *lux* genes to measure the mutagenic and genotoxic effects of various pollutants (Smith, 1985; Oda et al., 1985; Reifferscheid et al., 1991; Hill et al., 1993; Meighen and Dunlap, 1993; Van Dyk et al., 1995; Vollmer et al., 1997; Willardson et al., 1998)

We have previously shown that a *recA* mutant of *Xanthomonas campestris* pathovar citri (*X.c.* pv. citri), the bacterium responsible for citrus canker disease, exhibits increased sensitivity to ultraviolet (UV) irradiation and methylmethane sulfonate (MMS) and is unable to undergo homologous recombination. We have localized and characterized a region of the promoter of *recA* that is responsive to DNA damage. In addition, the *lexA* gene which product regulates the expression of *recA* has

also identified. In this study, *X. c.* pv. citri harboring the *recA* or *lexA* promoter region fused to *luxAB* originating from *Vibrio fischeri*, was used for detecting DNA-damaging agents by monitoring bioluminescent emission.

MATERIALS AND METHODS

1. Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used in this study are listed in Table 1. *Xanthomonas campestris* pv. citri strains were grown in Luria-Bertani medium at 28℃. Tetracycline or chloramphenicol was added to culture medium at final concentrations of 25 and 30 $\mu g\ ml^{-1}$, respectively. Competent *X. c.* pv. citri cells were prepared and subjected to electroporation as previously described (Yang et al., 1991). All recombinant DNA protocols and other DNA manipulations were performed by standard procedures (Sambrook et al., 1989). Southern blot analysis was performed with DNA probes labeled with digoxigenin by random priming.

2. Preparation of *recA-luxAB* and *lexA-lexA* fusion constructs

Promoter of *X. c.* pv. citri *recA* was excised by digestion with *Eco*RI and *Ava*II, and ligated into plasmid pMY3 digested with *Hin*dIII. pMY3 contains the promoterless *luxAB* genes just downstream of a multiple cloning site (Weng et al., 1996). A DNA fragment, containing the 5′ untranslated sequence of *X. c.* pv. citri *lexA*, was also isolated from plasmid pC1900 and fused to the pMY3 to generate the plasmid containing the promoter sequence of *lexA*. The presence of the desired inserts was confirmed by DNA sequencing of each fusion product. All plasmids containing the fusion constructs were introduced into *X. c.* pv. citri cells by electroporation and tetracycline-resistant transformants were selected.

3. Induction of luciferase by mitomycinC and other DNA-damaging agents

X. c. pv. citri XW47 cells harboring the pREP426 or pLEP320 plasmid were grown to log phase to an optical density at 600 nm (OD_{600}) of 1.0. Increasing amounts

Table 1. Bacterial strains and plasmids.

Strain or plasmid	Relevant genotype or characteristics	Source or reference
Strain		
XW47	Virulent citrus canker type strain of *X. c.* pv. citri	Yang et al., 1991
Plasmid		
pBC SK$^+$	Phagemid derived from pUC19. Used as a cloning vector.	Stratagene
pMY3	*Tc^r*, promoter-probe vector with *luxAB* markers	Weng et al., 1996
pC1900	pBC SK$^+$ with an inserted 1.9-kb ClaI fragment containing *X. c.* pv. citri *recA*	This study
pREP426	pMY3 with an inserted 426-bp *Eco*RV-*Ava* II fragment of *X. c.* pv. citri *recA* comprising 287 bp of 5′ untranslated sequence and 139 bp of 5′ coding sequence	This study
pLEP320	*lexA-luxAB* fusion vector, consisting of pMY3 with a 332-bp fragment of *lexcA* inserted into the *Hin*dIII sites	This study

of mitomycin C or other DNA-damaging agents including MMS, nalidixic acid, ethidium bromide and UV irradiation was added to induce the transcription of luciferase. Luciferase activity was detected with a lum inometer (LB953 AutoLumat; EG & G Berthold, Bad Wildbad, Germany). After addition of 100 μl of *n*-decylaldehyde (0.1% suspension in ethanol) to 500-μl samples, bioluminescence was measured over three 10-s intervals. Luminescence was expressed in relative light units (RLU).

4. Testing of industrial waste water

Waste water samples were collected from the effluent at Wu-ku industrial area, Tai-Sun. This area was known to contain various company and was suspeneded to be contaminated with chemicals. Samples were filtered through a disposable 0.2-μm membrane to remove bacteria and used in the *recA-lux* and *lexA-lux* assay on the same

day. These samples were stored at 4℃ and used for luciferase assays for up to 1 week after sampling. Water was tested by adding 1 ml of sample with equal amount of *X. c.* pv. citri cells harboring the pREP426 or pLEP320 plasmid in LB medium at a density of 0.6 OD_{600}. The cells were incubated at 28℃ and then luciferase activity was measured at various intervals. In addition, cell viability was also tested in the presence of the sample water.

RESULTS AND DISCUSSION

1. Time course of luminecent intensity

To test the luminesscent activity of the test bacteria, *X. c.* pv. citri cells carrying the plasmid pMY3-*lac* was measured. Luminescence wes observed upon addition of the luciferase enzyme substrate N-decyl aldehyde to cells grown for 12 h timeucorse. Viable cell counts and optical density of cultures were monitored during the time course to eliminate the possibility of false negative response due to cell death. A significant increase in luminescence level was observed over time within 24 h (Fig. 1). This indicates the need for cells to be in a metabolically active state for *lux* gene expression to occur. Thus, the luminescence observed in this system may be useful in providing quantitative data for detecting gene expression.

2. Induction of *recA-luxAB* and *lexA-luxAB* expression in *X. campestris* pv. citri

pC1900 was digested with *Eco*RV and *Ava*II, a DNA fragment of 426 bp was isolated and ligated into plasmid pMY3 digested with *Hind*III, thereby generating pREP426 (Fig 2). A 320-bp *Ava*II-*Pvu*II restriction fragment, containing the 5′ untranslated sequence of *X. c.* pv. citri *lexA*, was isolated from plasmid pC1900 and fused to pMY3 to generate the plasmid pLEP320 (Fig. 3). The vector pMY3 contains a promoterless copy of the *Vibrio fischeri luxAB* genes and can replicate in *X. c.* pv. citri, in the absence of a cloned promoter, no luciferase activity was detected when this vector was introduced into *X. c.* pv. citri. The pREP426 or pLEP320 plasmid was introduced into strain XW47 by electroporation, and the effects of DNA-damaging

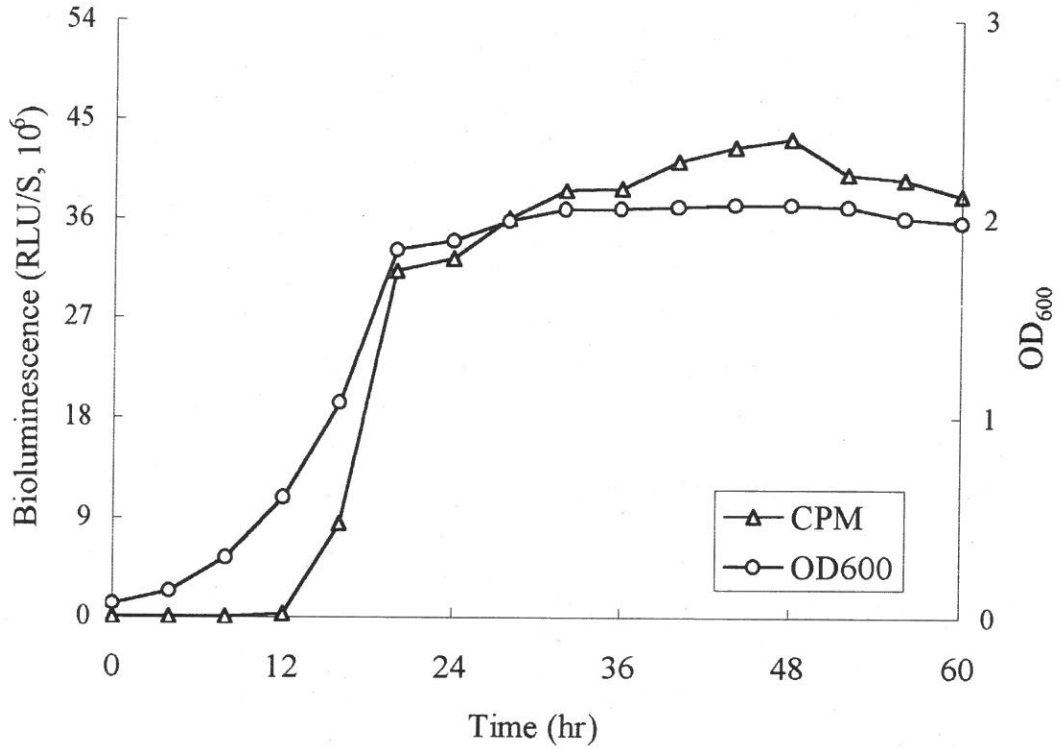


Fig. 1. Luminescence of *X. c.* pv. citri XW47 carrying pMY3-*lac* versus time in LB medium.

agents on luciferase activity were determined.

The expression of *recA-luxAB* was examined in *X. campestris* pv. citri as described previously (Yang and Yang). The time course of luciferase activity in *X. c.* pv. citri cells harboring pLEP320 plasmid is shown in Fig. 4. More than two-fold increase in the luciferase activity was detected by exposure of cells to mitomycin C (0.5 μg ml^{-1}) (Fig. 4A). Various other DNA-damaging agents such as MMS (an alkylating agent), nalidixic acid (a DNA gyrase inhibitor), ethidium bromide (an intercalating agent) and UV irradiation (which induces the formation of pyrimidine dimers) were also shown to increase the transcription of *lexA* in *X. c.* pv. citri cells (Fig. 4, B through E). These results indicate that the transcription of *recA* or *lexA* is increased by exposure of *X. c.* pv. citri to DNA-damaging agents.

3. Effect of waste water on the expression of a *recA-luxAB* fusion construct

To determine the utility of these two fusion constructs in measuring the

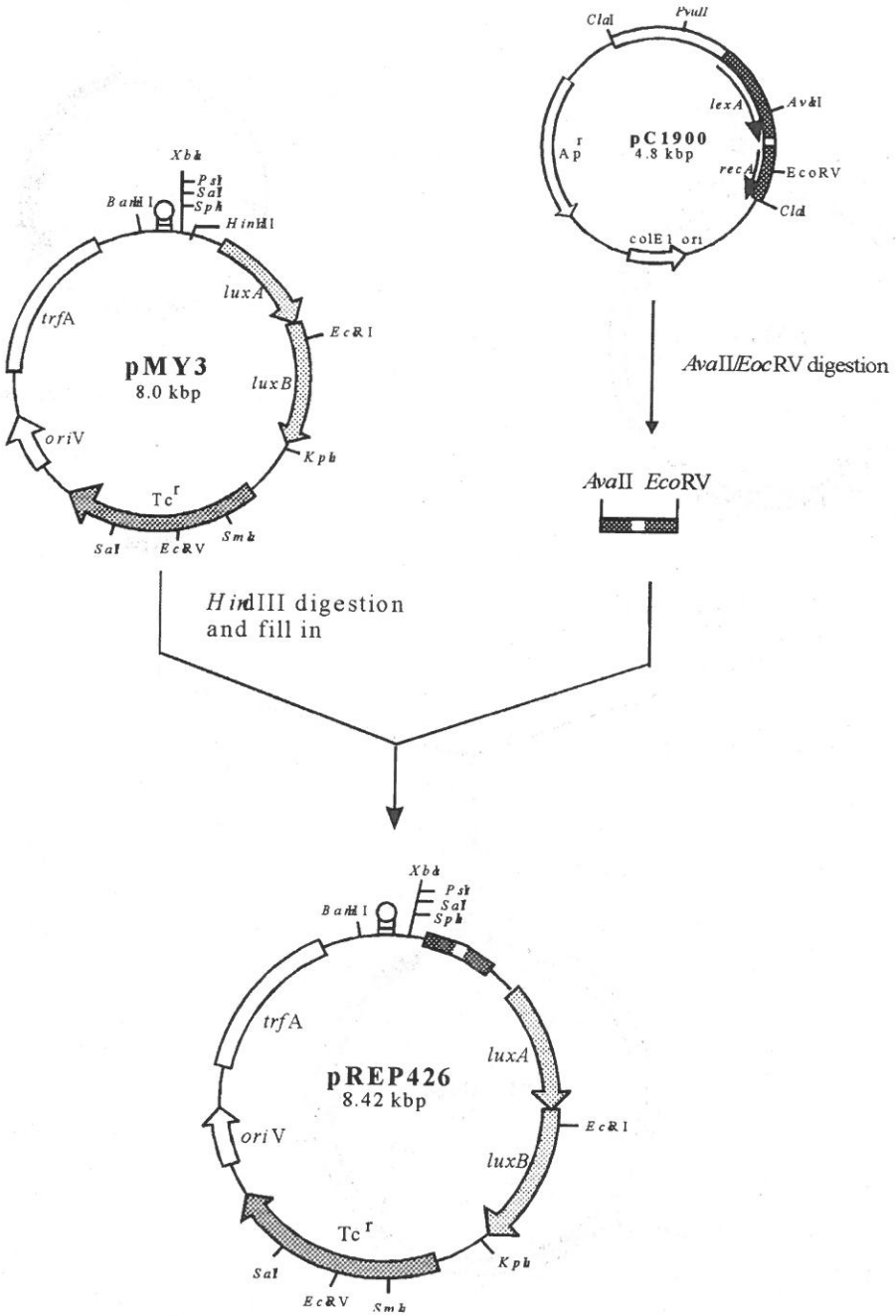


Fig. 2. Construction of a *recA-luxAB* fusion plasmid pREP426. A 426-bp *Eco*RV-*Ava*II fragment of pC1900 containing promoter sequence of *recA* was isolated and ligated to the *Hin*dIII site of plasmid pMY3 to generate pLEP426.

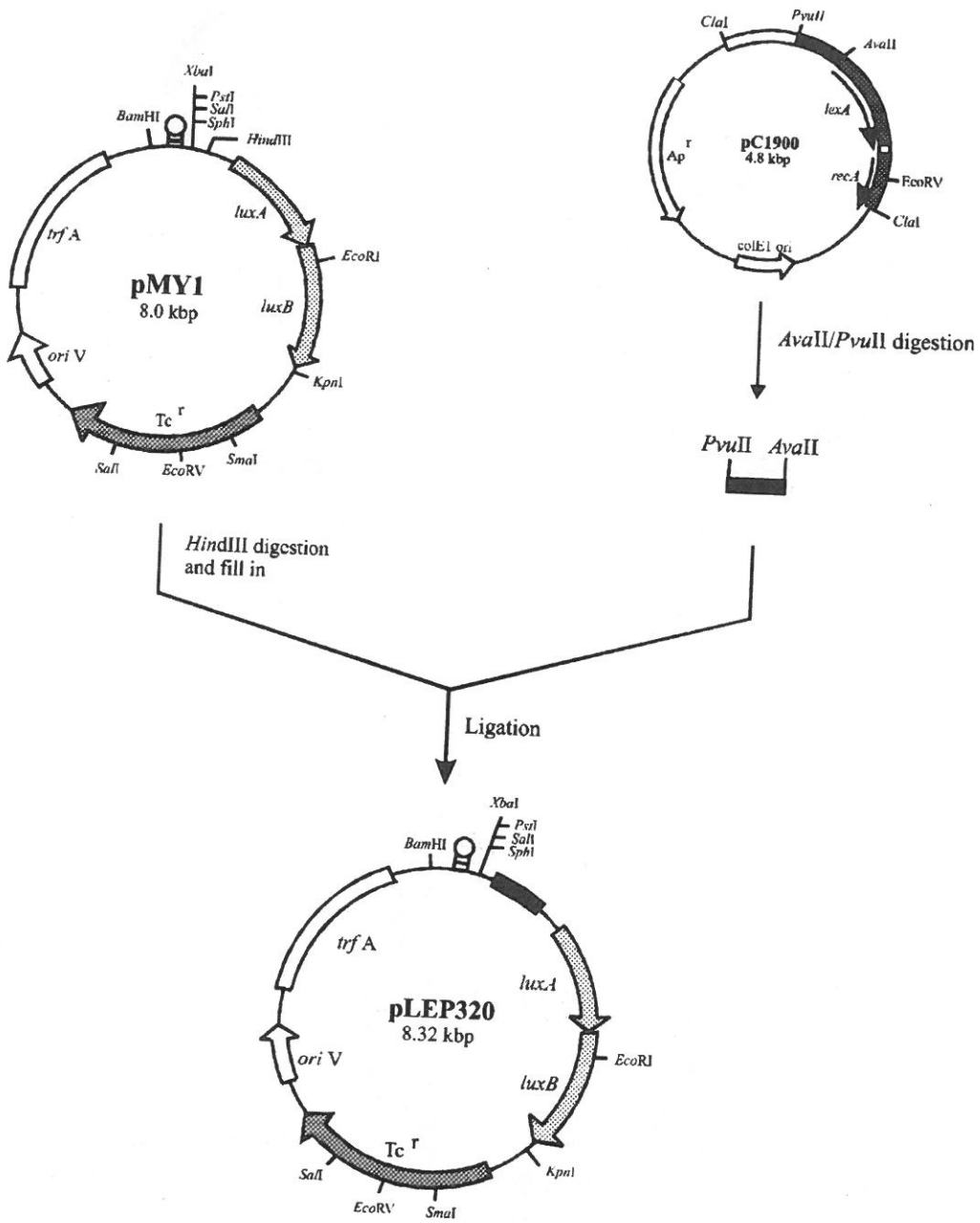


Fig. 3. Construction of a *lexA-luxAB* fusion plasmid pLEP320. A 320-bp *Ava*II-*Pvu*II fragment of pC1900 containing the promoter sequence of *lexA* was inserted into the *Hin*dIII site of plasmid pMY3 to generate pLEP320.

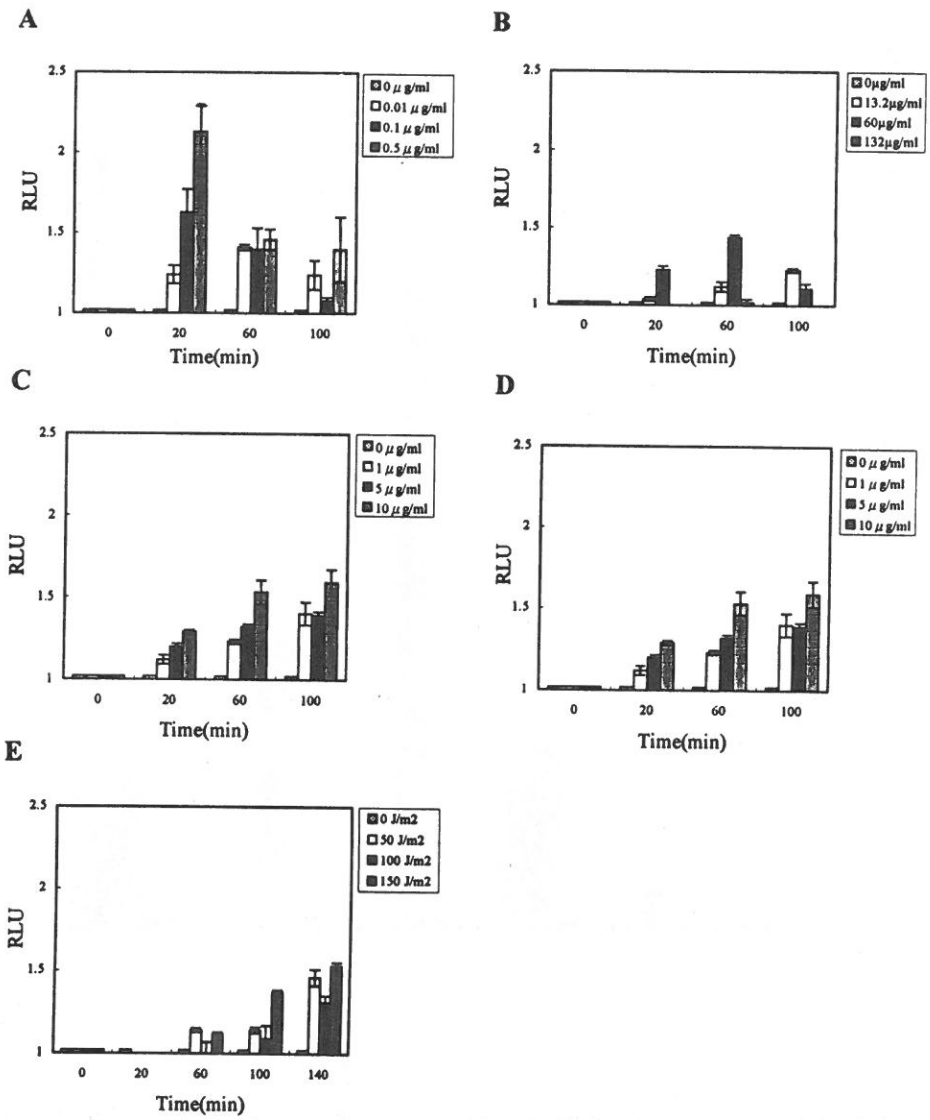


Fig. 4. Effects of DNA-damaging agents on expression of a *recA-luxAB* fusion construct in *X. c.* pv. citri. Exponential phase cultures of strain XW47 transformed with plasmid pLEP320 were incubated with various concentrations of mitomycin C (A), MMS (B), ethidium bromide (C), or nalidixic acid (D), or were exposed to the indicated doses of UV radiation (E), and luciferase activity was assayed at the indicated times. Data are means ± SEM of five independent experiments.

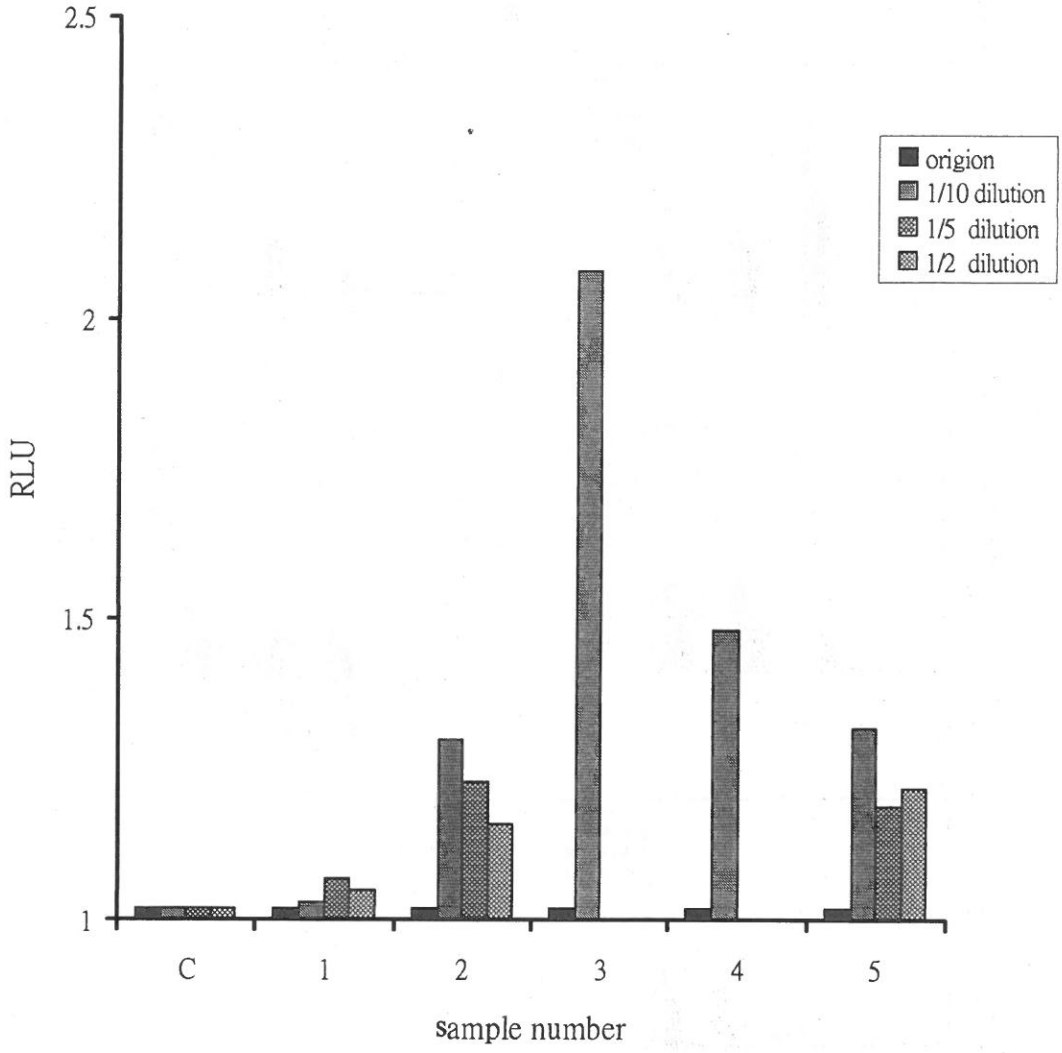


Fig. 5. Induction of *recA-luxAB* fusion in waste water samples collected from Tai-Pei Hsien area. Luciferase activity was assayed 60 min after adding various dilution of an estuarine water sample.

environmental contamination, waste water samples from industrial area were tested. Water was obtained from various factory at Hsin-Chuang and Tai-Sun site in Taipei Hsien, Taiwan. We examined the presence of DNA-damaging agents by determining

recA gene expression with the use of a *recA-luxAB* fusion construct. pREP426 containing 287 bp of the 5′ untranslated region and 139 bp of the 5′ coding region of *X.c.* pv. citri *recA* was introduced into strain XW47 by electroporation, and the effects of water samples on luciferase activity were measured. Maximal induction of the *recA-luxAB* fusion construct was apparent when XW47 cells were exposed to the DNA-damaging agents at a culture optical density at 600 nm of 1.0 to 1.2 (Yang and Yang, 1999). Bioluminescence of cells containing this construct indicated that certain mutagen was present in the waste water (Fig. 5). The maximal effects were apparent ~1 h after initial exposure to inducers. Induction of *lexA-luxAB* was also examined and the responses were similar to that observed in *recA-luxAB*. Thus, both reporter strains were useful for determination of DNA-damaging agents in the waste water.

ACKNOWLEDGMENTS

We thank Shu-Fen Weng (Chung Hsing University, Taichung) for providing the plasmid pMY3. This research was supported in part by the grant of 1998 from the the SVD section, Fu Jen Catholic University.

REFERENCES

(1) Hill, P. J., Rees, C. E. D., Winson, M. K. and Stewart, G. S. A. B. (1993) The application of *lux* genes. Biotechnol. Appl. Biochem. 17, 3-14.

(2) Justus, T. and Thomas, S. M. (1998) Construction of a *umuC′-lusAB* plasmid for the detection of mutagenic DNA repair via luminescence. Mut. Res. 398, 131-141.

(3) Manen, D., Pougeon, M., Damay, P. and Geiselmann, J. (1997) A sensitive reporter gene system using bacterial luciferase based on a series of plasmid cloning vectors compatible with derivatives of pBR322. Gene 186, 197-200.

(4) Meighen, E. A. (1991) Molecular biology of bacterial bioluminescence. Microbiol. Rev. 55, 123-142.

(5) Meighen, E. M. and Dunlap, P. V. (1993) Physiological, biochemical and genetic control of bacterial bioluminescence. Adv. Microbiol. Physiol. 34, 1-67.

(6) Miller, R. V. and Kokjohn, T. A. (1990) General microbiology of *recA*: environmental and evolutionary significance. Annu. Rev. Microbiol. 44, 365-394.

(7) Min, J., Kim, E. J., LaRossa, R. A. and Gu, M. B. (1999) Distinct responses of *recA-lux* CDABE Escherichia coli strain to direct and indirect DNA damaging agents. Mut. Res. 442, 61-68.

(8) Oda, Y., Nakamure, S, Oki, I., Kato, T. and Shinagawa, H. (1985) Evaluation of the new system (*umu* test) for the detection of environmental mutagens and carcinogens. Mut. Res. 147, 219-229.

(9) Quillardet, P., Huisman, O., D'Ari, R. and Hofnung, M. (1982) SOS chromotest, a direct assay of induction of an SOS function in *Escherichia coli* K12 to measure genotoxicity. Proc. Natl. Acad. Sci. USA 79, 5971-5975.

(10) Ptitsyn, L. R. (1996). Bioluminescent analysis of SOS response in *Escherichia coli*. Genetika 32, 354-358.

(11) Reifferscheid, G., Heil, J., Oda, Y. and Zahn, R. K. (1991) A microplate version of the SOS/*umu*-test for rapid detection of genotoxins anf genotoxic potentials of environmental samples. Mut. Res. 253, 215-222.

(12) Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(13) Smith, C. L. (1985) Response of recA-dependent operons to different DNA damage signals. J. Biol. Chem. 260, 10069-10074.

(14) Stewart, G. A. S. B. and Williams, P. (1992) *lux* genes and the applications of bacterial bioluminescence. J. Gen. Microbiol. 138, 1289-1300.

(15) Van Dyk, T. K., Majarian, W. R., Konstantinov, K. B., Young, R. M., Dhurjati, P. S. and LaRossa, R. A. (1994) Rapid and sensitive pollutant detection by induction of heat shock gene-bioluminescence gene fusion. Appl. Environ. Microbiol. 60, 1414-1420.

(16) Van Dyk, T. K., Smulski, D. R., Reed, T. R., Belkin, S., Vollmer, A. C. and LaRossa, R. A. (1995) Responses to toxicants of an *Escherichi coli* strain carrying a *uspA'-lux* genetic fusion and an *E. coli* strain carrying a

grpE′-lux fusion are similar. Appl. Environ. Microbiol. 61, 4124-4127.

(17) Vollmer, A. C., Belkin, S., Smulski, D. R., Van Dyk, T. K. and LaRossa, R. A. (1997) Detection of DNA damage by use of *Escherichia coli* carrying *recA′-lux*, *uvrA′-lux*, or *alkA′-lux* reporter plasmids. Appl. Environ. Micro. 63, 2566-2571.

(18) Walker, G. C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol. Rev. 48, 60-93.

(19) Weng, S. F., Shieh, M. Y., Lai, F. Y., Shao, Y. Y., Lin, J. W. and Tseng, Y. H. (1996) Construction of a broad-host-range promoter-probing vector and cloning of promoter fragments of *Xanthomonas campestris*. Biochem. Biophys. Res. Commun. 228, 386-390.

(20) Willardson, B. M., Wilkins, J. F., Rand, T. A., Schupp, J. M., Hill, K. K., Keim, P. and Jackson, P. J. (1998) Development and testing of a bacterial biosensor for toluene-based environmental contaminants. Appl. Environ. Micro. 64, 1006-1012.

(21) Yang, M. -K., Su, W. C. and Kuo, T. T. (1991). Highly efficient transfection of *Xanthomonas campestris* by electroporation. Bot. Bull. Acad. Sinica (Taipei) 32, 197-203.

received September 4, 1999
accepted October 8, 1999

Detection of DNA-damaging Agents in the Environment
by Using Bacteria Carrying a *recA-luxAB* or a *lexA-luxAB* Fusion

利用 *recA-luxAB* 與 *lexA-luxAB* 質體檢測環境中 DNA 突變劑的存在

楊美桂　吳姵儀　楊添鈞
輔仁大學生物系

摘　要

將柑橘潰瘍病菌 *recA* 與 *lexA* 基因之上游序列，具有可被誘導之啓動子區域，接在冷光基因 *luxAB* 之前，所形成之轉錄質體，可用於測定 DNA 突變劑的存在。取台北縣五股工業區之工廠排放水，與含此質體之細菌一同培養，可測得明顯的冷光反應。表示廢水中含有 DNA 突變劑，可誘導 *recA* 與 *lexA* 啓動子，促使 luxAB 基因的表現，產生 luciferase 而發出冷光。故知構築出之含 *luxAB* 之質體，可做爲生物感測器，以測定環境中造成 DNA 損害的因子。

關鍵詞： 生物檢測器，DNA 突變劑，工業廢水

Torsional Vibration Suppression By Excitation Control of Synchronous Motor Driving Mass-Spring Load Systems

Yuang-Shung Lee

Department of Electronic Engineering

Fu Jen Catholic University

Taipei, Taiwan, 242, R.O.C.

Abstract

This paper presents a torsional vibration suppression problem of a large-scale synchronous motor driving system under starting transient and steady-state disturbing conditions. The transient torsional vibration is suppressed by inserting a passive circuit element in the field circuit of a salient pole synchronous motor during the starting period. The circuit parameter in the starting field circuit control schemes is determined by the proposed optimization algorithm with the unified quadratic vibration performance index. The dynamic torsional vibrations are repressed by a proposed auxiliary excitation controller, is designed by optimal and suboptimal control theory. The LQR designed method does not need the prespecified weighting matrices. The eigenvalues of torsional modes are shifted to preassignment vertical strip, and does not destroy the other mode shapes of the system. For demostrating the effectiveness of torsional vibration suppression, eigenvalue analysis and nonlinear computer simulation results are used to show that the proposed starting control scheme and excitation controller give significant improvement in the torsional oscillation performance of the synchronous motor driving system.

Key Words : torsional vibration, starting control, strip eigenvalue

assignment, optimal and suboptimal control, stability

INTRODUCTION

Large synchronous motor has found application in numerous industrial activities, such as, multicompressor, pumped storage plants, drill presses, cement industry drivers, and synchronous compensators. The driving system composed of an inertia load through mass spring mechanical system brings a certain set of torsional vibrations under transient and steady state disturbance [1-5]. Torsional vibration is rarely a problem for small capacity utility motor, but may be of serious predicament to large-scale synchronous motor driving system during disturbance, if it is not provided with a well-design starting control scheme and a suppressed control method for the motor and shaft system [3,4]. The salient pole synchronous motor usually have a pulsating torque components with a frequency spectrum ranging from dc component to twice source-frequency during the starting transient period. If the frequency of one or more of these pulsating torque components becomes complementary with that of the torsional mode of the mass spring mechanical system, then they become mutually excited and are affected by the interaction with the electromagnetic torque of the synchronous motor. It will cause an extreme risk to the mechanical system and the production process, such as, tremendous shaft torque and shaft failure [4-8], the economic impact due to production loss [5], or increased risk to the system [6].

In practice, the above mentioned damaging oscillation torque usually occurs in the initial starting period and/or subject a disturbance. Therefore, the starting process and control of the motor driving system play a significant role for oscillation dynamics. Generally, a synchronous motor is started from a direct line voltage starter under short field circuit. As the starting current reaches a range of four to ten times of the full load motor current, then the torque surge can be in excess of 200% of the full load torque. The reduced voltage starting method is the most frequently adopted method to suppress the starting inrush in industrial motor application [10.11]. However, torsional oscillation components are unavoidably generated and interacted with the starting

transient or stead-state disturbance of the synchronous machine [8].

A synchronous motor, is excited from the direct current which is applied to it's field windings. The field circuit and excitation system supply a significant consideration that needs to be adopted either during the design or the operation phase of the synchronous motor for industrial applications [4,12]. A series connected field discharge resistor in the field circuit of a synchronous motor, is usually recommended by the industry for both field insulation protection, and increased starting torque during the starting period.

In the present study, a starting field circuit control scheme with a series connecting impedance is applied to compress the transient torsional vibrations. An unified optimization approach is proposed to determine the passive circuit parameters of the starting field circuit control schemes, according to the optimization of the novel quadratic vibration performance index. An auxiliary excitation controller with partial output feedback is used to damp the dynamic torsional vibration under the disturbance during the steady state conditions. The output feedback gains are obtained from suboptimal control approach of the strip eigenvalue assignment. For demonstrating the effectiveness of torsional vibration suppression, nonlinear computer simulation and eigenvalue analysis are illustrated to validate that the proposed starting control scheme and the excitation controller give significant improvement in the torsional vibration performance of the large-scale synchronous motor driving three-mass-spring load system.

SYSTEM MODELING

1. Nonlinear model

Figure 1 shows the schematic configuration of the three-mass-spring load driving system for torsional dynamic analysis. The system consists of a large-scale synchronous motor with a field circuit, and driving an inertia load through the coupling device with two elastic shafts. The three-phase, symmetrical stator winding of the synchronous motor whose arbitrary reference-frame d-q voltage equations can be written as:

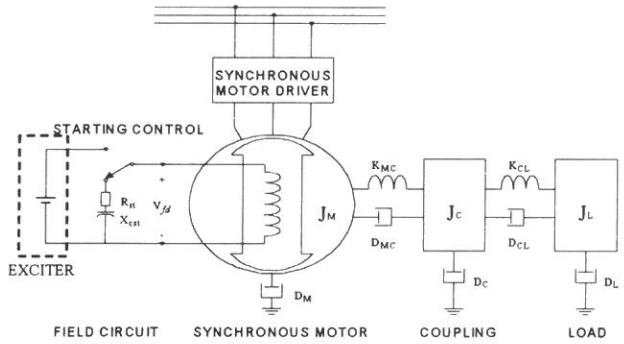


Fig. 1. Studied system.

$$\begin{bmatrix} E_{fd} \\ V_d \\ V_{kd} \\ V_q \\ V_{kq} \end{bmatrix} = \begin{bmatrix} R_{fd} + \frac{p}{\omega_b}X_{fd} & \frac{p}{\omega_b}X_{md} & \frac{p}{\omega_b}X_{md} & 0 & 0 \\ \frac{p}{\omega_b}X_{md} & R_s + \frac{p}{\omega_b}X_d & \frac{p}{\omega_b}X_{md} & \frac{\omega}{\omega_b}X_q & \frac{p}{\omega_b}X_{mq} \\ \frac{p}{\omega_b}X_{md} & \frac{p}{\omega_b}X_{md} & R_{kd} + \frac{p}{\omega_b}X_{kd} & 0 & 0 \\ \frac{\omega}{\omega_b}X_{md} & \frac{\omega}{\omega_b}X_d & \frac{\omega}{\omega_b}X_{md} & R_s + \frac{p}{\omega_b}X_q & \frac{p}{\omega_b}X_{mq} \\ 0 & 0 & 0 & \frac{p}{\omega_b}X_{mq} & R_{kq} + \frac{p}{\omega_b}X_{kq} \end{bmatrix} \begin{bmatrix} i_{fd} \\ i_d \\ i_{kd} \\ i_q \\ i_{kq} \end{bmatrix} \quad (1)$$

where $E_{fd} = R_{fd} \cdot V_{fd}/X_{md}$ is the field voltage, the applied d-q voltage of the damping winding is zero, i.e. $V_{kd} = V_{kq} = 0$ [13].

The external starting field voltage, $V_{fdst} = R_{st}i_{fd} + \omega_b X_{cst}\int i_{fd}dt$, where R_{st} and X_{cst} are the starting field discharge resistance and the starting field charge/discharge capacitance, respectively.

The expression for the electromagnetic torque in terms of the reference-frame states is:

$$T_e = (X_{md}i_{fd} + X_d i_d + X_{md}i_{kd})i_q - (X_q i_q + X_{mq}i_{kq})i_d \quad (2)$$

In order to analyze the dynamic torsional vibrations of the three-mass-spring motor driving systems, it requires to solve the simultaneous equations of the synchronous

motor and the mass-spring mechanical system, for several control schemes. The mathematical equation of the three-mass-spring mechanical system including the shaft torsional dynamics in terms of rotating angle θ, and angular velocity ω of each rotating mass can be expressed as:

$$\begin{bmatrix} 0 \\ T_e \\ 0 \\ 0 \\ 0 \\ -T_L \end{bmatrix} = \begin{bmatrix} p & -1 & 0 & 0 & 0 & 0 \\ K_{MC} & D_M + D_{MC} + J_M p & -K_{MC} & -D_{MC} & 0 & 0 \\ 0 & 0 & p & -1 & 0 & 0 \\ -K_{MC} & -D_{MC} & K_{MC} + K_{CL} & D_C + D_{MC} + D_{CL} + J_C p & -K_{CL} & -D_{CL} \\ 0 & 0 & 0 & 0 & p & -1 \\ 0 & 0 & -K_{CL} & -D_{CL} & K_{CL} & D_C + D_{CL} + J_L p \end{bmatrix} \begin{bmatrix} \theta_r \\ \omega_r \\ \theta_C \\ \omega_C \\ \theta_L \\ \omega_L \end{bmatrix} \tag{3}$$

where the subscripts of the states, r, c, and L denote the rotor, coupling, and load rotational mass, respectively. p denotes the time differential operator, and T_L is the load torque of the synchronous motor [14]. The shaft torques between the three mass are defined by:

$$T_{MC} = K_{MC}(\theta_M - \theta_C) \tag{4}$$

$$T_{CL} = K_{CL}(\theta_C - \theta_L) \tag{5}$$

where θ_M is the mechanical angle of shaft rotation of the synchronous motor.

By combining Eqs.(1)~(3) and taking the proposed control scheme of the field circuit of the synchronous motor into the system, the complete system model is composed of eleven nonlinear equations including the synchronous motor dynamics and the mass-spring dynamics. The torsional dynamics of high inertia load system interacts with the electromagnetic torque of synchronous motor during the disturbance under the different control schemes.

2. Linearized model

In order to regulate the excitation voltage of the synchronous motor under the system operating necessary. The excitation system for the synchronous motor is equipped as the static type exciter, and the s-domain block diagram is illustrated in Fig.2. Combining synchronous motor, exciter, and mass-spring mechanical system,

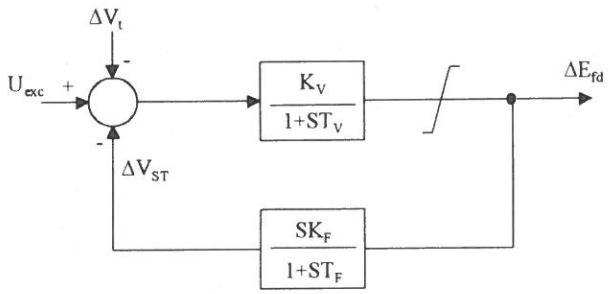


Fig. 2. Excitation system.

we can get a set of 13th-order nonlinear differential equations. At the nominal operating point, all the nonlinear differential equations can be linearized to obtained a set of linear differential state equations by the following compact from:

$$E\dot{X}(t) = FX(t) + GU(t)$$
$$\dot{X}(t) = E^{-1}FX(t) + E^{-1}GU(t) = AX(t) + BU(t) \tag{6}$$

where $A = E^{-1}F$, and $B = E^{-1}G$ are the constant system matrix, $X(t) = [\Delta i_{fd}, \Delta i_{ds}, \Delta i_{kd}, \Delta i_{qs}, \Delta i_{kq}, \Delta\theta_r, \Delta\omega_r, \Delta\theta_c, \Delta\omega_c, \Delta\theta_L, \Delta\omega_L, \Delta E_{fd}, \Delta V_{sT}]^T$ is the system state vector, $U(t) = [U_{exc}]$ is the supplementary control signal from the auxiliary excitation controller.

Torsional oscillation stability of the system can be determined by examining the eigenvalues obtained from Equ.(6). All of the eigenvalues must be located at the left-half plane in the s-domain and distant from the imaginary axis for better system damping effectiveness. The system data including synchronous motor, exciter, and mass-spring mechanical system are given in Appendix. The complete eigenvalues of the system without the auxiliary excitation controller are tabulated in the second column of Table 1. It is observed that there are two critical torsional modes, located at $-0.096 \pm j297.71$ rad/sec (47.38Hz) and $-1.101 \pm j698.73$ rad/sec (111.2Hz), denoted the torsional modes. The damping of those critical torsional modes are not sufficiency, and the damping ratio of the dominant mode is 0.003. In addition, the damping for the torsional modes to be increased, and the requirement for appropriate stabilization controller to damp out of the torsional oscillations of the motor driving system is evident.

Table 1. System eigenvalue.

	Without excitation controller	With optimal excitation controller	With suboptimal excitation controller
Synchronous motor and exciter	−54.88 −4.33±j12.23 −20.78±j375.09 −1.441 −1.742 −64.346	−54.88 −4.33±j12.23 −20.78±j375.09 −1.515 −1.742 −64.346	−60.41±j10.57 −5.55±j12.47 −17.48±j374.22 −1.574±j0.658
Mass-spring Mechanical System	−38.26 −0.096±j297.71 −1.101±j698.73	−38.26 −1.638±j297.71 −1.603±j698.73	−37.494 −1.85±j298.49 −1.89±j698.73

STARTING CIRCUIT DESIGN

The Salient-pole synchronous motor, which is used for most industrial applications, has a cage type damping winding on the surface of salient-pole, and there is a non-uniform magnetic reluctance. Most synchronous motors are started from rest and accelerated to near synchronous speed as an induction machine [2]. The starting production torque is to be constructed by the following three torque components. First, damping winding torque produced from the cage type damping winding as an induction machine. Second, reluctance torque produced from the variation between d-q axis that causes a pulsation of armature current and motor torque at twice-slip frequency [6]. Third, the transient torque produced from starting field circuit control schemes, usually shorts field circuit. During the starting transient disturbance, the torsional mode is excited, the transient torque component of a field circuit of the synchronous motor has a significant effect. To ensure that the starting transient disturbance does not excite the shaft torsional oscillation to approach the critical designed limit for torsional stress, it may produce a premature component failure. Hence, the starting techniques and proper switching coordination are necessary to reduce the corresponding torsional mode torque and mutual interactions. Thus a starting control scheme with a series conneting impedance in the field circuit of the synchronous motor is proposed in the present study, to reduce the starting transient

torsional vibrations. The series impedance is composed of a field discharge resister and/or a charge/discharge capacitor.

In order to optimize the field discharge resistance and charge/discharge capacitance, a quadratic vibration performance index is proposed to determine the optimum the starting field discharge resistance and the charge/discharge capacitance. By defined the motor-speed vibration error vector $\widetilde{X} = [\triangle\omega_r,\ td\triangle\omega_r]^T$ where $\triangle\omega_r$ is the motor-speed deviation and $td\triangle\omega_r$ is the time-weighted derivative of the motor-speed deviation, respecitively. Thus the quadratic vibration performance index can be expressed as [14] :

$$J_W = \frac{1}{2}\sum_{K=0}^{N}\widetilde{X}^T(t)Q\widetilde{X}(t) \tag{7}$$

where the superscript T denotes the transpose of a matrix, whereas, Q is the positive and diagonal weighting matrix of the vibration variable deviations, $t = K\triangle T$ where $\triangle T$ is the sampling time, N is the number of measured data points. The parameter optimization algorithm minimizes the vibration performance index such that the optimal parameter of the starting field control circuit can be calculated with a minimum vibration error and settling time response.

EXCITATION CONTROLLER DESIGN

To determine the gains of the controller, the linearized differential equations of the system, including synchronous motor, exciter, and the three-mass-spring load system, as mentioned in Eq.(6) are rewritten in state equations.

$$\dot{X}(t) = AX(t) + BU(t),\ X(t_0) = X_0 \tag{8}$$

$$Y(t) = CX(t) \tag{9}$$

where $X(t)$ is the $n \times 1$ state vector, $U(t) = [U_{exc}]$ is $m \times 1$ the supplementary control signal from the excitation controller, and $Y(t) = [\Delta\theta_r,\ \Delta\omega_r,\ \Delta\theta_c,\ \Delta\omega_c,\ \Delta\theta_L,\ \Delta\omega_L]^T$ is $p \times 1$ the output signal.

1. Optimal strip eigenvalue assignment

In the design of a conventionally optimal control system, it is desired to get $U(t)$

such that the performance index J is minimum, where

$$J = \frac{1}{2}\int_0^\infty [X^T(t)\widetilde{Q}X(t) + U^T(t)RU(t)]dt \tag{10}$$

The weighting matrices $\widetilde{Q}$ and R are $n \times n$ nonnegative and $m \times m$ positive definite symmetric matrices, respectively. The optimal control vector is $U^0(t) = -F^0X(t)$, where the feedback gain is $F^0 = R^{-1}B^TK$ with K being a symmetric positive definite solution of the matrix form Riccati equation in closed form:

$$A^TK + KA\text{-}KBR^{-1}B^TK + \widetilde{Q} = 0_n \tag{11}$$

and the closed loop eigenvalues, denoted by $\Lambda(A - BF^o) = [\lambda_1, \cdots\cdots \lambda_m, \lambda_{m+1}, \cdots \lambda_n]$, will lie in the open left-half plane of the complex s-plane as the desired damping for the system.

In conventionally *LQR* problems, the optimal gain is designed by redundantly selecting weighting matrixes in accordance with practical intention for simplicity, the matrices $\widetilde{Q}$ and R are usually chosen as diagonal matrices. To improve the system performance, the weighting matrix R is set to be an identity matric for equal weighted of the m control inputs, and the weighting matrix $\widetilde{Q}$ must be given previously, such that the controlled eigenvalues λ_1 to λ_m will be selected and shifted to a desired location.

For a large-scale system, the weighting matrices $\widetilde{Q}$ and R are hard to determine previously. A redundantly trial and error are the most usually approach to determine those weighting matrices. To solve this difficulty, a unified countermeasure will be proposed as follows, the closed-loop eigenvalues are shifted to a prespecified vertical strip without the need of weighting matrices.

Let $\xi \geqq 0$ be a real number, and represent the prescribed degree of relative stability. The unstable system eigenvalues of $\bar{A} = A + \xi I_n$ will be shifted to their coresponding locations with respect to the $-\xi$ vertical line, then the overall system eigenvalues of the closed loop system, $A_c = A\text{-}BR^{-1}B^T\overline{K}$, have lain on the left side of the $-\xi$ vertical line of the complex s-plane, where the $\overline{K}$ is the solution of the following closed form Riccati equation with $\overline{Q} = 0_n$ [15]:

$$\bar{A}^T\overline{K} + \overline{K}\bar{A} - \overline{K}BR^{-1}B^T\overline{K} + \overline{Q} = 0_n \tag{12}$$

Assume, ξ_1 and ξ_2 are two positive real numbers to denote the vertical strip of $[-\xi_2,$

$-\xi_1]$ on the negative real axis of the complex s-plane. Given $\tilde{A} = A + \xi_1 I_n$, the control law is modified to be $U(t) = -\rho\tilde{F}X(t) = -\rho R^{-1}B_T\tilde{K}$, where the matrix $\tilde{K}$ is the solution of the modified Riccati equation:

$$\tilde{A}^T\tilde{K} + \tilde{K}\tilde{A} - \tilde{K}BR^{-1}B^T\tilde{K} = 0_n \tag{13}$$

The gain ρ is slelected by [16]

$$\rho = 0.5(1 + \frac{\xi_2 - \xi_1}{tr(\tilde{A}^+)}) = 0.5 + \frac{\xi_2 - \xi_1}{tr(B\tilde{F}^+)} \tag{14}$$

where $tr(\tilde{A}^+) = -\sum_{i=1}^{n^+}\lambda_i^+ = \frac{1}{2}tr(B\tilde{F})$ and $\lambda_i^+(i = 1,\ldots,n^+)$ are the unstable eigenvalues of $\tilde{A}$. The optimal closed-loop system becomes $\dot{X}(t) = (A - \rho B\tilde{F})X(t)$, $\Lambda(A - \rho B\tilde{F})$ denoted a set of eigenvalues which located inside the vertical strip of $[-\xi_2, -\xi_1]$. In mentioned Eq.(13) for equal weighting of the m control inputs, R can be an unity matrix. The proposed optimal strip eigenvalue assignment can be used to design the damping controller without the need of weighting matrix $\overline{Q}$.

2. Suboptimal strip eigenvalue assignment

Since the control signal from the optimal controller is a linear time invariant combination of all states. It is impractical in real system. Some system states could not be measured. Consequently, we want a suboptimal controller with measurable output signals

$$U^S(t) = -P^SY(t) = -P^SCX(t) = -F^SX(t) \tag{15}$$

It is noted that $U^S(t)$ is also a linear time-invariant combination of the system states and near a optimal controller.

If $F^S = \tilde{F}^o = \rho\tilde{F}$, the near optimal controller is an optimal one. However, it is usually impossible in practical system. Therefore, we define the state error vector

$$e(t) = X^S(t) - X^0(t),\ e(t_0) = 0 \tag{16}$$

where $X^S(t)$ is the suboptimal state vector and $X^O(t)$ is the optimal state vector. Differentiating both sides of Eq.(16), and substituting Eq.(8) into Eq.(16), we get

$$\dot{e}(t) = (A - BF^s)e(t) + B(F^s - \tilde{F}^0)X^0(t) \tag{17}$$

Let the excitation error vector be

$$h(t) = (F^s - \tilde{F}^0)X^0(t) \tag{18}$$

then

$$\dot{e}(t) = (A - BF^s)e(t) + Bh(t),\ e(t_0) = 0 \tag{19}$$

To minimize the effect of $h(t)$ in the time span, a quadratic excitation performance index $\tilde{J}$ is define.

$$\tilde{J} = \frac{1}{2}\int_0^\infty h^T(t)\tilde{R}h(t)dt \tag{20}$$

Minimizing the excitation performance index obtains the output feedback suboptimal controller in Eq.(15) [11]. Where P^S is the suboptimal output feedback gain and

$$P^S = \tilde{F}^0 LC^T(CLC^T)^{-1} \tag{22}$$

L is the solution of the matrix equation

$$(A - B\tilde{F}^0)L + L(A - B\tilde{F}^0)^T = - I \tag{22}$$

The closed loop system with the suboptimal controller is

$$\dot{X}(t) = (A - BF^S)X(t),\ X(t_0) = X_0 \tag{23}$$

The suboptimal controller only guarantees a minimum excitation performance index, but does not ensure a stable system. It is necessary to check that all eigenvalues of $(A - BF^S)$ have negative real part, and the damping of torsional modes are adequate.

ANALYSIS AND SIMULATION

To evaluate the transient and dynamic torsional performance of the synchronous motor driving system with the proposed control scheme, a nonlinear computer simulation is performed under different starting field control schemes. The eigenvalue analysis is utilized to verify the amplitude and the location of the studied system are tabulated in Table 1. An inspection of the system eigenvalues shows that there are two critical torsional modes, located at 47.38Hz (297.71 rad/sec) and 111.2Hz (698.73 rad/sec). Besides, the electromagnetic mode and the slip frequency mode of the synchronous motor are also located at 1.95Hz (12.23 rad/sec) and 59.69Hz (375.09

rad/sec), respectively. The torsional dynamic of the studied system might be directly excited by the starting transient disturbance and interacted with the electromagnetic torque of the synchronus motor.

The Runge-Kutta fourth order integration routines are used to obtain the solutions of the nonlinear different equations, Eqs.(1)~(3), represent the synchronus motor driving inertia load system. Fig.3 (a) shows the transient torsional dynamic during the starting period under a short field circuit starting control scheme. The starting inrush electromagnetic torque is affected by the dc component and the line source frequency component. Whereas, the oscillation component pertaining to acceleration torque is excited by the starting transient disturbance which is located close to 47.38 Hz. Beside, the two starting inrush shaft torques are also excited by this disturbance, and pulsated at a torsional mode of 47.38Hz and 111.2Hz. The drastic peak of inrush torque on the shaft by using short field circuit directed starting control scheme produce an extreme risk of premature shaft failure due to excessive dynamic torsional stress. As a result, some additional countermeasures in the starting techniques must be employed to repress the shaft torsional vibrations under system disturbance.

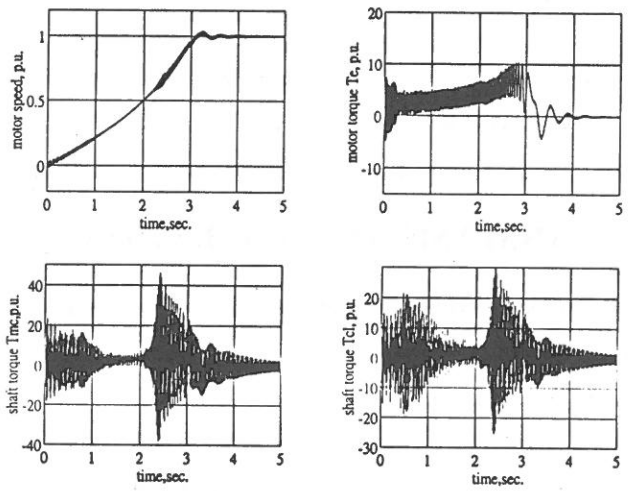


Fig. 3a. with directed short field circuit starting control scheme.

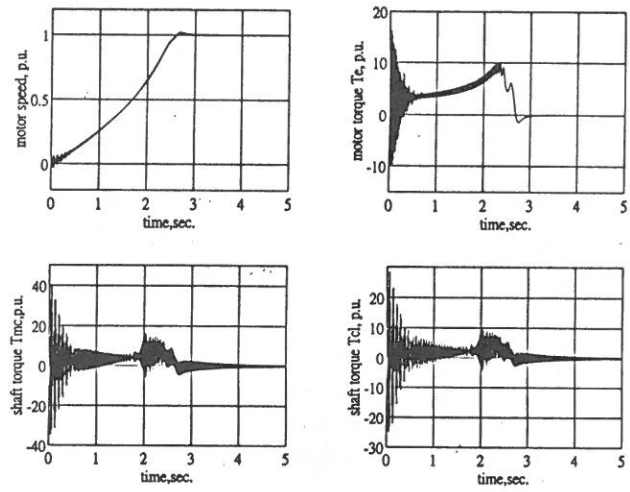


Fig. 3b. with optimum field capacitor starting control scheme

Fig. 3. Starting transient responses of synchronous motor driving system.

A simple starting control scheme with a series connecting capacitance (i.e, shorted the resistance, R_{st}) in the field circuit is proposed to compensate the shaft torsional vibrations during the starting period. The determination of starting circuit parameters for the field circuit of the synchronous motor is based on the applications of the parameter optimization algorithm as mentioned in Eq.(7). A numerical procedure is proposed for determining the starting circuit parameter, a positive weighting matrix Q = Diag [0.00001,0.00001] is selected. For each of the given starting circuit parameters and the system dynamic states, the quadratic cost function represents the vibration error, as computed in Eq.(7). The best parameter is the one that corresponds to the smallest of the vibration performance index. Fig.4 shows the vibration performance index versus different series starting field charge/discharge capacitance under different constant load torque. The minimum cost is located at zero constant load torque situation. Wherein, the optimum starting charge/discharge capacitance, X_{cst}, is 0.94482 p.u. The optimum charge/discharge capacitance is independent of the constant load torque. Since, the capacitance, which minimized the equivalent armature impedance, and resonated close to the torsional mode with respect to the field inductance for the shorted-circuited armature [12]. Reducing the

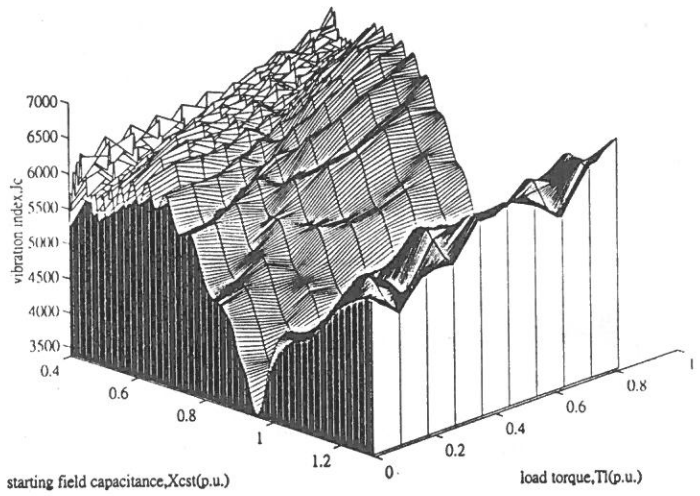


Fig. 4. Vibration performance index versus starting capacitance under different constant load torque.

synchronous impedance by the resonant field circuit, the synchronous machine could compensate the shaft torsional vibrations, which is excited by the starting transient disturbance.

With a well-designed optimum capacitance of the starting control schemes, one can not only improve the shaft torsional vibrations but also the acceleration of performance of the synchronous motor. Fig.3 (b) shows the transient responses of speed and torque for the system with an optimum field capacitance in the proposed starting control schemes. The essential features can be summarized as follows:

(a). The vibration of speed and electromagnetic torque of the synchronous motor can be compressed in the proposed starting control schemes with an optimum field charge/discharge capacitance. The minimum cost of the vibration performance for the synchronous motor with the optimum capacitance is almost less than that with the optimum resistance under the optimizating procedure [14].

(b). The shaft torsional vibrations are clearly improved for the choice of the proposed starting control schemes with the optimum series connecting capacitance. The peak amplitude of the transient shaft torque vibration was reduced to 40% of that of short field circuit starting control schemes.

(c). The speed accelerating performance and the maximum overshoot of the

proposed starting control schemes are better than the normal direct short field circuit starting control schemes.

Eigenvalue analysis and nonlinear computer simulation are used to ensure that the system damping is improved when the system with optimal and suboptimal excitation controller. To improve the torsional modes damping of the studied system, we select $[-\zeta_2, -\zeta_1] = [-2.0, -1.5]$ to design the excitation controller. There are two control schemes:

Case A: Optimal excitation controller is designed by optimal strip eigenvalue assgnment method.

$$U^{O}_{exc}(t) = -0.067\Delta i_{fd} - 0.076\Delta i_{ds} - 0.067\Delta i_{kd} - 0.001\Delta i_{qs} - 0.001\Delta i_{kq} - 3.38\Delta\theta_r - 0.041\Delta\omega_r + 29.71\Delta\theta_c + 0.37\Delta\omega_c - 26.34\Delta\theta_L - 0.334\Delta\omega_L + 0.373\Delta E_{fd} - 0.004\Delta V_{ST}$$

Case B: Suboptimal excitation controller is designed by suboptimal strip eigenvalue assignment method, that mentioned in above section.

$$U^{S}_{exc}(t) = -4.507\Delta\theta r - 0.065\Delta\omega_r + 44.432\Delta\theta_c + 0.596\Delta\omega_c - 39.924\Delta\theta_L - 0.531\Delta\omega_L$$

The eigenvalue of the system with the proposed optimal excitation controller is tabulated in the third column of Table 1. It is observed that all the eigenvalues are shifted to the vertical strip of $[-2.0, -1.5]$. The two critical torsional modes shift to $-1.638 \pm j297.1$ and $-1.603 \pm j698.73$, and the damping ratio of the dominant mode is improved to be 0.005. It also notes that the optimal excitation controller almost does not destroy the mode shapes of the system. The eigenvalues with suboptimal excitation controller are given in fourth column of Table 1. The torsional damping effect of the output feedback scheme is better than that of state feedback scheme. The eigenvalues of torsional modes are also near the vertical strip. Whereas, the mode shapes of the system have some destroyed more than that of the state feedback scheme. The damping effects are more acceptable than the original.

Eigenvalue analysis usually applies to investigate the small signal stability of a system. To examine the damping effect of the proposed controller during the dynamic period , time domain simulations are taken based on the nonlinear differential equations

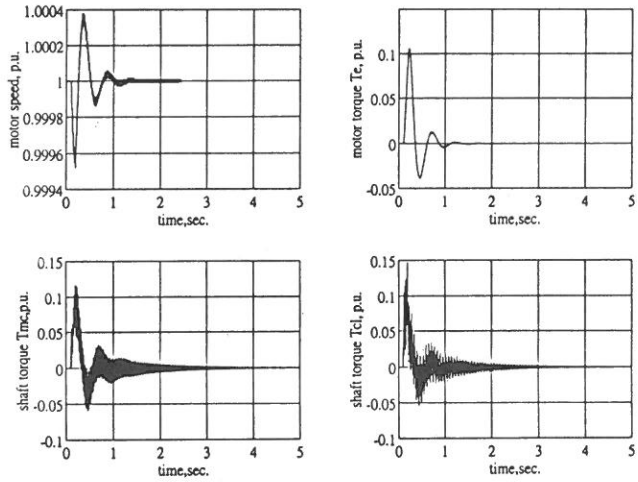


Fig. 5a. without auxiliary excitation controller.

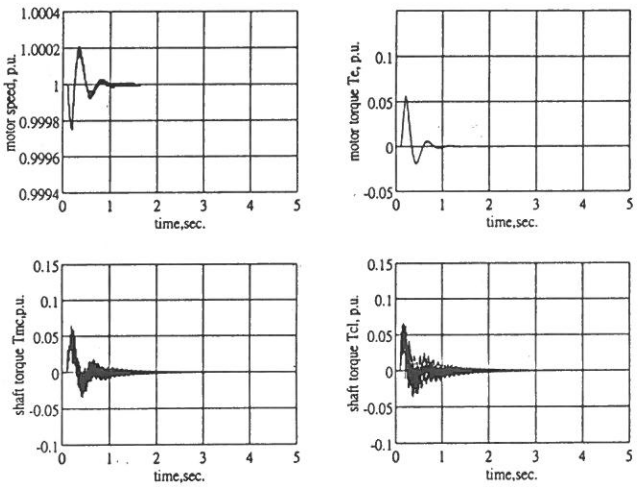


Fig. 5b. with suboptimal excitation controller.

Fig. 5. Steady state dynamic responses of synchronous motor driving system under constant load torque disturbance.

which describe the system behavior under disturbance conditions. The nonlinearity such as exciter ceiling voltage limit must be included. An 100msec, 10% load torque change is used to examine the effect on the synchronous motor shaft. The dynamic

responses are shown in Fig.5. The responses of the three-mass-spring motor driving system are obtained for following two cases: (a) without auxiliary excitation controller, and (b) with suboptimal excitation controller. The simulation results show that the proposed controller of the exciter can effectively damp torsional oscillations.

CONCLUSIONS

The study proposed a simple field excitation control scheme for a large-scale synchronous motor driving system during the transient and dynamic disturbance, to reduce the interaction between the torsional dynamic of mass-spring shaft and the electromagnetic torque of the syschronous motor.

A parameter optimization algorithm is proposed to determine the optimal field charge/discharge capacitance in the starting field circuit control schemes of the synchronous motor driving system. The nonlinear simulation results indicate that the proposed optimum starting field circuit control schemes for the synchronous motor driving system, can significantly reduce the amplitude of the shaft torsion vibrations. Furthermore, the transient interaction of the motor driving system excited by starting disturbance can also be minimized. The optimal field charge/discharge capacitance plays a more important role to limit the transient torsional vibrations and it is independent of the constant load torque of synchronous motor.

An output feedback excitation controller is designed to increase the dynamic stability of the three-mass-spring load synchronous motor driving system. The torsional modes can be shifted to a prespecified vertical strip without effecting the other mode shapes. The LQR design method is very simple and avoids the difficulty of choosing weighting matrices. Results of the analysis and digital simulation show that the proposed excitation controller can provide adequate torsional damping effect under disturbances.

ACKNOWLEDGMENTS

The auther would like to thank the National Council of R.O.C. (NSC 88-2213-E-030-010) for the financial support of the study, and the Fu-Jen Catholic University and SVD section for the support of computing facilities.

REFERENCE

(1) G. L. Godwin and E. F. Merrill, "Oscillatory torques during synchronous motor starting," IEEE Trans., IGA-6, no. 3, 1970, pp. 258-264, 1970.

(2) B. Mueller, T. Spinanger, and D. Wallstein, "Static variable frequency starting and drive system," Conf. Rec. IEEE/IAS, 1979 Annual Meeting, pp. 385-394, 1979.

(3) W. R. Finely and R. R. Burke, "Troubleshooting motor problems," IEEE Trans, IA-30, no. 5, pp. 1383-1397, 1994.

(4) R. M. Hamouda, M. A. Badr, and A. I. Alolah, "Effect of torsional dynamics on sailent pole synchronous motor-driven compressors," IEEE Trans., EC-11, no. 3, pp. 531-538, 1996.

(5) A. J. Willians, JR. and M. S. Griffith, "Evaluating the effect of motor starting on industrial and commercial power systems," IEEE Trans., IA-14, no. 4, pp. 292-305, 1978.

(6) C. B. Mayer, "Torsional vibration problems and analyses of cement industry drives," IEEE Trans., IA-17, no. 4, pp. 81-89, 1981.

(7) J. A. Bishop and C. B. Mayer, "A case for high-fidelity analysis of nonlinear electromechanical torsional dynamics, or 'Are you really sure it won't destroy itself? '", IEEE Trans., IA-15, no. 2, pp. 201-209, 1979.

(8) R. Yacamini and S. C. Chang, "Noise and vibration from induction machines fed for harmonic sources," IEEE Trans., EC-10, no. 2, pp. 286-292, 1995.

(9) A. A. Shaltout, "Analysis of torsional torque in starting of large squirrel cage induction motor," IEEE Trans., EC-9, no. 1, pp. 135-141, 1994.

(10) J. Nevelsteen and H. Aragon, "Starting of large motor-methods and economics," IEEE Trans., IA-25, no. 6, pp. 1012-1018, 1989.

(11) F. M. Bruce, III, R. J. Graefe, A. Lutg, and M. D. Panlener, "Reduced-voltage starting of squirrel-cage induction motors," IEEE Trans., IA-20, no. 1, pp. 46-55, 1984.

(12) F. Takase, M. Tominage, and Y. Ueda, "Harmonic compensation using a synchronous machine with resonant field circuits," IEEE Trans., EC-12, no.

2, pp. 143-150, 1997.

(13) P. C. Krause, *Analysis of electric machinery*, McGraw-Hill Book Co., 1987.

(14) Yuang-Shung Lee, "Transient torsional oscillations of a large synchronous motor driving system with field circuit control," Proceeding of the 19th Symposium on Electrical Power Engineering, pp. 433-438, 1998.

(15) L. S. Shieh, H. M. Dib and B. C. Mcinnis, "Linear quadratic regulators with eigenvalue placement," IEEE Trans. On Automatic Control, Vol. AC-31, NO. 3, pp. 241-243, 1986.

(16) Yuan-Chyuan Lee and Chi-Jui Wu, "Damping of power system oscillations with output feedback and strip eigenvalue assignment," IEEE Trans. on Power System, Vol. 10, No. 3, pp. 1620-1626, 1995.

APPENDIX

System data: Synchromous motor, exciter, and mass-spring mechanical system60Hz, 8-pole, 4KV, 6000hp

$R_s = 0.012$	$X_{kq} = 0.725$	$R_{kd} = 0.0302$	$X_{md} = 1.0$
$X_d = 1.17$	$X_{fd} = 1.297$	$X_q = 0.75$	$X_{kd} = 1.122$
$X_{mq} = 0.61$	$J_C = 0.012$	$J_m = 0.008$	$J_L = 0.014$
$R_{fd} = 0.0015$	$R_{kq} = 0.039$	$R_{kd} = 0.030$	$K_{MC} = 566.0$
$K_{CL} = 2952.0$	$D_M = D_C = D_L = 0$	$D_{MC} = 0.01$	$D_{CL} = 0.01$
$K_V = 2.0$	$T_V = 0.3$	$K_F = 0.125$	$T_F = 0.05$

received September 30, 1999
revised October 14, 1999
accepted November 20, 1999

Torsional Vibration Suppression By Excitation Control of Synchronous Motor Driving Mass-Spring Load Systems

同步電動機激磁控制以抑制負載一彈簧負載驅動系統的扭矩震動

李永勳
輔仁大學電子工程系

摘　要

本研究處理大型同步電動機起動暫態及穩態干擾下的扭矩震動抑制問題。暫態扭矩震動的抑制是在起動期間，在凸極式同步電動機場回路插入一被動電路元件，電路元件的常數是利用所提二項式震動特性指標的參數最佳化來求得。動態扭矩的抑制是利用所提的輔助激磁控制器來完成，控制器的參數是利用不需要權重矩陣的最佳和次佳帶狀特徵指定法來設計。扭矩模式及激磁模式特微值都被移到指定的帶狀區域而不破壞其他特微值模式。特微值分析和非線性計算機模擬證明所提起動控制架構及輔助激控制器都能很有效的抑制同步馬達驅動系統的扭矩震動特性。

關鍵詞： 扭矩震動、起動控制、帶狀特徵值指定法、最佳及次佳控制、穩定度

Random Position Orthography Halftoning

Shiang-Rung Ye and Brian K. Lien
Department of Computer Science and Information Engineering
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.

Abstract

Orthography halftoning is one of the old digital halftoning algorithms. It replaces each pixel in a continuos-tone image with an M by N font. The advantage of orthography algorithm is that when you need to scale up a continuous-tone image before applying halftoning algorithm to the image, orthography algorithm can do both things at the same time, which shortens the processing time considerably. But orthography halftoning algorithm does possess disadvantages. Because orthography algorithm uses only one font to reproduce one gray level, the textures, created from a uniform area, exhibit structure or fixed pattern. It is also common to see some dramatic texture in the halftoned image depending on selected fonts. In this paper, we propose an algorithm to solve these problems. The algorithm uses multiple fonts to reproduce one gray level. These multiple fonts are created at run-time for each pixel in the image by putting desired dots into the font randomly. Also a novel implementation data structure is provided to efficiently perform the algorithm.

Key Words: halftone, orthography, dither, error diffusion

INTRODUCTION

Digital halftoning is a process that creates a binary image from a continuous-tone image. The primary challenge is to create a binary image, which gives human's eyes an illusion of original continuous-tone image. There are many halftoning algorithms that

can deal with this problem. The well-known algorithms are random dither, ordered dither, error diffusion [1-2], constrained average halftoning [3] and orthography tone scale creation [4-5], etc. Some researches also reported that by incorporating human visual model into halftoning algorithms, the halftoned images can get improved visual effect too [6-12].

Orthography halftoning is one of the old digital halftoning algorithms. It replaces each pixel in a continuos-tone image with an M by N font. A font is an array consisting of black dots or white dots. The set of fonts is usually called book. Prior to applying orthography algorithm, we have to design a book for the algorithm. Some optimization has been done by Hamill [13], and the resulting fonts are similar to those shown in Fig.1 [14].

After applying orthography halftoning algorithm to images, the resulting images will be larger than original one. When we need to scale up a continuous-tone image before applying halftoning algorithm to the image especial for printing, orthography algorithm can do both things at the same time, which shortens the processing time considerably.

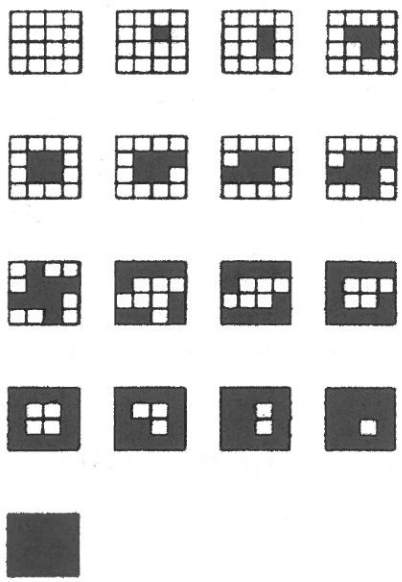


Fig. 1. 4×4 Fonts.

However orthography algorithm possess some disadvantages. Because orthography algorithm always uses the same font to reproduce the same gray level, it suffers from two problems. The first problem is fixed pattern. When we dither a uniform area, we will see a clear fixed patter. Human's eyes are sensitive to regular pattern, so we don't want that happening. The second problem of orthography algorithm is dramatic textures. Orthography algorithm may create some dramatic textures unexpectedly depending on selected font.

To solve these problems, we will use multiple fonts to reproduce one gray level. In orthography algorithm, fonts are previously defined in a book. In the proposed algorithm, fonts are not defined prior to processing but are created at run-time for each pixel in the image by putting desired dots into the font randomly.

FOURIER ANALYSIS

The characteristic of a halftoning algorithm can be determined by its ability to render a uniform gray area. This ability can be examined in the frequency by observing the power spectrum of the halftoning algorithm. The power spectrum of error diffusion shown in Fig.2 shows that this algorithm has a low frequency cutoff as denoted at 2 and there is a high peak denoted as 1. This type of dithering is called "blue noise" because the power spectrum of the halftoning algorithm is similar to that of blue color. Figure 3 is the estimated power spectrum of error diffusion algorithm created by averaging 1 0 periodograms of an image with 1 2 8 gray level . Another type of dithering

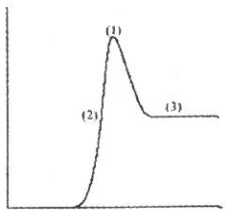


Fig. 2. The power spectrum of error diffusion.

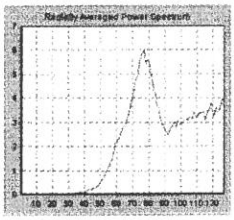


Fig. 3. The estimated power spectrum of error diffusion.

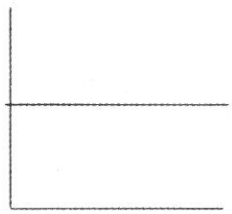


Fig. 4. The power spectrum of random.

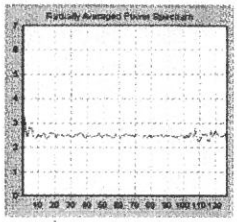


Fig. 5. The estimated power spectrum of random dither.

is random dither.

The power spectrum of random dither shown in Fig.4 shows that it add energy all over frequencies. This type of dithering is called white noise because the power spectrum is similar to that of white color. Figure 5 is the estimated power spectrum of random dither. The presence of energy at significant low frequencies is responsible for artifacts [1]. Random dither adds energy over the low frequencies of images, so its halftone images suffer from grainy appearance. By observing the power spectra of halftoned images, we can get an idea about how serious the artifact affects the halftoned images.

Halftoning processes can be classified into two classes; periodic and aperiodic. Periodic processes produce halftone images by thresholding images with a previously defined threshold array. The proposed algorithm doesn't dither images with threshold array, so it is an aperiodic process. Because the autocorrelation function of an aperiodic process won't be known in most case, we have to estimate the power spectrum. In this paper we estimate the power spectrum as the same way as described in reference [1]. In this method, Bartlett's periodograms are used to estimate the power spectrum [15]. A periodogram is the magnitude squared of the Fourier transform. It can be shown that the estimate power spectrum has an expectation equal to the actual power spectrum smoothed by convoluting the Fourier transform of a triangle function [16]. More information can be found in [1,16].

THE ALGORITHM

In Random Position Orthography halftoning (RPO), we create a font at run-time for each pixel according to its gray level. We divide a font into the same size of sub-blocks. The sub-blocks consist of at least 2 positions, which can be put a black dot or a white dot, but only one of them will be chosen to accommodate a dot. The purpose of division of a font is required in order to put dots uniformly in the font without lots of dots gathering together, while the purpose of randomly choosing position for a dot in a sub-block is to prevent dots from being put at same location periodically. In other word we define an extent in the font for a dot to be put randomly at any position in that extent instead an exact location in the font where a dot must be put. A font created in

this way will not be the same and dots distributed in this way will be more uniformly. Following is the proposed algorithm.

We first defined some parameters. Assume that the size of font is M by N and we want to put x white/black pixels in the font. Let Dots(M,N) denote the upper bound of dots that may be put in the font. If $M \times N$ is even, then $Dots(M,N) = M \times N/2$, else $Dots(M,N) = (M \times N - 1)/2$. Also let $S = \{x \in Integer \mid 0 < x < = Dots\ (M,N)\}$ and $P = \{x \in S \mid x \text{ divide } M \times N\}$ and $Q = S - P$

Step 1:

If $x < = Dots(M,N)$,

Set all dots in the font as black dots

Let $n = x$

Go to step 2.

If $x > Dots(M,N)$,

Set all dots in the font as white dots

Let $n = M \times N - x$

Go to step 2.

Step 2:

If $n \in P$,

Divide the font into n sub-blocks of the same size

Put one dot into each sub-block randomly.

Processing is completed

If $n \in Q$,

Find $z \in P$ and $z < n$ such that $\forall k \in P$, $k < n$, $k < > z$ and $n-z < n-k$

Divide the font into z sub-blocks of the same size

Put one dot into each sub-block randomly

Let $n = n - z$

Go to step 2

For example, If we want to generate a 4×4 font with 6 white dots, the following parameters can be derived:

$Dots(4,4) = 8$

$S = \{1,2,3,4,5,6,7,8\}$

P= {1,2,4,8}
Q= {3,5,6,7}
x=6

Step 1:

6 <= Dots(4,4)

Set all dots in the font as black dots

Let n = 6

Go to step 2

Step 2:

6 ∈ Q

Let z=4

Divide the font into 4 sub-blocks

Put one white dot into each sub-block randomly

Let n=6−4=2

Go to step 3

Step 3:

2 ∈ P

Divide the font into 2 sub-blocks of the same size

Put one white dot into each sub-block randomly

Processing is completed.

P is never an empty set because it must contain at least tow elements.

DATA STRUCTURE

In this section, we discuss the data structure that is used to present a font divided by sub-blocks. We use class of CsPattern to create an object, which represents one font. The font is divide into many sub-blocks. Each sub-block is a set of pixels. We call these pixels as the elements of the sub-block. The data structure is shown below:

```
class CsPattern
{
private:
```

```
    POINT PixelLocation [50] [100];
public:
    int PixelAmt [50];
    int TotalSubBlock;
    CsPattern ();
    void AddPixel (int SuBlock, int Row, int Column);
    POINT GetPixel (int SubBlock);
}
```

Setup:

Before we use the font, we have to decide how many sub-blocks we want. After we make up our mind, we set variable of TotalSubBlock to number of the sub-block. Next we have to decide how many pixels belong to each sub-block. We set the array element of PixelAmt [k] to the number of pixels that sub-block k has. Now we use the function, AddPixel (), to add a pixel to desired sub-block. Following example shows how we define a 4×4 font, which is divided into 4 sub-block:

```
CsPattern Book [9];
//Define a 4×4 font, which is divided into 4 sublocks
  Book [4]. TotalSubBlock=4;
  //Sub-block 0
    Book [4]. AddPixel (0,0,0); Book [4]. AddPixel (0,0,1);
    Book [4]. AddPixel (0,1,0); Book [4]. AddPixel (0,1,1);
  //Sub-block 1
    Book [4]. AddPixel (1,0,2); Book [4]. AddPixel (1,0,3);
    Book [4]. AddPixel (1,1,2); Book [4]. AddPixel (1,1,3);
  //Sub-block 2
    Book [4]. AddPixel (2,2,0); Book [4]. AddPixel (2,2,1);
    Book [4]. AddPixel (2,3,0); Book [4]. AddPixel (2,3,1);
  //Sub-block 3
    Book [4]. AddPixel (3,2,2); Book [4]. AddPixel (3,2,3);
    Book [4]. AddPixel (3,3,2); Book [4]. AddPixel (3,3,3);
```

Access:

Fig. 6. The Lena's eye.

After we defined all font, we can use function, GetPixel ( ), to get a pixel randomly in a sub-block. For example, the call of GetPixel (k) randomly choose a pixel in a sub-block k and return its location.

EXPERIMENTAL RESULTS

In the following experiments, we use the book shown in Fig.1 for orthography algorithm. The Lena's eye as shown in Fig.6 is used in the experiments. The Lena's eyes created by orthography algorithm and RPO are shown in Fig.7 (a) and Fig.7 (b) with resolution = 5000 pixel / meter. Both dramatic texture existing near the eyes and screen patterns are removed by RPO. The dither patterns from image with gray level 64 is also shown in Fig.8 and Fig.9. The fixed pattern in Fig.8 is broken by RPO in Fig.9.

The gray scale ramps created by orthography and RPO are also shown in Fig.10 and Fig.11 respectively. The estimated power spectrum of RPO is shown in the Fig.12 and Fig.13 by averaging periodograms of images with gray level 32 and 64 respectively. Much energy at low frequencies is pulled back to high frequencies because the dots are uniformly distributed . The power spectrum is similar to that of error

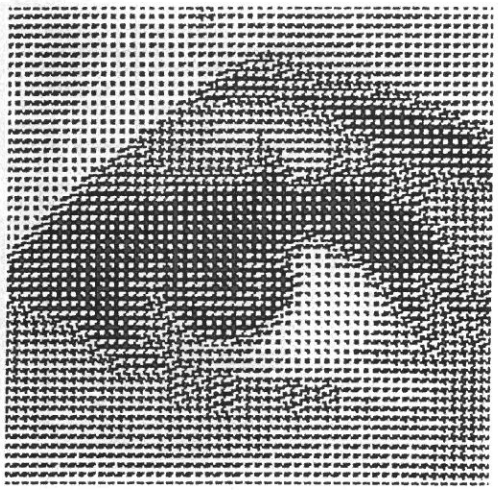
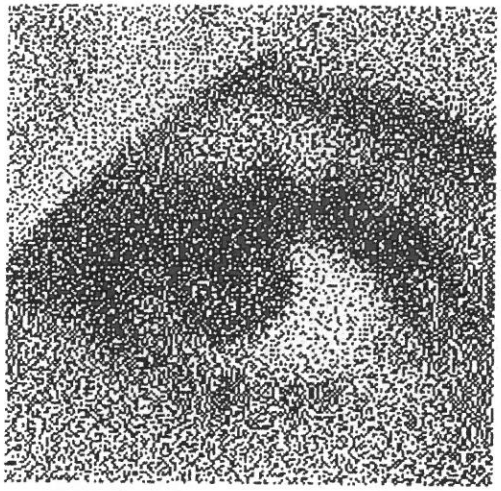


Fig. 7. (a) Lena's eye by orthography.



(b) Lena's eye by RPO.

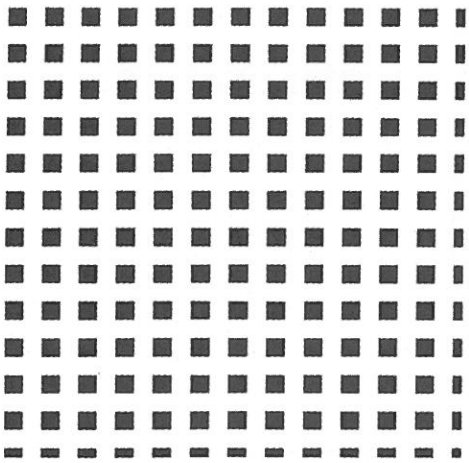


Fig. 8. Dither pattern by orthography.

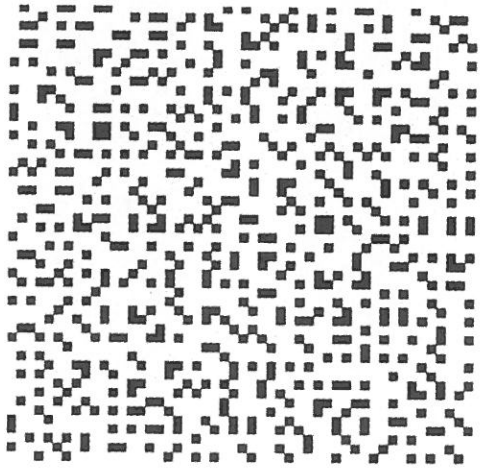


Fig. 9. Dither pattern by RPO.

diffusion except that the power spectrum doesn't posses the high peak in the power spectrum of error diffusion.

SUMMARY

Orthography algorithm can scale up an image and dither the image at the same

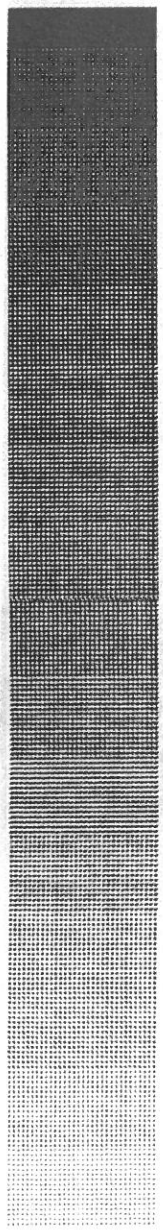


Fig. 10. The gray scale ramp of orthography.

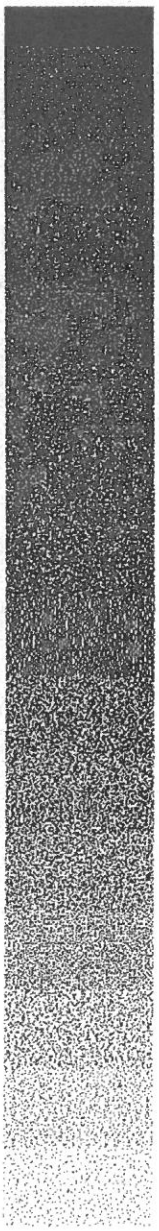


Fig. 11. The gray scale ramp of RPO.

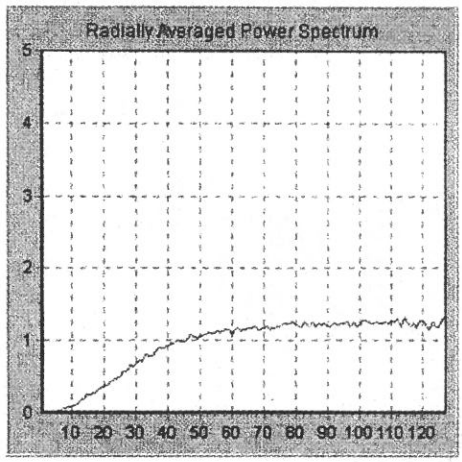


Fig. 12. The estimated power spectrum of RPO with 32 level.

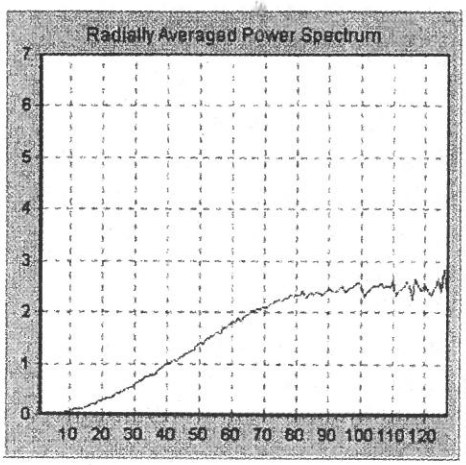


Fig. 13. The estimated power spectrum of RPO with64 level.

time, which will shorten whole processing time when we will need to scale up images before applying algorithm to images. But orthography algorithm does possess several drawbacks. Because orthography algorithm uses one font to reproduce one gray level, when we dither a uniform area, fixed pattern occurs. And it is also common to see some dramatic textures depending on selected font. To solve these problems, we use multiple fonts to reproduce one gray level. The fonts are created at run-time for each pixel in the image by putting desired dots into the font randomly.

REFERENCES

(1) R. Ulichney, Digital Halftoning, MIT Press, Cambridge, MA, 1987

(2) [2] J. F. Jarvis, C. N. Judice, and W. H. Ninke, "A survey of techniques for the display of continuous-tone pictures of bi-level displays," Computer Graphics Image Processing, pp. 13-40, 1976.

(3) J. F. Jarvis and C. S. Roberts, "A New Technique for Displaying Continuous Tone Images on a Believe Display," IEEE Trans. On Communications, pp. 891-897, 1976.

(4) C. Y. Wang, S. N. Chang, L. W. Chang, "Orthography for Image Halftoning," Conference on Computer Vision, Graphics and Image Proceeding, pp. 349-354, 1995.

(5) J. C. Stoffel and J. F. Moreland, "A survey of electronic techniques for Pictorial Image Reproduction," IEEE Trans. On Communication, pp. 1898-1952, 1981.

(6) T. Scheermesser and O. Bryngdahl, "Control of texture in Image Halftoning," J. Opt. Soc. Am. A, Vol 13, No 8, pp. 1645-1652, 1996.

(7) J. Sullivan, R. Miller , and G. Pios, " Image Halftoning using a visual model in error diffusion," J. Opt. Soc. Am. A , Vol. 10, pp. 1714-1724, 1993.

(8) Avideh Zakhor, "A New Class of B/W Halftoning Algorithms," IEEE Trans. On Image Processing, Vol. 2, No. 4, pp. 499-509, 1993.

(9) T Scheermesser and O. Bryngdahl, "Texture metric of halftone images," J. Opt. Soc. Am. A, Vol. 13, No. 1, pp. 18-24, 1996.

(10) J. Sullivan, L. Ray and R. Miller, "Design of Minimum Visual Modulation

Halftone Patterns," IEEE Trans. On Systems, Man. And Cybernetics, Vol. 21, No. 1, pp. 33-38, 1991.

(11) D. J. Granrath, "The Role of Human Visual Models in Image Processing," Proceedings of the IEEE, Vol. 69, pp. 552-561, 1981.

(12) D. Anastassiou, " Error Diffusion Coding for A/D Conversion," IEEE Trans. On Circuits and Systems, Vol. 36, No. 9, pp. 1175-1186, 1989.

(13) P. Hamill. "Line printer modification for better gray level pictures," Computer Graphics Image Processing, Vol. 6, 1974.

(14) J. F. Jarvis, C. N. Judice, and W. H. Ninke, "A survey of techniques for the display of continuous-tone pictures o bi-level displays," Computer Graphics Image Processing, pp. 13-40, 1976.

(15) Bartlett. M. S. , "An Introduction to Stochastic Process with Special Reference to Method and Applications," New York, Cambridge University Press, pp. 274-284

(16) Oppenheim, A. V. and R. W. Schafer, "Digital Signal Processing," New York, Prentice-Hall, pp. 548-549.

received September 1, 1999
revised November 8, 1999
accepted November 28, 1999

隨機位置拼字半色調法

連國珍　葉向榮
輔大資訊工程所

摘　要

拼字半色調法爲傳統半色調法的一種，它以一個 M 乘 N 的字形取代原本影像中的每一個 pixel。使用此半色調法的好處在於當我們必需先將影像放大後再使用半色調法於此影像時，拼字半色調法能同時完成這兩件事，如此一來就能省下許多處理時間。但拼字半色調法有它的缺點。因爲拼字半色調法是以同一個字型取代相同灰度值的 pixel，當我們對一張由相同灰度值的 pixel 所構成的影像使用半色調法時，產生的影像會含有規律性的圖案，因爲人的眼睛對規則性的圖案敏感，所以我們所設計的字形不夠好時，拼字半色調法有時也會產生一些突兀的圖案。在本篇文章裡我們將解決這兩個問題。我們將使用多個字型來取代同灰度值的 pixel。在拼字半色調法裡，字型必須事前決定好，在我們所提出的方法裡，字形是在 run-time 時經由隨機放入一些黑點或白點於字形中所產生的。

關鍵詞： 半色調，拼字，誤差擴散

Gain Scheduling Adaptive Control of Superconducting Magnetic Energy Storage Unit for Power System Stabilization

Yuang-Shung Lee and Chi-Jui Wu
Department of Electronic Engineering
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.

Abstract

A simultaneous active and reactive power (P-Q) modulation gain scheduling control scheme for the superconducting magnetic energy storage (SMES) controller is proposed to increase the damping of the power systems. The static load which is represented by the power polynomial function of the load bus voltage and the dynamic load which is contributed by a group of large induction motor are included to respond to the model fidelity of the studied system. In order to enhance the damping of the electromechanical mode, a gain-scheduled proportional-integral (PI) controller is designed to provide a supplementary damping signal to the firing circuit of the power conditioner of the SMES unit. The gains of the PI controller and the SMES current stabilizer are calculated for the different sets of model system parameters, and scheduled according to the operating state of the system, this can be adjusted in real time by using the on-line measurements and a lookup table of the gain scheduling adaptive algorithms. The eigenvalue analysis and the dynamic simulations for different operating conditions are demonstrated to evaluate the effectiveness of the adaptive control scheme. The results show that the proposed controller, based on gain scheduling approach is much better better than the fixed gain controller during different major disturbance.

Key Words: SMES unit, P-Q modulation, gain scheduling adaptive control, stability

INTRODUCTION

In power system oscillations occurs when there are disturbances in the system such as a change in load or a fault in the system. The damping of the system should be enough so that the synchronous generators can return to steady state after disturbance [1]. The most widely used countermeasures suggested in the literature studies are power system stabilizer [2] and static VAR compensator (SVC) [3].

Due to the rapid development in the technique of high temperature superconducting (HTSC) material, being more efficient and reliable with low cooling cost the HTSCs offer the potential for utility devices that are easy to operate. The HTSC applications become an important issue in electrical engineering [4]. The SMES is designed for efficiently storing energy in the magnetic field incoporated with a circulating current, having the highest round-trip efficiency, and an ability of responding quickly from a charge to discharge mode of the power conditioning system which provides a valuable power system stabilization features [5-7].

The flow of current in the shunt reactor is AC current and only the reactive power modulation is provided to the system. Whereas the flow of current in the SMES coil is DC current where both the active power and reactive power (P-Q) modulations are provided by using the forced-commutated GTO bridge [8] or a hybrid GTO/SCR bridge [9]. Both the active power and the reactive power can be absorbed or released from the unit according to the requirement of the system. The SMES unit can also be applied to damp out the low frequency electromechanical mode oscillations and increase the stability of the power system [6,10].

The load model plays a critical role in the study of dynamic problem of the power system. The static load is represented as the polynomial function of the load bus voltage which is recommended by the IEEE Power System Committee [11-13]. Dynamic load is contributed to the aggregation of large induction motors in the utility of industrial loads [14-16]. An accurate description of the load model is required for

evaluating the dynamic performance of a system.

SMES unit with only the active power modulation control scheme and a fixed proportional-integral (PI) controller is proposed to increase the damping of the turbogenerator [10]. The main drawback of the fixed gain controller is that the gain settings which are suitable for one operating condition may not be suitable for other operating conditions. To apply the SMES in load levelling, both active power modulation and reactive power modulation are required. The controller becomes more effective if its gains can be regularly updated according to the operating conditions of the system.

In this study, a SMES controller and a SMES current stabilizer with very simple gain scheduling scheme are incorporated into the active power and reactive power modulation control of the energy storage unit. The gains of the adaptive controllers are adjusted in real time using the on-line measured operating conditions of the system and with a look-up table stored in the memory of a computer. For a more complete model fidelity, the static load and the dynamic load are included. Results from the eigenvalue analysis and the digital simulation show that the gain scheduling adaptive controller is better than the fixed-gain controller and can always provide a good damping effect to the synchronous generator over a wide range of operating condition.

SYSTEM FORMULATION

The model system consists of a synchronous generator connected with two parallel lines through a single line to the infinite bus as shown in Fig.1. A static load, a dynamic load, and a SMES unit are located at the load bus. The dynamic behavior of the generator can be described by the two-axis model [1], where the transient voltage equations are

$$\dot{E}'_d = [-E'_d - (X_q - X'_d)I_q]/T'_{qo} \tag{1}$$

$$\dot{E}'_q = [E_{FD} - E'_q + (X_d - X'_d)I_d]/T'_{do} \tag{2}$$

The swing equation of the generator can be formulated by

$$\dot{\omega} = (P_m - D_g\omega - P_e)/M_g \tag{3}$$

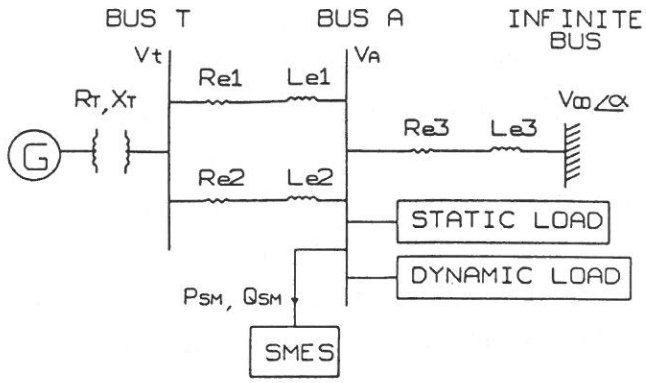


Fig. 1. System model.

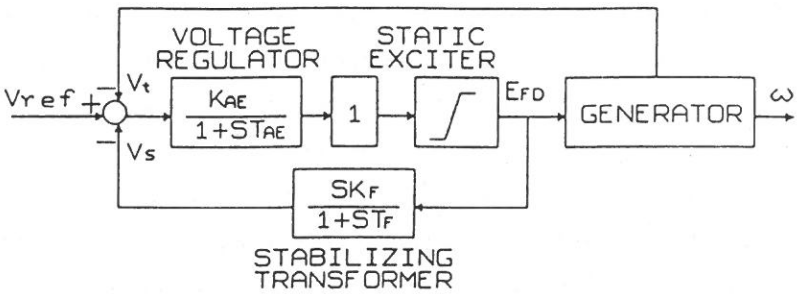


Fig. 2. Exciter.

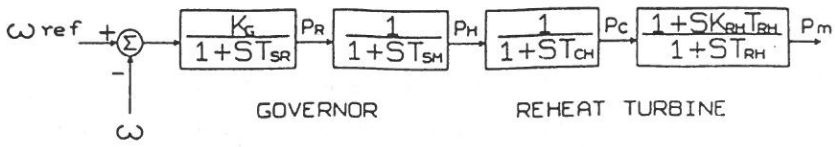


Fig. 3. Governor and turbine.

$$\dot{\delta} = \omega_b(\omega - 1) \tag{4}$$

where $P_e = E'_d I_d + E'_q I_q$ is the electromagnetic power of the generator.

The static excitation system shown in Fig.2 supplies the field current to the generator. Whereas the governor and the turbine system is shown in Fig.3. The static load is generally represented by the voltage dependent polynomial functions [11-13].

$$P_L = P_{LO}(a_p + b_p V_A + c_q V_A^2)$$
$$Q_L = Q_{LO}(a_q + b_q V_A + c_p V_A^2) \tag{5}$$

where a, b, and c are the weighed percentage of the constant power, constant current, and constant impedance in the static load, respectively. The weighed percentage of coefficient a, b, and c are calculated by field test using a mixture data determined from the load bus of the power system.

The dynamic load is constructed by a group of induction motors [13-15]. For the purpose of stability study, these motors can be combined into an equivalent one by the aggregation method [16]. A 5th order dynamic model of the induction motor is used in this study to obtain a more detailed results. The dynamic equations are

$$\frac{d}{dt}[I] = [L]^{-1}\{-[R][I] + [V]\} \tag{6}$$

$$\frac{d\omega_r}{dt} = (T_{me} - T_{mm})/2H_m \tag{7}$$

where

$$[I] = [I_{qsm} \quad I_{dsm} \quad I_{qrm} \quad I_{drm}]^T$$
$$[V] = [V_{qA} \quad V_{dA} \quad 0 \quad 0]^T$$
$$[L] = \begin{bmatrix} L_{SS} & 0 & L_M & 0 \\ o & L_{SS} & 0 & L_M \\ L_M & 0 & L_{rr} & 0 \\ 0 & L_M & 0 & L_{rr} \end{bmatrix}$$
$$[R] = \begin{bmatrix} R_S & -\omega L_{SS} & 0 & -\omega L_M \\ \omega L_{SS} & R_S & \omega L_M & 0 \\ 0 & -(\omega - \omega_r)L_M & R_r & -(\omega - \omega_r)L_{rr} \\ (\omega - \omega_r)L_M & 0 & (\omega - \omega_r)L_{rr} & R_r \end{bmatrix}$$
$$T_{me} = L_M(I_{dsm} I_{qrm} - I_{drm} I_{qsm})$$

Fig.4 gives the configuration of the SMES unit, containing an Y-Δ/Y-Y connected transformer, a 12-pulse cascaded bridge type converter/inverter, and a dc superconducting inductor. The use of forced-commutated GTO converter/inverter [8], the control of the firing angles α_1 and α_2 of the cascaded converter/inverter,

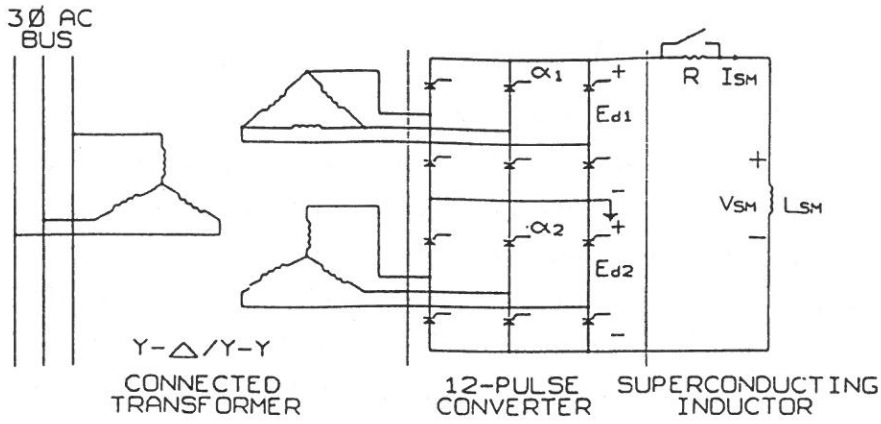


Fig. 4. Schematic configuration of SMES unit.

provides the SMES with the ability to control active and reactive power independently within a circular range containing four quadrants domain.

In order to derive an independent control of the active and the reactive power, let us assume here, to simplify, the equivalent commutating resistance be neglected and the converters operate in equal firing angle mode, i. e., $\alpha_1 = \alpha_2 = \alpha$. The voltage in the DC side of the 12-pulse converter/inverter is expressed by

$$V_{SM} = E_{d1} + E_{d2} = 2V_{SMO}\cos\alpha \tag{8}$$

where E_{d1} and E_{d2} denote the DC voltage output of each 6-pulse converter/inverter, and V_{SMO} is the ideal no-load maximum DC voltage.

The active and the reactive power transferred in the 12-pulse converter /inverter is given by

$$\begin{aligned} P_{SM} &= 2V_{SMO}I_{SM}\cos\alpha \\ Q_{SM} &= 2V_{SMO}I_{SM}\sin\alpha \end{aligned} \tag{9}$$

where the current and voltage of the superconducting inductor is related by

$$I_{SM} = \frac{1}{L_{SM}} \cdot \int_{to}^{t} V_{SM}dt + I_{SMO} \tag{10}$$

where I_{SMO} is the initial current of inductor. The amount of active power and the reactive power drawn from or delivered to the power system can be obtained by

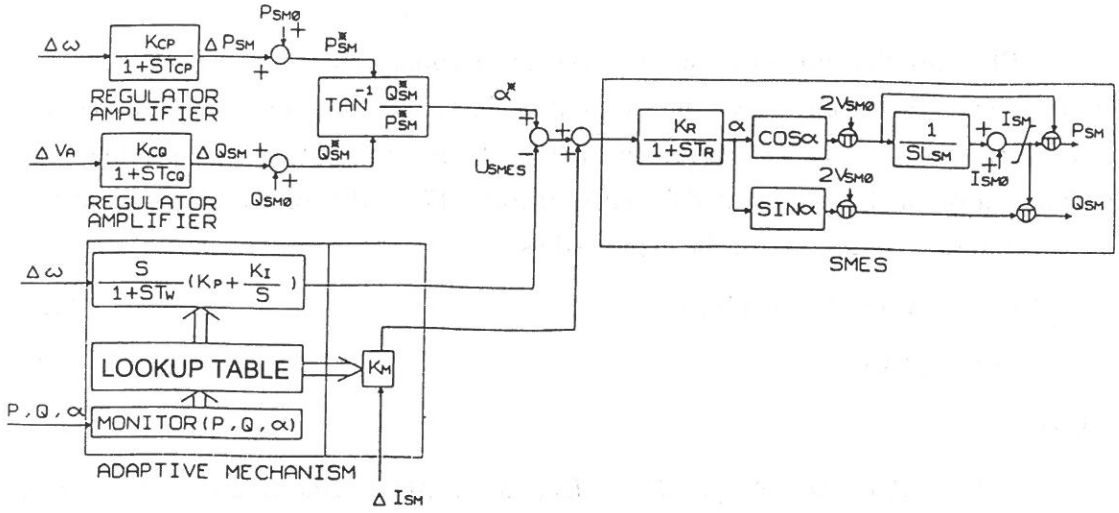


Fig. 5. P-Q modulation scheme of SMES unit with gain scheduling algorithm.

controlling the firing angle of the converter/inverter within the full range between 0° to 360°. The P-Q simultaneous control scheme of the SMES unit is shown in Fig.5.

In order to use the SMES unit as a stabilizer, the active power transferred to the converter/inverter is controlled depending on the measured speed deviation of the generator rotor as shown in Fig.5, where

$$\Delta P_{SM} = \frac{K_{CP}}{1 + sT_{CP}} \Delta\omega \tag{11}$$

The reactive power control is usually for the purpose of voltage stabilization. Thereafter the reactive power transferred is to the converter/inverter controlled depending on the measured voltage deviation of the load bus, i.e.,

$$\Delta Q_{SM} = \frac{K_{CQ}}{1 + sT_{CQ}} \Delta V_A \tag{12}$$

A proportional SMES current stabilizer with a gain K_M is also used to stabilize the SMES dc current during the charging or discharging procedure.

CONTROLLER DESIGN

A PI controller with a washout term is used here to provide a supplementary damping signal to the firing circuit input as shown in Fig.5. Since the two parameters

of the PI controller provides two degrees of freedom, therefore one pair of the electromechanical modes, that is two eigenvalues, can be controlled by the PI controller. In order to determine the gains of the PI controller using the pole assignment method based on modal control theory [17], the linearized system at the initial operating conditions can be expressed as

$$\dot{X}(t) = A\,X(t) + B\,U(t) \tag{13}$$

$$Y(t) = C\,X(t) \tag{14}$$

where

$$X(t) = [\Delta\omega,\ \Delta\delta,\ \Delta E'_d,\ \Delta E'_q,\ \Delta E_{FD},\ \Delta V_s,\ \Delta P_{SM},\ \Delta Q_{SM},\ \Delta I_{SM},\ \Delta\alpha,\ \Delta P_m,\ \Delta P_c,\ \Delta P_H,\ \Delta P_R,\ \Delta I_{qsm},\ \Delta I_{dsm},\ \Delta I_{qrm},\ \Delta I_{drm},\ \Delta\omega_r]^T$$

is the state vector, $U(t) = \Delta U_{SMES}$ is the control signal, and $Y(t) = \Delta\omega$ is the output signal. A, B, and C are all system matrices. The transfer function of the PI SMES controller is given by

$$H(s) = \frac{\Delta U_{SMES}(s)}{\Delta\omega(s)} = \frac{s}{1 + sT_w}\left(K_p + \frac{K_I}{s}\right) \tag{15}$$

where the first term with the known time constant T_w is a washout term used to eliminate the dc offset. The two parameters K_P and K_I, needs to be determined while the electromechanical mode, needs to be stabilized, the controller gain can be solved by substituting the prespecified eigenvalue corresponding to electromechanical mode to the closed-loop system with a characteristic equation:

$$|\,SI - (A + BH(s)C)\,| = 0 \tag{16}$$

where I is an identity matrix. The designed procedure and the results are presented as follows:

Prespecified eigenvalues: $-2.1 \pm j9.3$ (electromechanical mode)

Gains of the PI SMES controller: $K_P = -31.29\ K_I = -759.37$

The values of the prespecified eigenvalues are chosen arbitrarily according to the expected damping of the system. If the gain values of the PI controller is not within a reasonable range, then the assigned eigenvalues must be adjusted until the gain values

Table 1. System eigenvalues at the nominal operating condition.

control scheme / mode	without SMES	with SMES but without PI controller	with SMES and PI controller
generator	$-0.129 \pm j9.899^{*}$ $-1.549 \pm j0.692$	$-0.154 \pm j9.891^{*}$ $-1.545 \pm j0.695$	$-2.1 \pm j9.3^{*}$ $-1.547 \pm j0.695$
exciter	-218.402 -41.638	-218.403 -41.627	-218.446 -45.314
govemor and turbine	-19.940 -10.213 -4.898 -0.1249	-19.939 -10.214 -4.897 -0.1249	-19.936 -10.254 -4.880 -0.1249
induction motor	$-31.399 \pm j403.838$ $-9.361 \pm j34.553$ -15.548	$-31.409 \pm j403.835$ $-9.351 \pm j34.533$ -15.577	$-31.684 \pm j403.763$ $-9.209 \pm j34.653$ -15.576
SMES		-999.996 -38.461 -38.450 -0.0058^{**}	-992.562 -38.461 -38.453 -0.0059^{**} -39.521

*: electromechanical mode
**: SMES current mode

are located in the suitable range.

At the nominal operation point, the eigenvalues of the open loop system, that is without a SMES unit, are tabulated in the second column of Table 1. From the values it is seen that the damping of the electromechanical mode is not adequate enough. The eigenvalues of the system with a SMES unit but without the PI controller are given in the third column of Table 1, where the damping of the electromechanical mode is still not enough. While for a closed loop system with the SMES unit and the proposed PI controller, the eigenvalues are given in the fourth column of Table 1. It can be observed that the eigenvalues of electromechanical mode are exactly at the prespecified position with the proposed control scheme.

Table 2. Electromechanical mode and SMES current stabilizer mode at different operating conditions.

operating condition / mode	$P_0=0.95$, $Q_0=0.25^*$				$P_0=1.15$, $Q_0=0.35$				$P_0=0.75$, $Q_0=0.15$			
	$\alpha_0=-90^\circ$	$\alpha_0=0^\circ$	$\alpha_0=90^\circ$	$\alpha_0=180^\circ$	$\alpha_0=-90^\circ$	$\alpha_0=0^\circ$	$\alpha_0=90^\circ$	$\alpha_0=180^\circ$	$\alpha_0=-90^\circ$	$\alpha_0=0^\circ$	$\alpha_0=90^\circ$	$\alpha_0=180^\circ$
electromechanical mode	−2.1± j9.3	0.542± j0.843	1.985± j0.009	−0.779± j8.907	−1.839± j9.238	1.228± j10.784	2.278± j9.814	0.814± j8.670	−2.280± j9.171	−0.159± j10.664	1.571± j0.001	−0.741± j8.923
SMES crurrent stabilizer mode	−0.0059	0	0.0059	0	−0.0059	0	0.0059	0	−0.0059	0	0.0059	0

*: nominal operating point

The dynamic performance of the system can be maintained at the desired level as long as the system is fixed at a special operation point. However, this can never be achieved in a practical system as the operation conditions may change as a result of major disturbances or a load changes. Whereas the simultaneous P-Q control scheme provides the SMES unit with the ability to control the active power and the reactive power independently in the four guadrants domain. The active power and/or the reactive power can be drawn or released from the SMES unit according to the requirement of the system. For example, if $-90^\circ < \alpha < 0^\circ$, then the active power is transferred to the SMES unit of the power system. On the other hand, if $-180^\circ < \alpha < -90^\circ$, then both the active power and the reactive power are released from the SMES unit to the power system.

In order to survey the system damping at various operating conditions with the SMES controller designed at the nominal operating condition, the electromechanical mode and the SMES current stabilizer mode are given in Table 2. It is observed that these modes damping have not be adequate enough at some of the operating conditions. The gains of the PI SMES controller and the SMES current stabilizer should be adjusted according to the operating conditions of the system.

A simple gain scheduling adaptive control scheme is employed in this study as shown in Fig.6 [18]. In the dynamic period, some of the auxiliary measured system signals are sampled and sent to the computer memory. The gains of the controllers are selected as functions of the auxiliary measured signals. Normally all the possible gains of the controllers are determined off-line and constructed in the look-up table. The advantage of this type of adaptive control scheme is that the gains can be changed

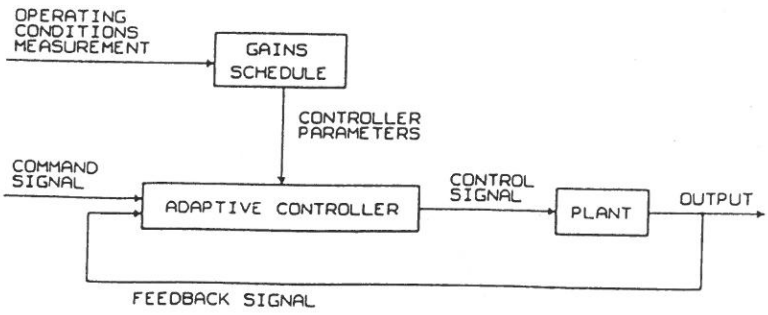


Fig. 6. Gain scheduling adaptive controllers.

quickly in response to the operating condition of the system. The limiting factors depend on how quickly the auxiliary measured signals responds to the changes in the operating condition.

From the eigenvalues analysis it is procured that for positive damping of the SMES current stabilizer mode, the gain K_M can be adjusted according to the following rules.

If $0°<\alpha<180°$ $K_M=0.9$

If $-180°<\alpha<0°$ $K_M=-0.9$

The gains of the PI SMES controller can be obtained by repeating the pole assignment procedure at different operating conditions. As shown in Fig.7, the P-Q plane is divided into several square zones, and in each P-Q zone, the firing angle plane of the converter/inverter is also divided into several sectors. The procedure of pole assignment is repeated in each sector of some P-Q zone using the same prespecified eigenvalues . The gains of the PI SMES controller are determined off-line at the various loading conditions and the modulation mode.

All the gains of the PI SMES controller and current stabilizer at different operating conditions are stored in the computer memory as a look-up table. At each of the sampling instant, from the computer monitor the system's active power (P), reactive power (Q), and the converter/inverter firing angle (α), selects the corresponding gain settings from the look-up table. Thus, the gains of the controller are regularly updated in real time in order to cope with the dynamic operating

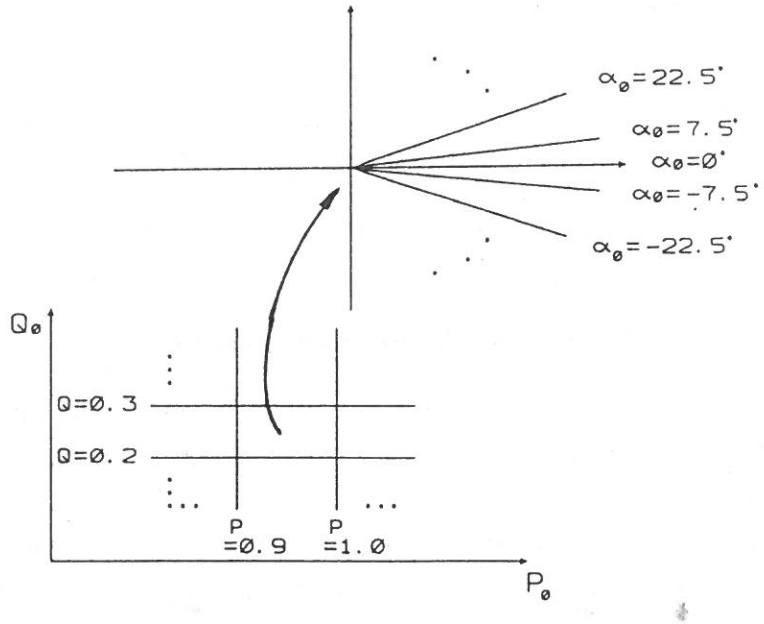


Fig. 7. P-Q zones and α sectors in determining the gains of PI SMES controller.

conditions. The damping effect by the PI SMES controller and the current stabilizer can be ensured through this simple adaptive control scheme. Table 3 gives the gain settings at some of the operating conditions. The gain table of the proposed PI SMES controller for some prespecified firing angle of the converter/inverter are displayed in Fig.8. The prespecified eigenvalues of the electromechanical mode are $-2.1 \pm j9.3$.

NONLINER SIMULATIONS

In order to illustrate the damping effect of the SMES unit with the adaptive control scheme, computer simulations are taken on the basis of a nonlinear model system which includes all kinds of nonlinearites such as the exciter ceiling voltage and the SMES converter current limits, etc.

The dynamic responses of the system subjected to a 100 ms, 0.22 pu mechanical torque changes are shown in Fig.9. Whereas the dynamic responses of the system subjected to a 200 ms open line fault at one of the parallel lines are also given in

Table 3. Gains of the PI SMES controller and the SMES current stabilizer at different operating condition with the electromechanical mode assigned at (−2.1 ± j9.3).

operating condition / mode	$P_0=0.95$, $Q_0=0.25$				$P_0=1.15$, $Q_0=0.35$				$P_0=0.75$, $Q_0=0.15$			
	$\alpha_0=-90°$	$\alpha_0=0°$	$\alpha_0=90°$	$\alpha_0=180°$	$\alpha_0=-90°$	$\alpha_0=0°$	$\alpha_0=90°$	$\alpha_0=180°$	$\alpha_0=-90°$	$\alpha_0=0°$	$\alpha_0=90°$	$\alpha_0=180°$
K_P	−31.29	−45.13	34.07	40.16	−25.99	−81.13	28.47	73.02	−28.73	−32.04	31.39	29.47
K_I	−759.37	1068.78	788.79	−1047.29	−673.59	940.18	697.99	−953.15	−852.67	981.85	887.98	−955.63
K_M	−0.9	0.9	0.9	0.9	−0.9	0.9	0.9	0.9	−0.9	0.9	0.9	0.9

*: nominal operating point

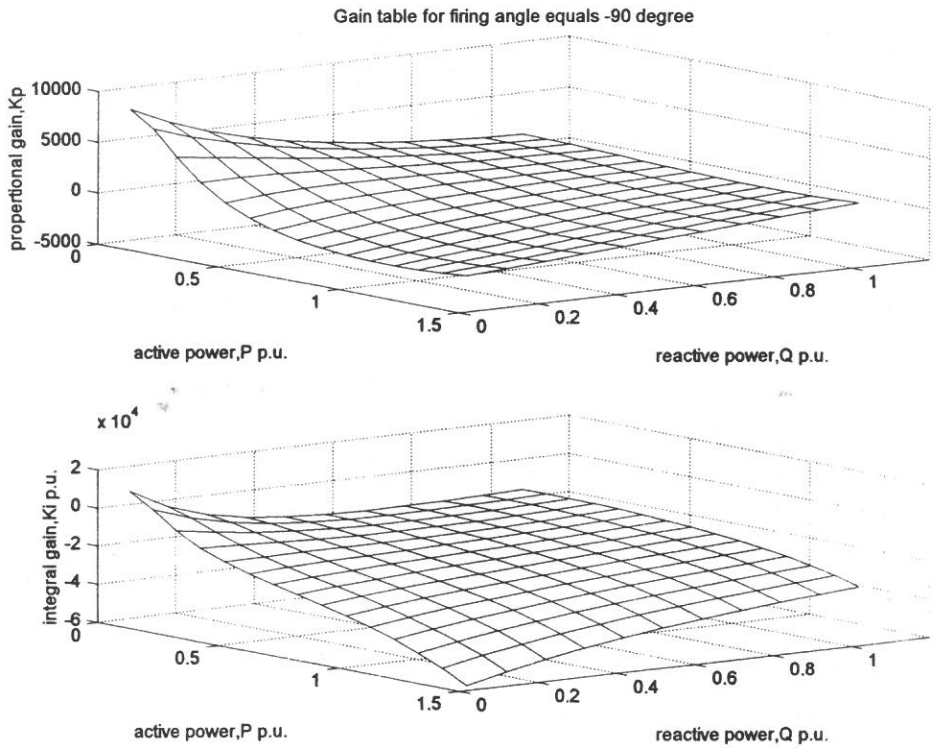


Fig. 8. Gain table for firing angle $\alpha = 90°$.

Fig.10. The responses of speed deviation ($\Delta\omega$) and load bus voltage (V_A) with adaptive controllers or fixed-gain controllers are drawn together for comparison. Fig.11. shows the performance of the system under a series of a wide range of mechanical torque changes when applied to the system. Several observations can be obtained from the Figs.9-11.

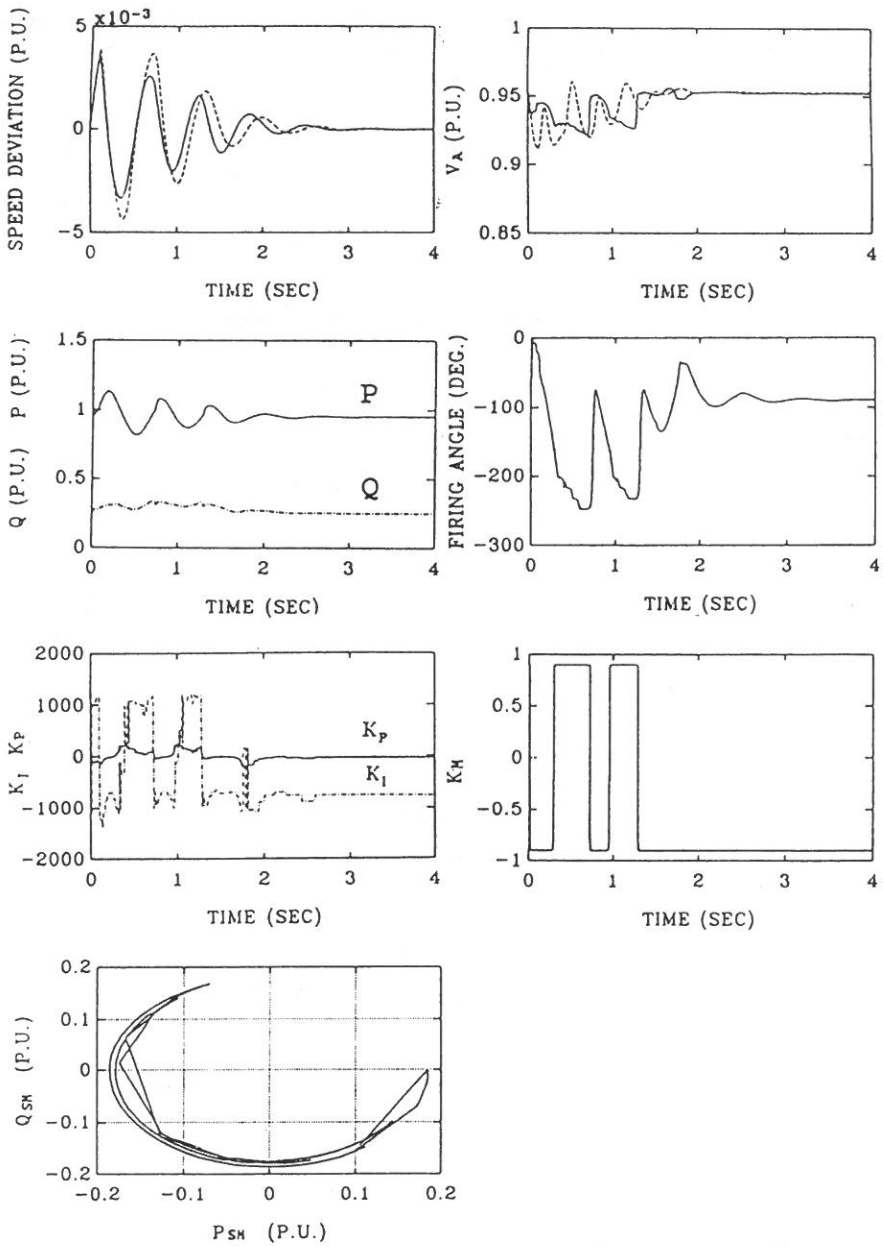


Fig. 9a. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = -90^\circ$.

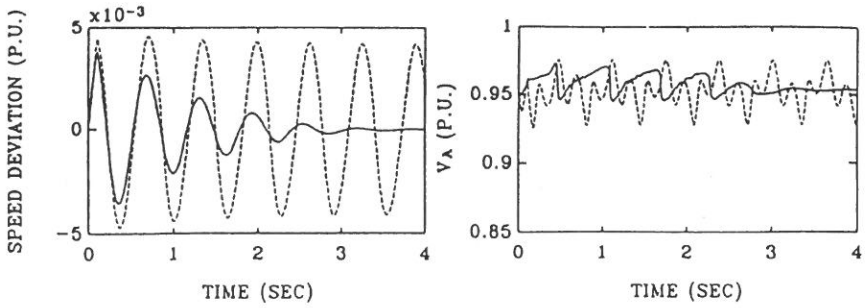


Fig. 9b. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 0^\circ$.

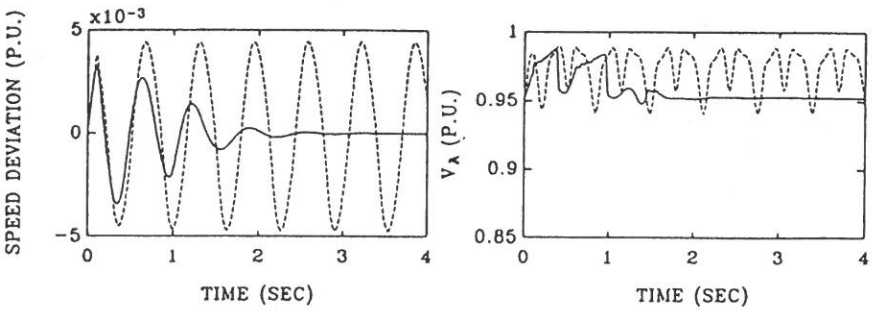


Fig. 9c. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 90^\circ$.

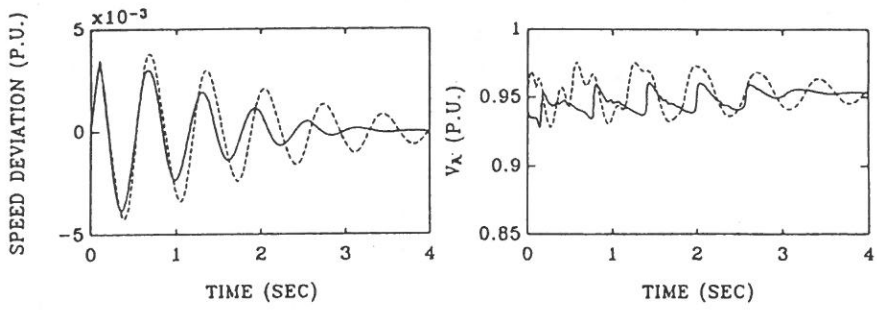


Fig. 9d. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 180^\circ$.

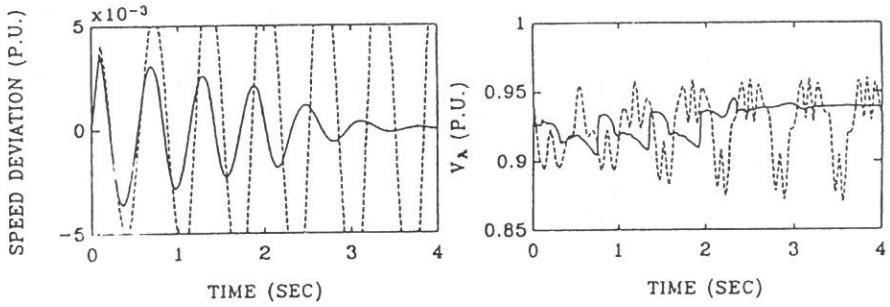


Fig. 9e. $P_0 = 1.15$, $Q_0 = 0.35$, $\alpha_0 = -90^\circ$.

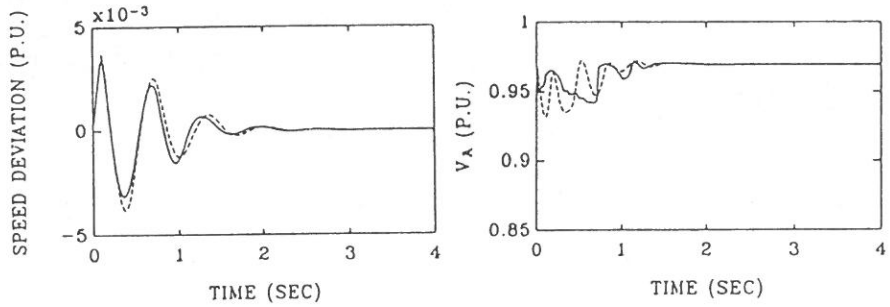


Fig. 9f. $P_0 = 0.75$, $Q_0 = 0.15$, $\alpha_0 = -90^\circ$.

Fig. 9. Response of system subjected to a mechanical torque change
—— with gain scheduling adaptive controllers
------ with fixed-gain controllers.

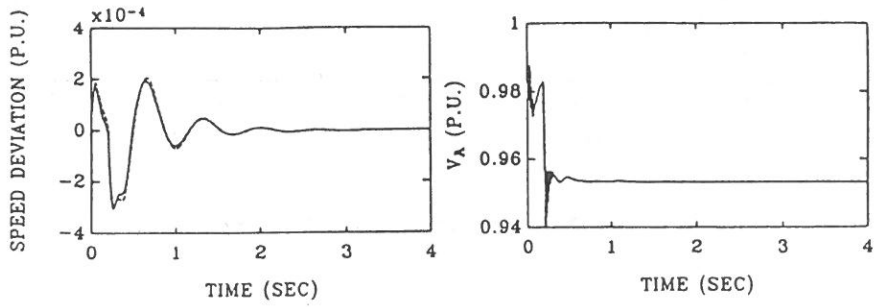


Fig. 10a. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = -90^\circ$.

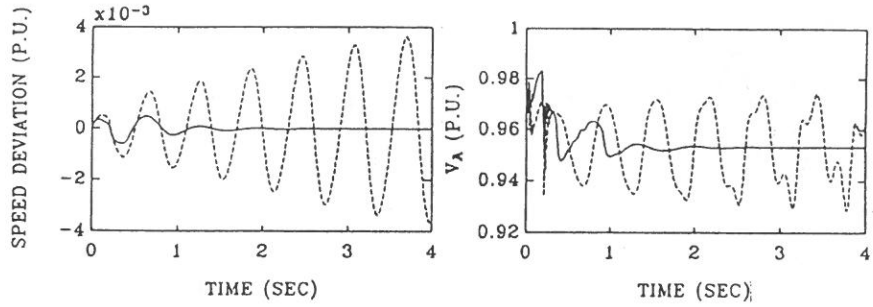


Fig. 10b. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 0°$.

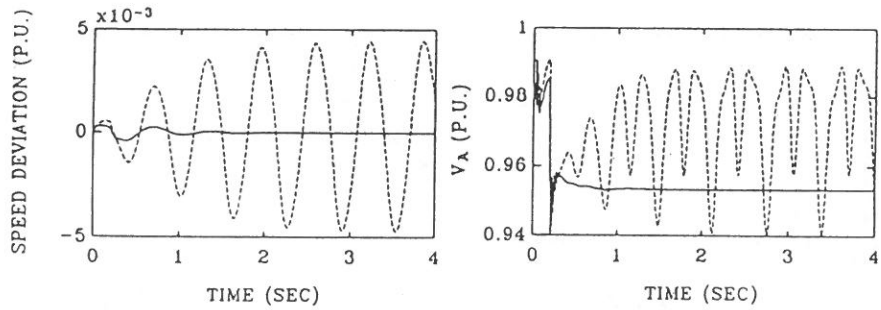


Fig. 10c. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 90°$.

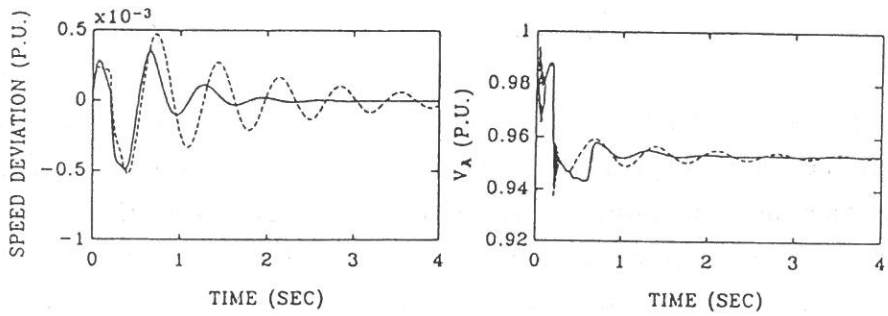


Fig. 10d. $P_0 = 0.95$, $Q_0 = 0.25$, $\alpha_0 = 180°$.

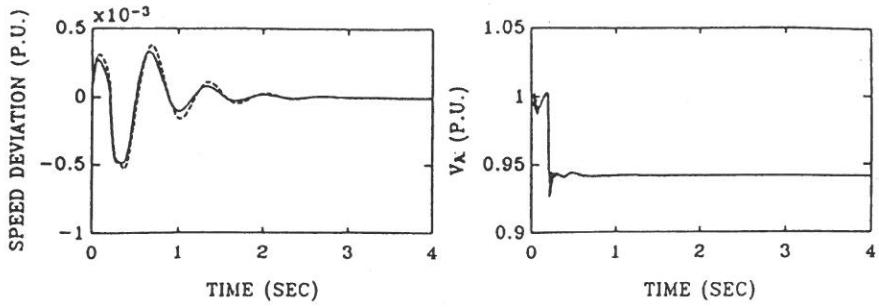


Fig. 10e. $P_0 = 1.15$, $Q_0 = 0.35$, $\alpha_0 = -90°$.

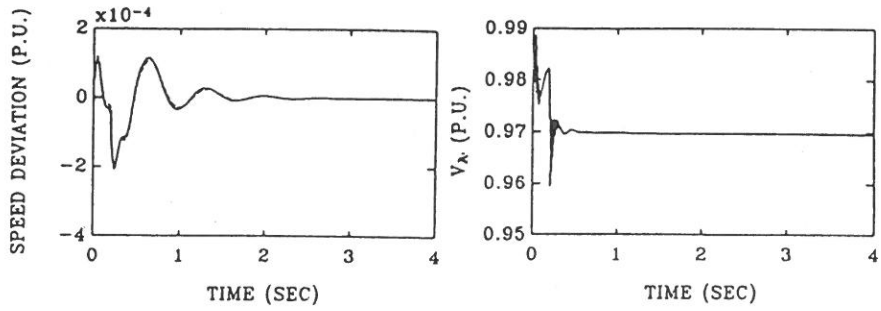


Fig. 10f. $P_0 = 0.75$, $Q_0 = 0.15$, $\alpha_0 = -90°$.

Fig. 10. Response of system subjected to an open line fault
—— with gain scheduling adaptive controllers
------ with fixed-gain controllers.

(a) Although the PI SMES controller and the current stabilizer with fixed gains can provide the desired damping effect at the nominal operating condition, they can not provide the desired damping effect at other operating conditions. The SMES unit with the adaptive PI controller and the SMES current stabilizer can always suppress the system's oscillations caused by major disturbances in the system.

(b) The loading conditions (P, Q, α) of the system varies during the dynamic period. The gains of the PI SMES controller and the SMES current stabilizer are regularly updated at each sampling instant by measuring the conditions of the system.

(c) The SMES unit can provide both the active power and the reactive power

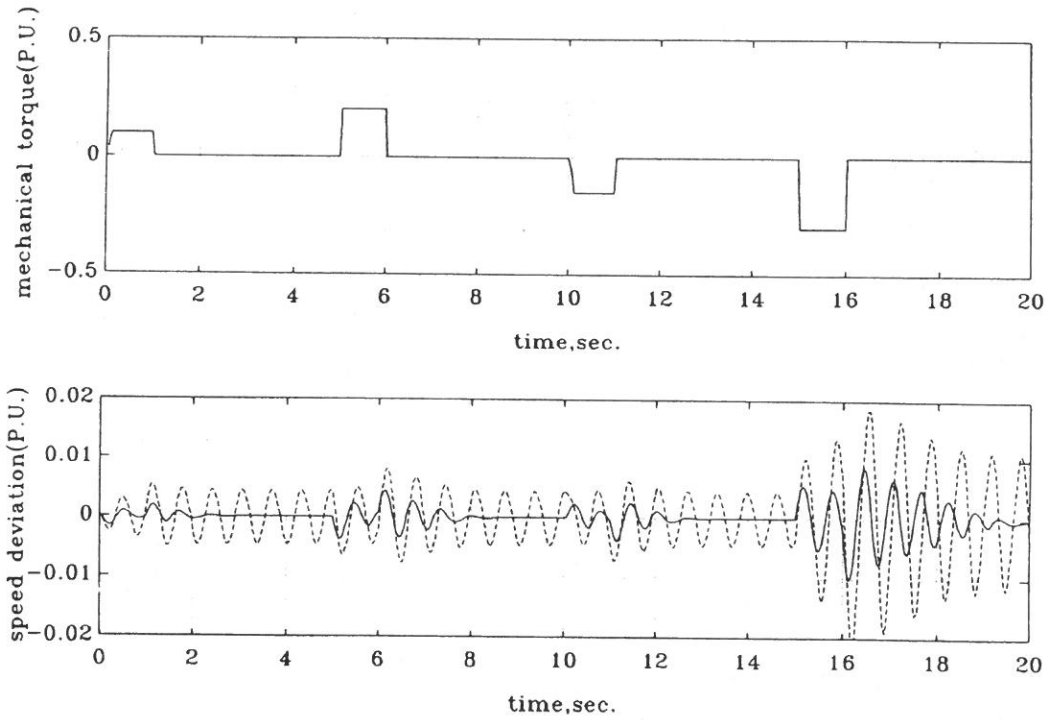


Fig. 11. Responses under a series of heavy load disturbance
------ with gain scheduling adaptive controllers
------ with fixed-gain controllers.

modulations to the power system for disturbance or load leveling. The P_{SM}-Q_{SM} locus is a part of the circle diagram.

(d) The voltage profile of the load bus voltage is also improved by the proposed control scheme with reactive power modulation.

(e) The smaller size of the P-Q zone and the α sector, the shorter is the sampling time, and thus a more pronounced damping effect can be obtained from the adaptive controllers. However they are limited by the hardware of computer memory and sampling device. Hence compromises should be made. In this studied, a zone size of 0.1 p.u. × 0.1 p.u. in P-Q plane, and that of 15° for α plane are selected. Sampling time of 1 ms is used.

CONCLUSION

The gain scheduling approach for the PI SMES controller to damp electromechanical oscillation is investigated in a power system incorporated with the static and dynamic load. The variations in the states of the system makes it difficult to get a fix set of feedback parameters which gives an adequate dynamic performance over a wide range of operations. The proposed algorithm can store the preassigned gains, which corresponds to the operating conditions of the system and schedule them according to the on-line system measurements at each sampling instant. The results show that a substantial enhancement of the system damping can be obtained using the gain scheduling SMES controller for various any operating conditions and disturbance. The SMES unit can operate in the full range of hour quadrant for the desired of the dynamic system. The performance of the gain scheduling adaptive controllers is much better than that of the fixed gain controllers especially during the dynamic load leveling. The on-line tuning of fuzzy-neural network for adaptive control of the power system incorporated with the SMES unit can be suggested for later study.

LIST OF PRINCIPAL SYMBOLS

generator

ω	angular speed
δ	torque angle
E'_d, E'_q	d-axis and q-axis transient voltage
E_{FD}	field voltage
V_s	stabilizing transformer output voltage
P, Q	active power and reactive power outputs
V_t, V_A	terminal voltage of bus T and bus A

SMES unit

V_{SM}, I_{SM}	voltage and current of superconducting inductor

P_{SM}, Q_{SM} active and reactive power transferred to SMES

L_{SM} inductance of superconducting inductor

K_{CP}, K_{CQ} gains of measurement device

T_{CP}, T_{CQ} time constant of measurement device

K_R, T_R firing circuit loop gain and time constant

Induction motor

V_{dsm}, V_{qsm} d and q-axis stator voltage

I_{dsm}, I_{qsm} d and q-axis stator current

I_{drm}, I_{qrm} d and q-axis rotor current

ω_r angular rotor speed

T_{me}, T_{mm} electrical and mechanical torque

L_{ss}, L_{rr} stator and rotor self-inductances

R_s, R_r stator and rotor resistances

L_M mutual inductance between stator and rotor

H_m inertia constant

REFERENCES

(1) P. M. Anderson and A. A. Fouad, Power system control and stability, Iowa State University Press, 1977.

(2) E. V. Larsen and D. A. Swann, "Applying power system stabilizers," IEEE Transactions on power apparatus and systems, Vol. 100, 1981, pp. 3017-3046

(3) M. O'Briem and G. Ledwich, "Static reactive power compensation controls for improved system stability," IEE Proceeding, Vol. 134, Pt. C, No. 1, 1987, pp. 38-42.

(4) R. F. Giese, T. P. Sheahen, A. M. Wolsky, and D. K. Sharma, "High-temperature superconductors: their potential for utility applications," IEEE Transaction on energy conversion, Vol. 7, NO. 3, 1992, pp. 589-597.

(5) J. B. Simo and I. Kamwa, "Exploratory assessment of the dynamic behavior

of multimachine system stabilized by a SMES unit," IEEE Trsans. on Power System, Vol. 10, No. 3, 1995, pp. 1566-1571.

(6) Y. Mitani, K. Tsuju, and Y. Murakami, "Applications of superconducting magnetic energy storage unit to improve the power system dynamic performance," IEEE Transaction on power system, Vol. 3, 1988, pp. 1418-1425.

(7) S. Bartos and G. T. Heydt, "Increased operating flexibility for superconducting magnetic energy storage system through the use of self-commutation," IEEE Transaction on power system, Vol. 3, 1988, pp. 944-948.

(8) T. Ise, Y. Murakami, and K. Tsuji, "Simultaneous active and reactive power control of superconducting magnetic energy storage using GTO converter," IEEE Transaction on power delivery, Vol. 1, No. 1, 1986, pp. 143-150.

(9) J. Wang, R. L. Kustom, J. J. Skiles, and F. Vong, "Results of the P and Q control with a model hybrid GTO/SCR bridge for SMES," International Conference on High Technology in the Power Industry, IASTED, 1991, pp. 239-241.

(10) C. J. Wu and Y. S. Lee, "Application of superconducting magnetic energy storage unit to improve the damping of the synchronous generator," IEEE Transaction on energy conversion, Vol. 6, No. 4, 1991, pp. 573-578.

(11) W. J. Lee, M. C. Chen, and L. B. Williams, "Load model for stability studies," IEEE Transaction on industry application, Vol. 23, 1987, pp. 159-165.

(12) M. A. Choudary, M. A. Reza, and K. A. Ellithy, "Design of a robust modulation controller over a wide range of load characteristics for the AC/DC system," IEEE Transaction on power systems, Vol. 5, 1990, pp. 212-218.

(13) K. A. Ellithy and M. A. Choudary, "Effect of load models on the AC/DC system stability and modulation control design," IEEE Transaction on power systems, Vol. 4, 1988, pp. 411-418.

(15) A. E. Hammaf. and M. A. El-Sadek, "Prevention of transient voltage instability due to induction motor loads by static var compensators," IEEE Transation on power systems, Vol. 9, 1984, pp. 1182-1190.

(16) K. O. Abledu and A. A. Mahmoud, "Equivalent load model of induction machines," Electric machine and power systems, Vol. 9, 1984, pp. 435-452.

(17) J. J. Cathey, R. K. Cavin III, and A. K. Ayoub, "Transient load model of an induction motor," IEEE Transaction on power apparatus and system, Vol. 92, 1973, pp. 1399-1406.

(18) T. Kailath, "Linear System," (Prentice-Hall, 1980).

(19) K. J. Astrom and B. Wittenmark, "Computer Controlled Systems," Theory And Designing, (Prentice-Hall, 1984).

APPENDIX

The data used for th system are given as follows [10, 11, 16]:

Generator and transmission line

$X'_d = 0.245$ p.u. $X_d = 1.7$ p.u. $X_q = 1.64$ p.u.

$R_a = 0.001096$ p.u. $D_g = 0$ $M_g = 4.74$ s

$T'_{qo} = 0.075$ s $T'_{do} = 5.9$ s $\omega_b = 377$ rad/s

$R_T = 0.0002$ p.u. $X_T = 0.02$ p.u. $R_{e1} = R_{e2} = 0.03$ p.u.

$L_{e1} = L_{e2} = 0.6$ p.u. $R_{e3} = 0.005$ p.u. $L_{e3} = 0.05$ p.u.

exciter, governor, and turbine

$K_{AE} = 400$ $T_{AE} = 0.05$ s $K_F = 0.025$ s

$T_F = 1$ s $K_G = 3.5$ $T_{CH} = 0.05$ s

$T_{RH} = 8$ s $T_{SR} = 0.1$ s $T_{SM} = 0.2$ s

$K_{RH} = 0.3$

50% static load

$c_p = 0.25$ $c_q = 0.12$

Induction motor load

Rated 1500 HP Rated voltage 2.3 KV

$R_5 = 0.056$ W $R_r = 0.037$ W $L_s = 0.0537$ H

$L_r = 0.0537$ H $M = 0.0527$ H $J = 44.548$ kg-m^2

20 motors are used to represent 50% of the dynamic load

SMES unit

$I_{SMO} = 0.21109$ p.u. $V_{SMO} = 0.4387$ p.u. $L_{SM} = 0.5$ H

$R_{SM}=0$ $K_{CP}=1.83$ $T_{CP}=0.026$ s

$K_{CQ}=1.83$ $T_{CQ}=0.026$ s $K_R=1$

$T_R=0.001$ s $T_W=0.025$ s

Nominal operating condition

$P_o=0.95$ p.u. $Q_o=0.25$ p.u. $V_t=1.0$ p.u.

$\alpha_o=-90°$ $K_M=-0.9$

received September 30, 1999
revised October 17, 1999
accepted November 30, 1999

超導磁能儲存系統之增益規劃適應性控制以改善電力系統穩定度

李 永 勳　　吳 啓 端
輔仁大學電子工程系

摘　要

本研究提出具同時調變實功率及虛功率之超導磁能儲存控制器增益規劃控制法以增進電力系統之阻尼。爲顯現模型精確性，系統包含負載匯流排電壓的多項式函數表示的靜態負載，和由一群感應電動機所組成的動態負載。爲增強電磁模式之阻尼，設計一增益規劃比例—積分控制器用以提供一額外阻尼訊號到超導磁能儲存系統的電力調製觸發電路。PI 控制器和 SMES 電流穩定器的增益是在不同系統參數組合和依系統工作狀態下規劃計算出來的，然後利用線上量測和增益規劃查表的適應性演算法作即時調整。不同工作點的特徵值分析和動態模擬用以評估適應控制法的效果，結果證明所提具有增益規劃的適應控制器，在遭遇不同干擾下的阻尼效果比固定增益控制器來得好。

關鍵詞：超導磁能儲存系統、P-Q 調變、增益規劃適應性控制、穩定度

A Modified M/M/2 Queueing Model for Tri-state Power Management Methods in a Dual Processor System

Jung-ren Hu and Ying-Wen Bai
Department of Electronic Engineering
Fu Jen Catholic University
Taipei, Taiwan, 242, R.O.C.

Abstract

Using the proposed tri-state power management methods, a dual processor system can work during the waiting state, during which there is only one processor "on", with low power consumption to accumulate sporadically arriving information from networks. In addition, the system can also work during the busy state, during which there is one processor "on", at high-power consumption when the amount of the accumulated information reaches a preset level. Furthermore, this system can also work during a congested state when a burst of multimedia information arrives from networks or when heavy computing loads are needed. During the congested state, the second processor will be turned on to provide the highest processing ability. Based on the analysis of the stationary non-time-varying state transition diagram from a modified M/M/2 queueing model, this paper shows that this tri-state power management method provides the relationship between power efficiency and latency of a dual processor information processing system. In addition, both the analysis and the simulation results show that this method also provides a potential to save approximately 70% power consumption with a latency of about five times based on a homogeneous arrival information.

Key Words: power management, mobile terminal, dual processor system, state probability.

INTRODUCTION

Over the past few decades, numerous researchers have been working on different levels of a low power computing system design [1,2]. Power reducing design is especially attractive due to its potential of extending the recharge period of mobile terminals. The purpose of power management not only provides maximum battery life for a mobile working environment with a limited battery energy capacity [3,4], but also provides optimum performance during burst information arriving from networks.

Traditionally, the power management mechanism is implemented by counters or timers which can count the length of the waiting period. The machine uses counters as a reference to notify each subsystem whenever there is a switching in the state of the machine. Therefore, using this control mechanism, the machine can turn subsystems on and off when the timer reaches its limits. However, if the counter setting is not reasonable, the switching frequency will be too high and may degrade the system performance and power efficiency [6,7,8,9].

To deal with "busy" and "idle" dynamic interleaf switching in a mobile computing device with a dual low-power processor system, we propose a tri-state power management method as the control mechanism, containing three machine states. To fulfill this control mechanism, we use the accumulated amount of information as a driving event for the machine state transition. This is used as the trigger of the control mechanism of machine states. In this system, different machine states have different levels of power consumption, therefore the average power consumption can be reduced by the machine state control and the use of the low-power state.

The latest power management model, "Advanced Configuration and Power Interface Specification" [3], also specifies a way that the OS and peripheral devices "talk" to each other about power usage. Thus, ACPI can be used to build a tri-state power management model directly which can enable or disable power activities-providing power devices only on a as-needed-basis at different machine states.

Based on the analysis of a modified M/M/2 queueing model, this paper shows the relationship between power efficiency and latency of mobile information processing under a tri-state power management method. The latency is a result of the probability distribution of the machine states. The machine state can be determined by the user operation model, the terminal characteristics and the variation of the information flow rate of multimedia networks. Furthermore, the current technology trend can be used to estimate both the power consumption at each machine state and the average power consumption for all machine states [10,11,12].

The rest of this paper is organized as follows. In section 2, the typical palmtop terminal operation model is discussed. In section 3, the tri-state power management method in a dual processor system is presented. Section 4 shows how the power consumption is estimated based on the tri-state power management method. Section 5 shows how to find the relationship between the power efficiency and the latency for a tri-state power management method. Section 6 shows some simulation results. The final section presents conclusions.

TYPICAL PALMTOP TERMINAL OPERATION MODEL

The operation of a typical palmtop multimedia terminal can be modeled by a two-dimensional state transition diagram, as shown in Fig.1. This model has two major characteristics. The y-direction represents the amount of received commands stored in the command queue and the x-direction represents the amount of received information stored in the information queue. While the palmtop terminal is executing the current command from the mobile user, any new commands from a user will be temporarily stored in the command queue and executed later. Similarly, the machine will temporarily store the information arriving from the network in a information queue for processing later.

To determine the state probability of each machine state, we define p_{ij} as the joint probability so that command length is i and information length is j on:

$$P_{ij} = P[\text{command queue length} = i, \text{ information queue length} = j] \quad (1)$$

To simplify the analysis, we assume that the arrival process is a Poisson distribution and the service process is an exponential distribution. There are a couple of

ways to solve p_{ij} in the two-dimensional state transition diagram of Fig.1. Using moment-generating functions is one possibility. Another approach, used here, is called the matrix geometric method. All the solution techniques rely on setting up two dimensional balance equations for the two-dimensional state transition diagram. We equate the rate of leaving a state to the rate of entering it. We then get first, for, $i = 0$,

$$i = 0,\ j = 0\text{:}\ (\lambda_C + \lambda_I)P_{00} = \mu_I P_{01} + \mu_C P_{10} \tag{2}$$

$$i = 0, 1 \leq j \leq N - 1\text{:}\ (\lambda_C + \lambda_I + \mu_I)P_{0j} = \lambda_I P_{0j-1} + \mu_C P_{1j} + \mu_I P_{0j+1} \tag{3}$$

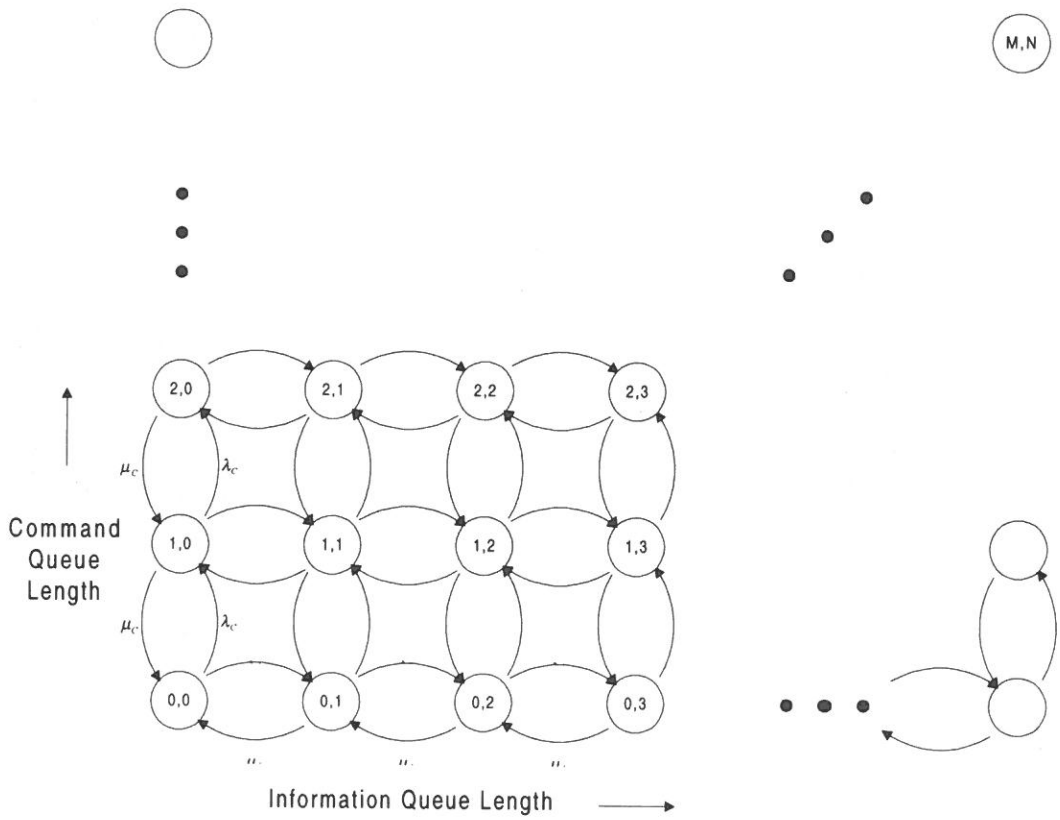


Fig. 1. Two-dimensional state transition diagram.

$$i = 0,\ j = N:\ (\lambda_C + \mu_I)P_{0N} = \lambda_I P_{0N-1} + \mu_C P_{1N} \tag{4}$$

For the case, $1 \leq i \leq M$, the state balance equation can be determined. Usually, the system of equations will not be inconsistent, and there will be an existing solution. The solution $P = [P_{ij}]$, $0 \leq i \leq M$, $1 \leq j \leq N$, shows the state probability that the machine will stay on.

For a mobile-computing environment, the power consumption can be different at different machine states. Therefore, the average power consumption of a palmtop terminal can be smaller if there is a low-power state controlled by the multiple state power management method.

$$POWER_{average} = \sum_{\substack{i=0 \\ j=0}}^{\substack{i=M \\ j=N}} P_{ij} power_{ij} \tag{5}$$

We shall also mention a simple approximation technique which is a useful condition during which the two random variables i and j may be decoupling. This approach is particularly applicable in the case of multimedia information sources. In the following sections, we will discuss how to group the "waiting", "busy", and "congested" states to form a tri-state power management method.

THE TRI-STATE POWER MANAGEMENT METHOD IN A DUAL PROCESSOR SYSTEM

In the mobile operation environment, due to the stochastic property of the information arrival process from computer networks, the delay time variance of information transmission can be difficult to estimate precisely. Therefore, we use a certain probability distribution to make the estimation. Basically, we will focus on the x-direction shown in Fig.1, which is an information queue based on our previous discussion. By setting the length and the utilization of the information queue, we can estimate the waiting probability and power consumption efficiency.

As previously mentioned, a tri-state power management method provides the control mechanism so that the dual processor system will stay in the "waiting state" until the amount of received information reaches a predefined level. Then the machine will switch one processor to the "busy state" and process the received information. If

the machine has received a large amount of information, the control mechanism will turn on both processors to process information. The state transition diagram in Fig.2 shows the scalable batch processing of the information queue.

In Fig.2, we consider the states from 0 to L-1 as a group of states, called a waiting state; the state L as a group of states, called a busy state; and the other states as another group of states, called a congested state. Then Fig.2 can be simplified and represented by a "tri-state" transition diagram, as shown in Fig.3.

We use a modified $M/M/2$ queuing model to represent our tri-state power management method, which is a Poisson arrival distribution and an Exponential departure distribution with two parallel servers that operate at the same rate μ. Based on the discussion above, we will switch a processor into high-speed state when

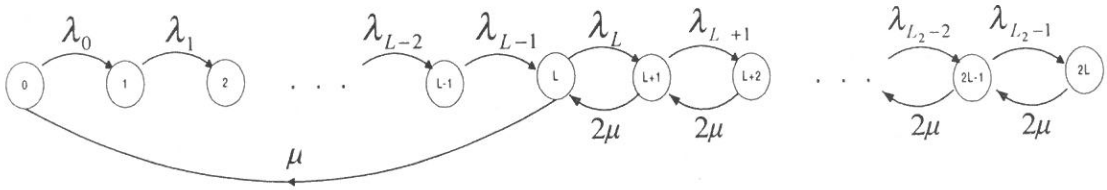


Fig. 2. State transition diagram of the scalable batch model of a dual process or system.

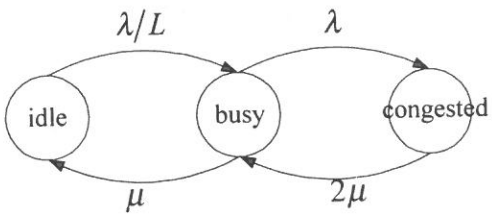


Fig. 3. Tri-state transition diagram of the dual processor system.

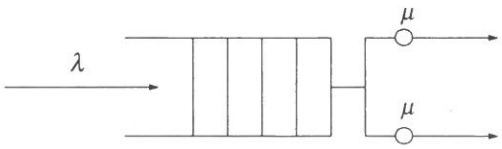


Fig. 4. A modified $M/M/2$ queuing model.

information in the queue reaches length L and turn on another processor when information in the queue exceeds length L. From the statistical point of view, we use the state transition model shown in Fig.4, where λ are the information arrival rates of the information queue, and $\mu_1 = \mu$, $\mu_2 = 2\mu$ are the information service rates of the dual processor system. We assume that the arrival rate is not influenced by the state of the system.

POWER CONSUMPTION ESTIMATION BASED ON A TRI-STATE POWER MANAGEMENT METHOD

Using Fig.3, we set up a balance equation at each state as follows:

$$\mu P_{busy} = \frac{\lambda}{L} P_{waiting} \tag{6}$$

$$(\mu + \lambda) P_{busy} = \frac{\lambda}{L} P_{waiting} + 2\mu P_{congested} \tag{7}$$

where $\frac{\lambda}{L}$ is an equivalent information arrival rate, μ is an average service rate, P_{budy} is the "busy probability" of the machine, $P_{congested}$ is the "congested probability", and $P_{waiting}$ is the "waiting probability" of the machine. In order to simplify the analysis, we assume the machine stands only at the "waiting state", "busy state" or "congested state", so that we have

$$P_{waiting} + P_{busy} + P_{congested} = 1 \tag{8}$$

During the waiting state, the control mechanism can turn off some subsystems to reduce power consumption, and during the busy state, there is only one CPU on for processing. The average power consumption, $POWER_{average}$, can be expressed as the summation of the power consumption during the "waiting state", "busy state" and "congested state".

$$POWER_{average} = P_{waiting} POWER_{waiting} + P_{busy} POWER_{busy} + P_{congested} POWER_{congested} \tag{9}$$

From the basic analysis of a modified M/M/2 queuing model, we can get

$$D_{latency} = W_s + L/2 \tag{10}$$

where $W_s = \frac{1}{\mu} + \frac{\rho^3}{4 \cdot \lambda (1 - \frac{\rho}{2})^2} \cdot P_0$ (W_s: average packet waiting time)

$\rho = \frac{\lambda}{\mu}$

L: *queue length*

If we input equation (6), (7) and (10) into equation (9), we then obtain that

$$POWER_{average} = \frac{1}{1 + \varepsilon + \rho\varepsilon/2} \cdot POWER_{waiting} + \frac{\varepsilon}{1 + \varepsilon + \rho\varepsilon/2} \cdot POWER_{busy} + \frac{\rho\varepsilon/2}{1 + \varepsilon + \rho\varepsilon/2} POWER_{congested} \quad (11)$$

$$\varepsilon = \frac{\rho}{D_{latency} - W_s}$$

Usually, the service rate μ is greater than the arrival rate , the queue length L is greater than 1, and $POWER_{waiting}$ is smaller than $POWER_{busy}$ and $POWER_{congested}$ because storing information uses less power than processing information. According to the previous discussion and equation (11), we see that

$$POWER_{average} \leq POWER_{congested} \quad (12)$$

Equation (11) shows that a tri-state power management method can save power consumption because the machine can switch itself into the "waiting state" with lower power consumption.

RELATIONSHIP BETWEEN POWER EFFICIENCY AND LATENCY IN A TRI-STATE POWER MANAGEMENT METHOD

From the handheld computing technology trend, we have determined that the amount of power consumption of the waiting state $POWER_{waiting}$ is usually in proportion to the length of the information queue. Furthermore it is in proportion to the size of the RAM chips as shown in the Fig.5 [11]. Hence, $POWER_{waiting}$ can be expressed as equation (13)

$$POWER_{waiting} = KL + POWER_{sleeping} \quad (13)$$

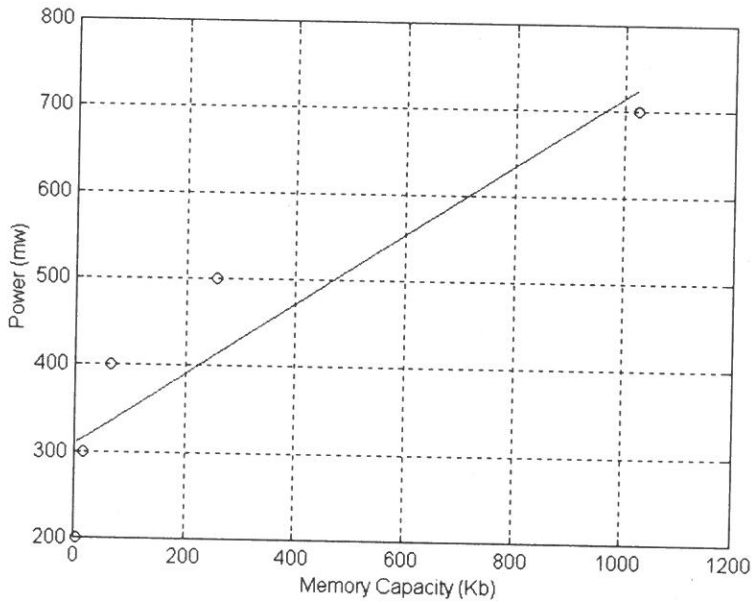


Fig. 5. Power consumption with respect to the size of the RAM chips.

where K is a proportion constant and L is the length of queue

$$POWER_{sleeping} \leq 0.25 \cdot POWER_{busy}$$

We then input equation (13) $POWER_{waiting}$ into equation (11), and we obtain

$$POWER_{average} = \frac{1}{1+\varepsilon+\rho\varepsilon/2} \cdot \left(POWER_{sleeping} + K \cdot \frac{\rho}{\varepsilon}\right) + \frac{\varepsilon}{1+\varepsilon+\rho\varepsilon/2} \cdot POWER_{busy} + \frac{\rho\varepsilon/2}{1+\varepsilon+\rho\varepsilon/2} POWER_{congested} \quad (14)$$

$$\varepsilon = \frac{\rho}{2 \cdot (D_{latency} - W_s)}$$

$$\rho = \frac{\lambda}{\mu}$$

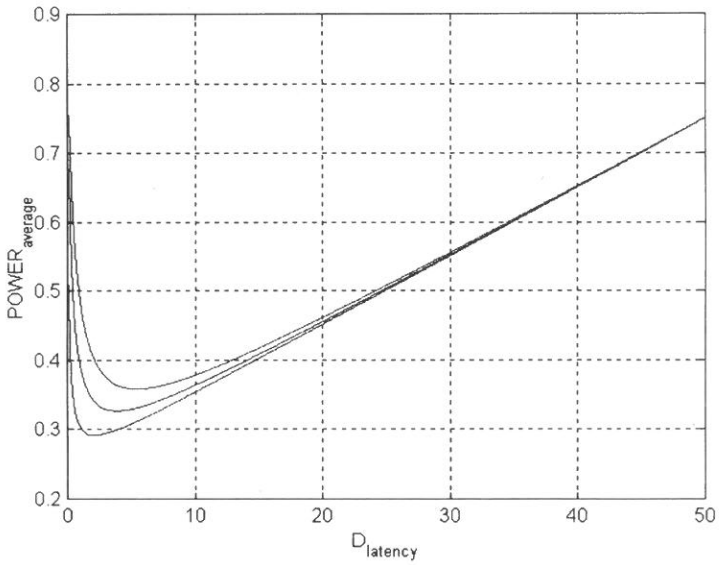


Fig. 6. Average power consumption with respect to the latency $D_{latency}$.
($K = 0.005$, $POWER_{sleeping} = 0.25$, $POWER_{busy} = 0.75$, $POWER_{congested} = 1$, $\rho = 0.167, 0.5, 0.875$)

We plot the equation (14) $POWER_{average}$ with respect to latency $D_{latency}$ shown in Fig.6 and with respect to latency $D_{latency}$ and utilization ρ shown in Fig.7.

From Fig.6 and Fig.7, we see that the dual processor system can reduce the average power consumption to 30% in an ideal situation, where we can tolerate three to five times of latency based on a homogeneous arrival information. However, when the latency $D_{latency}$ increases, mobile users may feel the system is sluggish, because, due to buffering, the received information will be stored in information queues longer before it is processed. Fortunately, the application we mention here is multimedia information from computer networks and a user could observe a piece of information instead of just seeing fragments of a picture. So this power management policy should be acceptable to mobile users, especially since it can improve the power efficiency.

THE SIMULATION RESULTS OF THE TRI-STATE POWER MANAGEMENT MODEL

According to the previous discussion, we implement a simulation module , which

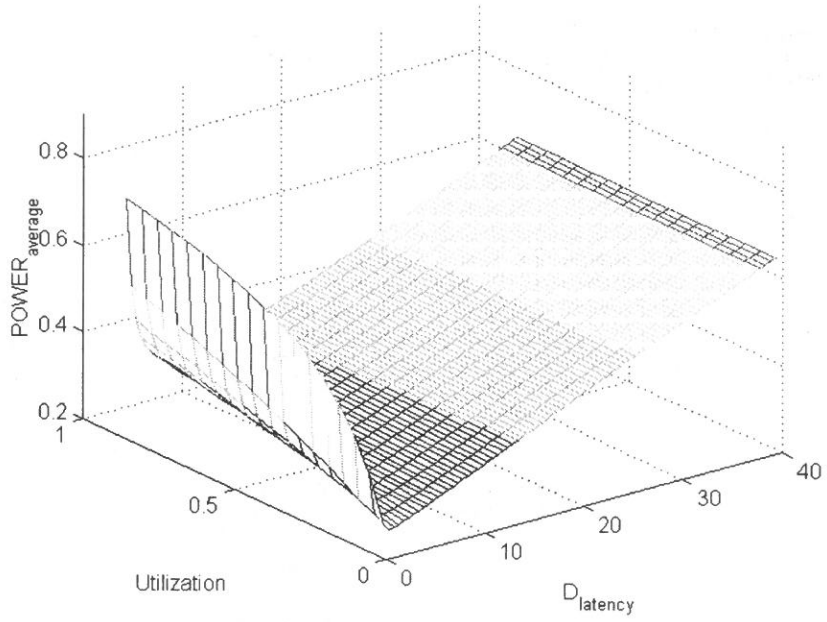


Fig. 7. Average power consumption with respect to the latency $D_{latency}$ and utilization ρ.
($K = 0.005$, $POWER_{sleeping} = 0.25$, $POWER_{busy} = 0.75$, $POWER_{congested} = 1$)

follows the flowchart of the tri-state power management model at Fig.8. The communication subsystem will receive the multimedia packets from the multimedia server, and store them in the buffer temporarily. The OS will monitor the buffer subsystem, and retrieve the buffer status to handle system states, "idle state", "busy state", and "congestion state". While the buffer subsystem is empty or the length of the buffer does not reach the preset "level 1", the system will be in the "idle state". While it reaches the preset "level 1" but does not reach preset "level 2", the system will be in the "busy state". The machine will be in the "congestion state" when the processors can not handle the arrival packets fast enough.

Using the tri-state power management model, with typical multimedia information arrival and service patterns, such as Poisson and deterministic service probability distribution, we simulate the characteristics for the event driving mechanism in a modified $M/M/2/N$ queueing model. The simulation results show that when the state transition is decreased , the average power consumption is

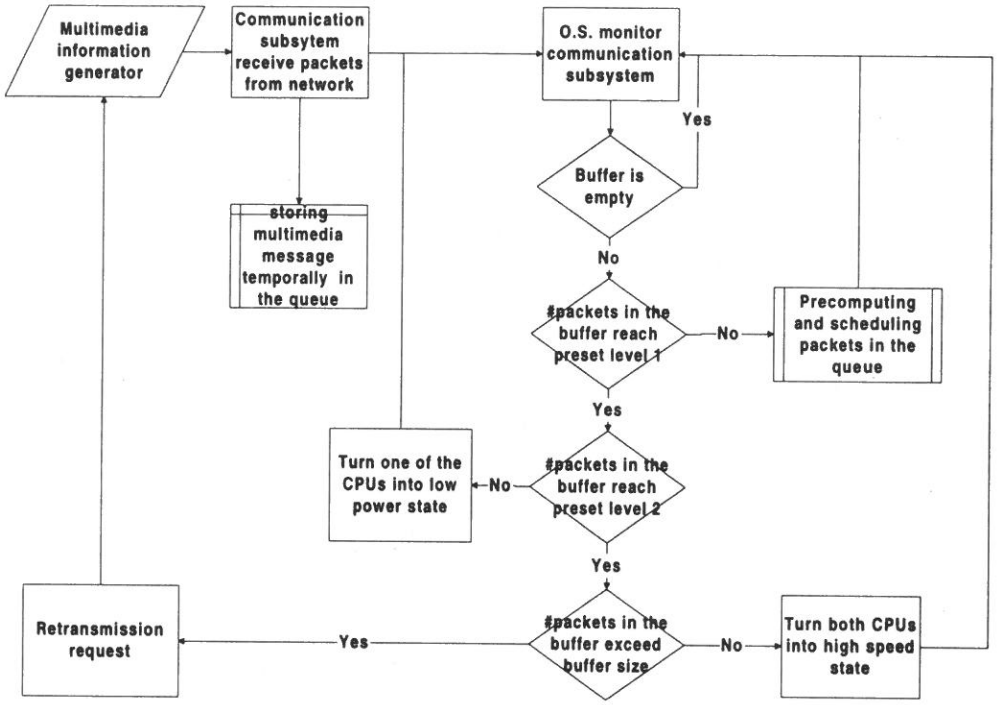


Fig. 8. The simulation flowchart of the tri-state power management model.

decreased. In addition, the power consumption will be increased when users excessively increase the length of queue.

In the simulation result, as shown in Fig.9, its curves are similar with the theoretical results in Eq. (14). The performance of the latency represents the time spent for information accumulation which is usually in proportion with the queue length in Eq. (14). Therefore if we can tolerate the latency within certain ranges, we can potentially save around 65% of the total power consumption with providing the utilization $\rho = 0.5$. In Fig.9, we can also obtain the best power efficiency at five times of the latency to see multimedia information from computer networks.

CONCLUSION

From the state transition diagram of the typical palmtop terminal operation model, we see that a desktop PC consumes almost the same power at each machine

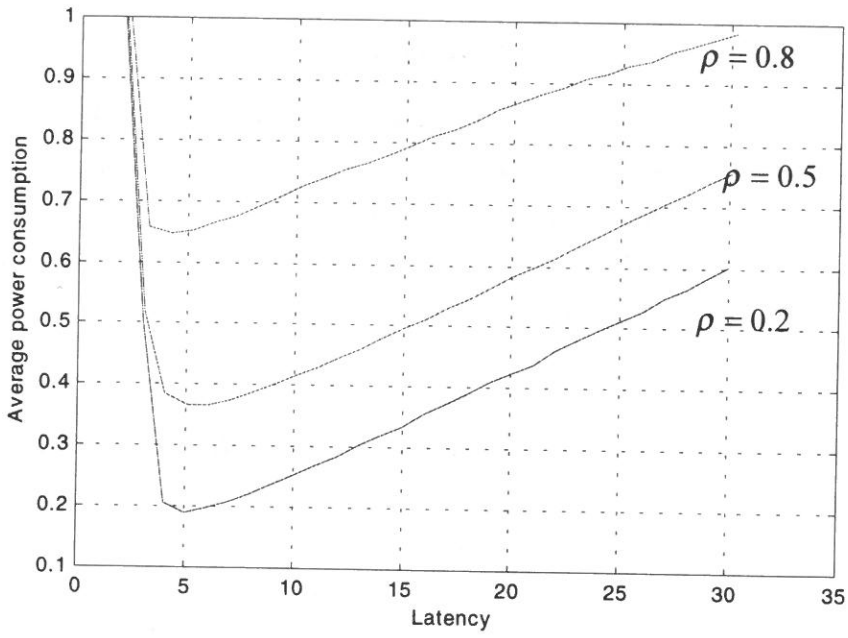


Fig. 9. The relationship between power consumption and latency.

state, even the idle state. This is the reason that a desktop PC power consumption policy is not good for a palmtop multimedia terminal. Instead of wasting power during each machine state, we propose a tri-state power management method which provides low power states with event-driving to reduce state switching frequency by buffering the arriving information.

The tri-state power management method can reduce the average power consumption based on reducing the busy state and the congested state probability with certain system parameters. Through the machine state control mechanism, the machine can stay in a waiting state and turn off some subsystems and then switch itself into a low power state. From both the analysis and the simulation of the modified M/M/2 queueing model, this paper shows that there is a potential to save 70% of the average power consumption for a dual processor system by providing a latency of five times to see multimedia information from computer networks.

References

(1) T. H. Meng, B. M. Gordon, E. K. Tsern and A. C. Hung, "Portable Video-on Demand in Wireless Communication," Proceedings of the IEEE, Vol. 83, No. 4 PP. 659-681, April (1995).

(2) J. M. C. Stork, "Technology Leverage for Ultra-Low Power Information Systems," Proceedings of the IEEE, Vol. 83, No. 4 PP. 607-617, April (1995).

(3) Intel Corporation, Microsoft Corporation and Toshiba Coporation, "Advanced Configuration and Power Interface Specification," Revision 1. 0b, February 8, 1999.

(4) R. Rebhan and S. Olsson, "Multimedia Goes Mobile in Broadcast Networks," IEEE Multimedia, PP. 14-22, April-June 1997.

(5) J. Monteriro, S. Devadas, P. Ashar and A. Mauskar, "Scheduling Techniques to Enable Power Management," Design Automation Conference 1996. IEEE, PP. 349-352, 1996.

(6) J. M. Rulnick and N. Bambos, "Mobile Power Management for Maximum Battery Life in Wireless Communication Network," Proceedings IEEE INFOCOM V2, PP. 443-450, 1996.

(7) J. R. Lorch and A. J. Smith, "Software Strategies for Portable Computer Energy Management," IEEE Personal Communications, PP. 60-73, June 1998.

(8) S. Shigematsu, S. Mutoh and Y. Matsuya, "Power Management Technique for 1-V LSIs Using Embedded Processor," IEEE 1996 Custom Integrated Circuits Conference, PP. 111-114.

(9) H. -P. Juan, V. Chaiyakul and D. D. Gajski, "Condition Graphs for High-Quality Behavioral Synthesis," ACM, PP. 170-174, 1994.

(10) Y. -W. Bai, "Semi-batch Power Management for a Palmtop Multimedia Terminal," Fu-Jen Studies Science and Engineering, ISSN 1028-5679, PP. 19-31, 1998.

(11) K. Itoh, K. Sasaki, and Y. Nakagome, "Trends in Low-Power RAM Circuits

Technologies," Proceedings of the IEEE, VOL. 83, NO. 4, PP. 524-543, April 1995.

(12) E. P. Harris, S. W. Depp, W. E. Pence, S. Kirkpatrick, M. Sri-Jayantha, and Ronald R. Troutman, "Technology Directions for Portable Computers," Proceedings of the IEEE, VOL. 83, NO. 4 PP. 636-657, April 1995.

received September 10, 1999
revised October 8, 1999
accepted December 1, 1999

雙重處理器系統三態電源管理方法之修正 M/M/2 模型分析

胡中仁　白英文
輔仁大學電子工程系

摘　要

雙重處理器系統三態電源管理方法可用來控制行動手機狀態轉移，手機三態是指「等待」、「忙碌」和「擁擠」三個狀態。當手機處於「等待」狀態時，只有開啓一個處理器並且處於低速度低功率操作模式，等待累積從網路抵達手機之資訊，當資訊量累積至預設值時，手機從「等待」狀態轉移至「忙碌」狀態進行處理所累積的資訊，「忙碌」狀態是指開啓一個處理器而且處於高速度高功率操作模式。再者當猝發資訊從網路抵達手機或高計算負擔需求產生時，手機則可由「忙碌」狀態轉移至「擁擠」狀態，此時兩個處理器全被啓動提供最高速資訊處理能力，同時手機也處於消耗最高功率之狀態。根據三態電源管理方法，以修正 M/M/2 排隊模型之靜態非時變狀態轉移圖分析，本文提出雙重處理器系統電源效率與資訊處理遲延關係，經由分析與模擬結果顯示，在網路資訊均勻抵達情形下，如果可以接受五倍資訊處理遲延時間，則行動手機有潛力節省 70％功率消耗。

關鍵詞： 電源管理，行動手機，雙重處理器系統，狀態機率

A Graph-Based Approach to Discovering Multiple-Level Association Rules from Large Databases

Show-Jane Yen

Department of Computer Science and Information Engineering

Fu Jen Catholic University

Taipei, Taiwan, R.O.C.

Abstract

Mining association rules is an important task for knowledge discovery. We can analyze past transaction data to discover customer behaviors such that the quality of business decision can be improved. Various types of association rules may exist in a large database of customer transactions. The strategy of mining association rules focuses on discovering large itemsets which are groups of items which appear together in a sufficient number of transactions.

In this paper, we propose a graph-based approach to generate various types of association rules from a large database of customer transactions. This approach is to construct an *association graph* to indicate the associations between items, and then traverse the graph to generate *large itemsets*. Empirical evaluations show that our algorithm outperforms other algorithms which need to make multiple passes over the database.

Key Words: Data Mining, Knowledge Discovery, Association Rules, Association Graph.

INREOSUXRION

From a large amount of data, potentially useful information may be discovered. Techniques have been proposed to find knowledge (or rules) from databases [1, 4, 11,

13]. The knowledge discovered can be used to answer cooperative queries [5], handle null values [10] and facilitate semantic query optimization [9, 14].

Data mining has also high applicability in retail industry. The effective management of business is significantly dependent on the quality of its decision making. It is therefore important to improve the quality of business decision by analyzing past transaction data to discover customer purchasing behaviors. In order to support this analysis, a sufficient amount of transactions needs to be collected and stored in a database. A transaction in the database typically consists of customer identifier, transaction date (or transaction time) and the items purchased in the transaction. Because the amount of these transaction data can be very large, an efficient algorithm needs to be designed for discovering useful information.

An *association rule* describes the associations among items in which when some items are purchased in a transaction, the others are purchased too. In order to find association rules, we need to discover all *large itemsets* (*or association patterns*) from a large database of customer transactions. A large itemset is a set of items which appears often enough within the same transactions, that is, an itemset that is contained in a number of transactions above a certain minimum threshold.

The following definitions are adopted from [2]. A transaction t *supports* an item x if x is in t. A transaction t supports an itemset X if t supports every item in X. The support for an itemset is defined as the ratio of the total number of transactions which support this itemset to the total number of transactions in the database. To make the discussion easier, occasionally, we also let the total number of transactions which support the itemset denote the support for the itemset. Hence, a large itemset is an itemset whose support is no less than a certain user-specified *minimum support*. An itemset of length k is called a k-itemset and a large itemset of length k a large k-itemset.

After discovering all large itemsets, the association rules can be generated as follows: If the large itemset $Y = I_1 I_2 \cdots I_k$, $k \geq 2$, all rules that reference items from the set I_1, I_2, $\cdots$, I_k can be generated. The antecedent of each of these rules is a subset X of Y, and the consequent Y-X. The confidence of $X \Rightarrow Y$-X in database D is the probability that when itemset X occurs in a transaction in D, itemset Y-X also occurs

in the same transaction. That is, the ratio of the support for itemset Y to the support for itemset X. A generated rule is an association rule if its confidence achieves a certain user-specified *minimum confidence*. Assuming {coffee, sugar, milk} is a large itemset, an example of an association rule is "95% of the transactions in which coffee and sugar are purchased, milk is purchased too." This rule can be specified as "coffee, sugar ⇒ milk 95%." The antecedent of this rule consists of coffee and sugar, and the consequent consists of milk alone. The percentage 95% is the confidence of the rule.

Another type of association rules is called *multiple-level association rules* [6]. The multiple-level association rules are discovered from a large database of customer transactions in which all items are described by a set of relevant attributes. Each attribute represents a certain concept, and these relevant attributes form a set of multiple-level concepts. The concept level for an attribute is defined by domain experts. For example, food items can be described by the relevant attributes "category", "content" and "brand," and attribute "category" represents the first-level concept (i.e., the highest level concept), attribute "content" for the second-level concept and attribute "brand" for the third-level concept. There is a set of domain values for an attribute. Each item in the database contains a domain value for each relevant attribute. For example, if the "category", "content" and "brand" of an item have the domain values "bread," "wheat" and "Wonder", respectively, then this item is described as "bread wheat Wonder" in the database.

From the items in the database, we can derive other items at different concept levels. The domain values of the attribute at the first (i.e., the highest) concept level are the items at the first concept level. An item at the kth concept level can be formed by combining a domain value of the attribute at the kth concept level with an item at the $(k-1)$th concept level.

Hence, item "bread" is at the first concept level, item "bread wheat" at the second concept level, and item "bread wheat Wonder" at the third concept level. For an item, the items at the corresponding higher (lower) concept level of this item are more general (specific) than this item. For example, items "bread" and "bread wheat" are the items at the corresponding higher level of item "bread wheat Wonder," and items "bread wheat Wonder" and "bread wheat" are the items at the corresponding

lower level items of item "bread."

In general, items at higher concept levels have a larger support than those of items at lower concept levels. If we want to find associations among items at relatively low concept levels, many uninteresting associations among items at higher concept levels may also be generated. Hence, the minimum supports specified at higher concept levels should be larger than the minimum supports specified at lower concept levels. A multiple-level association rule [6] is an association rule which describes the associations among items at the same concept level. For each concept level, both minimum support and minimum confidence are specified.

A *multiple-level association pattern* is a large itemset in which all items are at the same concept level. For an item I to be in a multiple-level association pattern, the items at the corresponding higher concept levels of the item I need to be large at their corresponding concept levels. This is to avoid the generation of many meaningless combinations formed by the items at the corresponding lower concept level of the non-large items. For example, if "bread" is not a large item, item "bread wheat" which is at the corresponding lower concept level of item "bread" need not be further examined.

We propose a data mining framework to discover various types of association rules:

1. Numbering phase: in this phase, all items are assigned an integer number.
2. Large item generation phase: this phase generates large items and records related information.
3. Association graph construction phase: this phase constructs an association graph to indicate the associations between large items.
4. Association pattern generation phase: this phase generates all association patterns by traversing the constructed association graph.
5. Association rule generation phase: the association rules can be generated directly according to the corresponding association patterns.

This paper focuses on the association pattern generation, because after generating the association patterns, the association rules can be generated from the corresponding association patterns. In this paper, we propose two algorithms for generating association patterns and multiple-level association patterns.

The rest of this paper is organized as follows: Section 2 describes the related work. The algorithms to discovering association patterns and multiple-level association patterns are presented in Section 3 and 4, respectively. Section 5 evaluates the performance of our data mining algorithms. Finally, we conclude this paper and present directions for future research in Section 6.

RELATED WORK

Various algorithms [2, 3, 7] have been proposed to generate all large itemsets from a large amount of transaction data. These algorithms generate candidate k-itemsets for large k-itemsets, scan each transaction in a database to count the supports for these candidate k-itemsets and find all large k-itemsets in the kth iteration based on a pre-determined minimum support. However, because the size of the database can be very large, it is very costly to repeatedly scan the database to count supports for candidate itemsets. Park, Chen and Yu [8] pointed out that the key issue to improve the performance of large itemsets discovery is the initial candidate set generation, especially for the candidate 2-itemsets, and the amount of data that has to be scanned for large itemset generation in each iteration. They utilized hash method to reduce the number of the generated candidate 2-itemsets and employed pruning techniques to progressively trim the transaction database. The pruning techniques are described as follows: an item in a transaction can be trimmed if it does not appear in at least k candidate k-itemsets in the kth iteration. However, in order to reduce the number of candidate 2-itemsets, the overhead for building the hash table is large. Moreover, in order to trim the database, it is necessary to scan each transaction in the database to determine which items in the transaction can be trimmed.

Yen and Chen [12] proposed an efficient approach to generate large itemsets. This algorithm constructs an association graph to indicate the associations between every two large items, and then traverses the graph to generate all large itemsets. This algorithm need not repeatedly scan the database to count supports for the candidates, and need to scan the database only once, which has been demonstrated to have a better performance than other approaches [2, 3, 7, 8].

The problem of mining multiple-level association rules was introduced by Han and

Fu [6], in which four similar algorithms are presented. They apply the same method as [3] on the transactions for each concept level (from the highest concept level to the lowest concept level) to generate multiple-level association patterns, and reduce the size of the database after discovering multiple-level association patterns at a concept level. This approach has the same drawback as the approach presented in [3], that is, for each concept level, many candidate itemsets need to be counted and a large number of transactions needs to be repeatedly scanned.

MINING ASSOCIATION RULES

An algorithm APG (Association Pattern Generation) is presented to generate association patterns (or largeitemsets). In the following, we describe the four phases discussed in Section 1 for the algorithm APG.

1. Association graph construction

In the numbering phase, algorithm APG arbitrarily assigns each item a unique integer number. Suppose item i represents the item whose item number is i. In the large item generation phase, algorithm APG scans the database and builds a bit vector for each item. The length of each bit vector is the number of transactions in the database. If an item appears in the ith transaction, the ith bit of the bit vector associated with this item is set to 1. Otherwise, the ith bit of the bit vector is set to 0. The bit vector associated with item i is denoted as BV_i. The number of 1's in BV_i is equal to the number of transactions which support the item i, that is, the support for the item i.

Example 1: consider the database TDB1 in Table 1. Each record is a <TID, Itemset > pair, where TID is the identifier of the corresponding transaction, and Itemset records the items purchased in the transaction. Assume that the minimum support $\Im$ is 50% (i.e., 2 transactions).

After the numbering phase, the numbers of the items A, B, C, D and E are 1, 2, 3, 4 and 5, respectively. In the large item generation phase, the large items found in the database TDB1 are items 1, 2, 3 and 5, and BV_1, BV_2, BV_3 and BV_5 are (1010), (0111), (1110) and (0111), respectively.

Table 1. A database TDB1 of transactions.

TID	Itemset
100	C A D
200	E C B
300	A B C E
400	E B

Property 3.1: The support for the itemset i_1, i_2, $\cdots$, i_k is the number of 1's in $BV_{i1} \wedge BV_{i2} \wedge \cdots \wedge BV_{ik}$, where the notation "$\wedge$" is a logical AND operation.

In the association graph construction phase, APG constructs an association graph to indicate the associations between large items. For every two large items i and j ($i < j$), if the number of 1's in $BV_i \wedge BV_j$ achieves the user-specified minimum support, a directed edge from item i to item j is created. Also, itemset (i, j) is a large 2-itemset. Note that an ordered list notation is used to indicate the order of the items in an itemset for the following discussion.

The algorithm AGC (Association Graph Construction) is proposed to construct an association graph, which is described as follows:

every two large items i and j ($i < j$) **do**
begin
if (the number of 1's in $BV_i \wedge BV_j$) $\geq$minimum-support **then**
begin
create a directed edge from item i to item j
generate a large 2-itemset (i, j)
end
end

The association graph for the Example 1 is shown in Fig.1, and the large 2-itemsets are (1, 3), (2, 3), (2, 5) and (3, 5).

The large k-itemsets are generated by traversing the association graph constructed in the association graph construction phase. The data structure used to implement the association graph is a linked list.

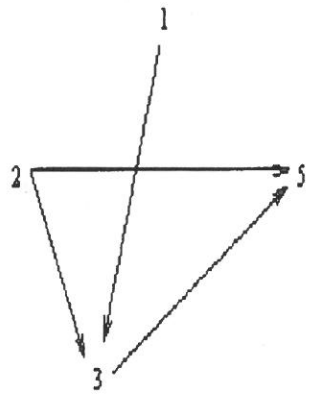


Fig. 1. The association graph for Example 1.

2. Association pattern generation

The large 2-itemsets are generated after the association graph construction phase. In the association pattern generation phase, the algorithm LGDE (Large itemset Generation by Direct Extension) is proposed to generate large k-itemsets ($k > 2$). For each large k-itemset ($k \geq 2$), the last item of the k-itemset is used to extend the large itemset into $k+1$-itemsets.

Lemma 3.1: If an itemset is not a large itemset, then any itemset which contains this itemset cannot be a large itemset.

Rationale: Because the itemset is not a large itemset, the support for the itemset is less than the minimum support. Hence, the support for an itemset which contains this itemset must be also less than the minimum support.

Lemma 3.2: For a large itemset $(i_1, i_2, \cdots, i_k)$, if there is no directed edge from item i_k to an item v, then itemset $(i_1, \cdots, i_k, v)$ cannot be a large itemset.

Rationale: Because there is no directed edge from item i_k to an item v, the itemset (i_k, v) is not a large 2-itemset. Hence, by Lemma 3.1, itemset $(i_1, \cdots, i_k, v)$ is not a large itemset.

Suppose $(i_1, i_2, \cdots, i_k)$ is a large k-itemset. If there is no directed edge from item

i_k to an item v, then the itemset need not be extended into $k+1$-itemset, because $(i_1, \cdots, i_k, v)$ must not be a large itemset according to Lemma 3.2. If there is a directed edge from item i_k to an item u, then the itemset $(i_1, i_2, \cdots, i_k)$ is extended into $k+1$-itemset $(i_1, i_2, \cdots, i_k, u)$. The itemset $(i_1, i_2, \cdots, i_k, u)$ is a large $k+1$-itemset if the number of 1's in $BV_{i1} \wedge BV_{i2} \wedge \cdots \wedge BV_{ik} \wedge BV_u$ achieves the minimum support. If no large k-itemsets can be generated, the algorithm LGDE terminates. The algorithm LGDE is described as follows:

```
k is set to 2.
there are large k-itemsets generated do
  begin
  foreach large k-itemset do
    begin
    if the large k-itemset can be extended based on the association graph then
      begin
      extend the large k-itemset into k + 1-itemsets.
      foreach extended k + 1-itemset do
        begin
            apply the logical AND operations on the bit vectors associated with
              the k + 1 items contained in the k + 1-itemset.
            if (the number of 1's in the resultant bit vector ≥ minimum-support
            then
            generate the k + 1-itemset which is a large k + 1-itemset.
        end
      end
    end
  increment k by 1
  end
replace the item numbers in each large k-itemset (k ≥ 1) by its item names, and
generate all association patterns.
```

For example, consider Example 1. After the association graph construction phase, the large 2-itemsets (1, 3), (2, 3), (2, 5), (3, 5) are generated. For the large 2-itemset (2, 3), there is a directed edge from the last item 3 of the itemset (2, 3) to item 5 in the association graph shown in Fig.1. Hence, the 2-itemset (2, 3) is extended into 3-itemset (2, 3, 5). The number of 1's in $BV_2 \wedge BV_3 \wedge BV_5$ (i.e., (0110)) is 2. Hence, the 3-itemset (2, 3, 5) is a large 3-itemset, since the number of 1's in its bit vector is no less than the minimum support threshold. The LGDE algorithm terminates because no large 4-itemsets can be further generated. After completing the APG algorithm, the generated association patterns are {A, C}, {B, C}, {B, E}, {C, E} and {B, C, E}.

MINING MULTIPLE-LEVEL ASSOCIATION RULES

Mining multiple-level association rules is to find associations at multiple concept levels. For each concept level, both minimum support and minimum confidence are specified.

A multiple-level association pattern is a large itemset in which all items are at the same concept level. For an item to be in a multiple-level association pattern, the items at the corresponding higher concept levels of this item need to be large at their corresponding concept levels. In the following, we present the algorithm MLAPG (Multiple-Level Association Pattern Generation) to generate all multiple-level association patterns.

1. Mining algorithm

In Section 1, we have mentioned that each item in the database contains domain values of the relevant attributes for the problem of mining multiple-level association rules. For an attribute, each domain value is arbitrarily given a unique number in the numbering phase. Besides, each item in a transaction is numbered according to its domain values.

After the numbering phase, MLAPG performs the remaining three phases for each concept level (from the highest concept level to the lowest concept level), which are similar to the associated three phases in APG algorithm. The large item generation

phase scans the database once to build a bit vector for each item and find large items at the current concept level. Also, the size of the database is progressively trimmed by eliminating the items which are not large items at the previous concept level. Notice that for the last concept level, the database need not be further trimmed. The association graph construction phase constructs an association graph by applying AGC algorithm for the large items found at the current concept level. The association pattern generation phase performs algorithm LGDE on the constructed association graph to generate all large itemsets at the current concept level.

2. An example

Example 2: Consider the database TDB2 in Table 2 in which there are three concept levels defined, and the items at each concept level are numbered. For example, in Table 2, item "121" can be the item "bread wheat Wonder", where the first number "1" represents the domain value "bread" of the attribute "category" at level-1, the second number "2" for the domain value "wheat" of the attribute "content " at level-2 and the third number "1" for the domain value "Wonder" of the attribute "brand" at level-3. Assume that the minimum supports $\mathfrak{I}_1$, $\mathfrak{I}_2$ and $\mathfrak{I}_3$ specified at the levels 1, 2 and 3 are 4, 3 and 3 transactions, respectively.

Because there are three concept levels defined, MLAPG needs to perform the above three phases three times to generate all multiple-level association patterns. For the first level, only the level-1 items in the transactions are considered. After the large item generation phase , the found level - 1 large items are "1**" and "2**" and the

Table 2. A database TDB2 of transactions.

TID	Itemset
100	{111, 121, 211, 221}
200	{111, 211, 222, 323}
300	{112, 122, 221, 411}
400	{111, 121}
500	{111, 122, 211, 221, 413}
600	{211, 323, 524}
700	{323, 411, 524, 713}

associated bit vectors are (1111100) and (1110110), respectively, where the notation " * " represents any item. In the association graph construction phase, the association graph for the level-1 large items is constructed as shown in Fig.2 (a) and the level-1 large 2-itemset (1**, 2**) is generated.

For the second level, MLAPG scans the database TDB2 to build a bit vector for each level-2 item and find level-2 large items in the large item generation phase. The level-2 large items found are 11 * , 12 * , 21 * and 22 * , and the associated bit vectors are (1111100), (1011100), (1100110) and (1110100), respectively. By the way, MLAPG eliminates non-large items at the first level from the database TDB2. The trimmed database is shown in Table 3.

Table 3. The trimmed database of TDB2.

TID	Itemset
100	{111, 121, 211, 221}
200	{111, 211, 222}
300	{112, 122, 221}
400	{111, 121}
500	{111, 122, 211, 221}
600	{211}

In the association graph construction phase, MLAPG performs AGC algorithm for the level-2 large items. The constructed association graph is shown in Fig.2 (b), and the generated level-2 large 2-itemsets are (11 * , 22 *), (11 * , 21 *), (11 * , 12 *), (12 * , 22 *) and (21 * , 22 *). After the association pattern generation phase, the level-2 large 3-itemsets (11 * , 21 * , 22 *) and (11 * , 12 * , 22 *) are generated by performing the algorithm LGDE.

For the last level, after scanning the trimmed database in Table 3, the found level-3 large items are 1 1 1 , 2 1 1 and 2 2 1 , and the associated bit vectors are (110110), (110011) and (101010), respectively. The association graph constructed for the level-3 large items is shown in Fig.2 (c), and there is only one level-3 large 2-itemset (111, 211) generated.

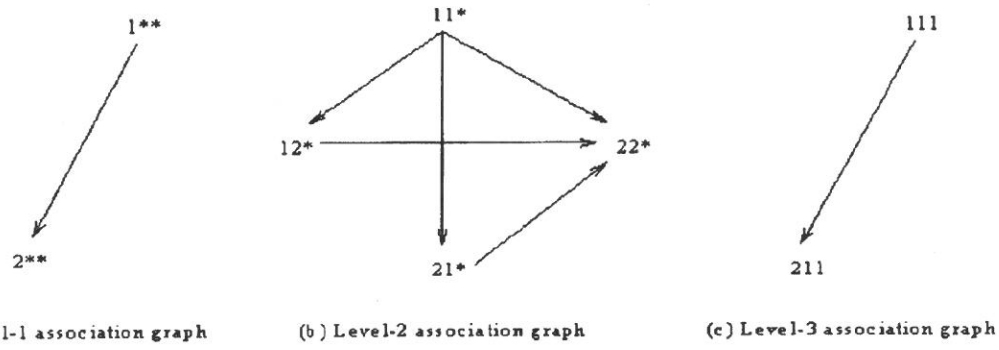


Fig. 2. The association graph for Example 2.

PERFORMANCE EVALUATION

In this section, we evaluate the performance of the two algorithms APG and MLAPG which are proposed to discover the two types of association patterns: association patterns and multiple-level association patterns, respectively. In [12], we have evaluated the performance of APG and demonstrated that APG has a better performance than other approaches [2, 3, 7, 8]. In the following we analyze the performance of algorithm MLAPG.

For MLAPG algorithm, Han and Fu [6] presented four similar algorithms to generate multiple-level association patterns, in which the algorithm ML_T2L1 is used to compare with our algorithm MLAPG.

The synthetic data generation method is the same as the method described in [6]. The parameters used to generate synthetic data are set as follows: the number of transactions is 100,000, the number of items is 1000, the number of potentially large itemsets is 2000, the average size of the transactions is 5 and the average size of the potentially large itemsets is 10. Besides, the number of the concept levels is set to 4 and the fanouts for levels 2, 3 and 4 are set to 5, 5 and 5, respectively. Hence, the number of nodes at level-1 is 8.

Fig.3 shows the relative execution time for MLAPG algorithm and ML_T2L1 algorithm for various database sizes. In this experiment, we set the minimum supports for levels 1 , 2 , 3 and 4 to 50% , 10% , 5% and 2% , respectively , because in this case,

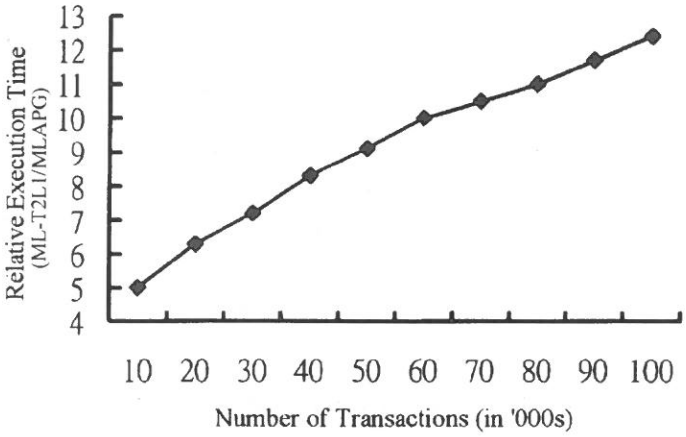


Fig. 3. Threshold (50%, 10%, 5%, 2%).

the algorithm ML_T2L1 has the best performance among the four algorithms presented in [6].

For each concept level, ML_T2L1 applies Apriori algorithm [3] to generate large itemsets at the current concept level. Hence, for the concept level i, ML_T2L1 needs to generate level-i candidate k-itemsets ($k \geq 1$) and scan the database to count the support for each level-i candidate k-itemset in the kth iteration. If there are m concept levels defined and Apriori needs to perform ri iterations to find all level-i large itemsets, then there are database scans needed to generate all large itemsets at each concept level.

For our MLAPG algorithm, it applies APG algorithm for each concept level. Hence, MLAPG needs only one database scans to generate all large itemsets at every concept level. Because MLAPG and ML_T2L1 apply algorithms APG and Apriori, respectively, for each concept level, we analyze the performance of the two algorithms APG and Apriori.

Suppose in the kth iteration, the set L_k of the large k-itemsets is generated. For the first iteration, APG and Apriori both need to scan the database to count the support for each item. By the way, APG builds the bit vector for each item.

In the second iteration, APG performs $\frac{|L_1|(|L_1|-1)}{2}$ logical AND operations on the bit vectors to construct an association graph and generate large 2-itemsets. However, Apriori needs to generate $\frac{|L_1|(|L_1|-1)}{2}$ candidate 2-itemsets. After generating candidate 2-itemsets, Apriori scans the database to combine every two items in each transaction and search these candidate 2-itemsets to count their supports.

In the kth iteration ($k > 2$), APG applies LGDE algorithm to generate large k-itemsets. Suppose the average out-degree of each item in the association graph is q. LGDE performs $(k\text{-}1) \times |L_{k-1}| \times q$ logical AND operations to find all large k-itemsets. However, Apriori needs to generate candidate k-itemsets from large k-1-itemsets. After generating candidate k-itemsets, Apriori scans the database to combine every k items in each transaction and search these candidate k-itemsets to count their supports. Hence, the execution time of Apriori depends on the number of generated candidate itemsets and the amount of data that has to be scanned.

Since for each concept level, the number of logical AND operations performed by MLAPG is much less than the number of the candidate itemsets to be counted and the sizes of the databases to be scanned by ML_T2L1 algorithm, ML_T2L1 takes much more time than MLAPG for each concept level. Fig.3 shows that MLAPG outperforms ML_T2L1 significantly and the performance gap increases as the size of the database increases because the number of candidate itemsets and the number of database scans increase for ML_T2L1 algorithm.

CONCLUSION AND FUTURE WORK

We propose a uniform graph-based approach to discover association rules and multiple-level association rules. The approach includes the five phases: numbering phase, large item generation phase, association graph construction phase, association pattern generation phase and association rule generation phase.

The numbering phase numbers each item in the database. The large item generation phase generates large items and build a bit vector for each large item. The association graph construction phase constructs an association graph to indicate the associations between large items. The association pattern generation phase generates

all association patterns by traversing the constructed association graph. The association rule generation phase generates association rules according to the corresponding association patterns.

We present two algorithms APG and MLAPG to generate association patterns and multiple-level association patterns, respectively. In [12], the algorithm APG has been demonstrated to have a better performance than other approaches [2, 3, 7, 8]. In this paper, we compare MLAPG algorithm to the previously known algorithm ML_T2L1 [6]. The experimental results show that MLAPG outperform ML_T2L1. When the minimum support decreases or the size of the database increases, the performance gap increases because the number of candidate itemsets generated by ML_T2L1 increases and the number of database scans also increases.

For our approach, the related information may not fit in the main memory when the size of the database is very large. In the future, we shall consider this problem by reducing the memory space requirement. Also, we shall apply our approach on different applications, such as document retrieval and resource discovery in the world-wide web environment.

ACKNOWLEDGMENT

This work is partially supported by the Republic of China National Science Council under Contract No. 89-2213-E-030-003. The author wishes to express his thanks for the financial support of the Societas Verrbi Divini.

References

(1) R. Agrawal and et al., "Database Mining: A Performance Perspective", *IEEE Transactions on Knowledge and Data Engineering*, 1993, pp. 914-925.

(2) R. Agrawal and et al., "Mining Association Rules Between Sets of Items in Large Databases", *Proceedings of the ACM SIGMOD*, 1993, pp. 207-216.

(3) R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules", *Proceedings of the International Conference on Very Large Data Bases*, 1994, pp. 487-499.

(4) J. Han and et al.," Data-Driven Discovery of Quantitative Rules in Relational

Databases, *IEEE Transactions on Knowledge and Data Engineering*, 1993, pp. 29-40.

(5) J. Han and et al., "Intelligent Query Answering by Knowledge Discovery Techniques", *IEEE Transactions on Knowledge and Data Engineering*, 1996, pp. 373-390.

(6) J. Han and Y. Fu. Discovery of Multiple-Level Association Rules from Large Databases. In Proceedings of the International Conference on Very Large Data Bases, pages 420-431, 1995.

(7) M. Houtsma and A. Swami, "Set-Oriented Mining for Association Rules in Relational Databases", *Proceedings of the International Conference on Data Engineering*, 1995, pp. 25-33.

(8) J. S. Park and M. S. Chen and P. S. Yu, "An Effective Hash-Based Algorithm for Mining Association Rules", *Proceedings of ACM SIGMOD*, Vol. 24, No. 2, 1995, pp. 175-186.

(9) M. S. E. Sciore and et al., "A Method for Automatic Rule Derivation to Support Semantic Query Optimization", *ACM Transactions on Database Systems*, 1992, pp. 563-600.

(10) S. J. Yen and A. L. P. Chen, "Neighborhood/Conceptual Query Answering with Imprecise/Incomplete Data", *Proceedings of the International Conference on Entity-Relationship Approach*, 1993, pp. 151-162.

(11) S. J. Yen and A. L. P. Chen, "An Efficient Algorithm for Deriving Compact Rules from Databases", *Proceedings of the International Conference on Database Systems for Advanced Applications*, 1995, pp. 364-371.

(12) S. J. Yen and A. L. P. Chen, "An Efficient Approach to Discovery Knowledge from Large Databases", *Proceedings of the IEEE/ACM International Conference on Parallel and Distributed Information Systems*, 1996, pp. 8-18.

(13) S. J. Yen and A. L. P. Chen, "The Analysis of Relationships in Databases for Rule Derivation", *Journal of Intelligent Information Systems*, *Kluwer Academic Publishers*, Vol. 7, No. 3, 1996, pp. 235-260.

(14) C. Yu and W. Sun, " Automatic Knowledge Acquisition and Maintenance for

Semantic Query Optimization", *IEEE Transactions on Knowledge and Data Engineering*, 1989, *pp.* 362-375.

received ***October 6, 1999***
revised ***November 20, 1999***
accepted ***December 3, 1999***

從大型資料庫中發掘多階層關聯法則

顏 秀 珍
輔仁大學資訊工程學系

摘 要

從大型資料庫中發掘多階層關聯法則是資料探勘領域重要的研究方向，我們分析過去的交易資料，以發掘顧客的購物行爲，有助於提升商業管理與決策的品質。發掘關聯法則的重點在於如何找出所有常常“一起”被購買的商品。不同類型的關聯法則存在於顧客的交易資料中。在這篇論文中，我們提出一個有效率的演算法來解決這個問題。此演算法建立一個「關聯圖」來標示商品之間的相關性；然後走訪這個「關聯圖」來產生所有的關聯法則。有別於其它已提出的演算法需多次的掃瞄整個交易資料庫，我們所提出的演算法只需要掃瞄整個交易資料庫一次。而從實驗的數據上也顯示我們的演算法比其它演算法有更好的效率。

關鍵詞：資料探勘，知識發掘，關聯法則，關聯圖

高效能鐵電單晶$(PbMg_{1/3}Nb_{2/3}O_3)_{0.67}$ $(PbTiO_3)_{0.33}$其介電性質的研究

趙鳳強　蘇文仲　杜繼舜
輔仁大學物理學系

摘　要

本論文對弛豫鐵電單晶$(PbMg_{1/3}Nb_{2/3}O_3)_{0.67}(PbTiO_3)_{0.33}$(PMN-33%PT)沿著晶體［001］方向（也就是晶體的C軸），進行介電常數隨著溫度變化的研究。在430 K附近時觀察到一個陡峭的鐵電相變，又在340 K附近時看到一個“瀰散相變（diffuse phase transition)”。在上述兩個溫度點附近，有明顯的熱滯現象，也就是說，上述兩個相變都是屬於一階鐵電相變。從介電常數ε_c^*（f, T）和頻率的關係，證實了在360 K以下晶體系統有一個電偶極的鬆弛過程（relaxation process)。我們同時也計算了此鬆弛過程中的活躍能（*activation energy*），“*Vogel-Fulcher*”溫度，及“*attempt frequency*”。

簡　介

弛豫鐵電材料（relaxor ferroelectric materials）一般是指具有ABO_3形式晶格（unit cell）的相關鐵電系統，在它的晶體內正離子（A或B離子）不是位於一個固定的位置（A或B），使正負電荷成比例出現，而是所有的正離子隨機分佈在任何可能出現正離子的位置。那些隨機分佈的正離子產生了不規則及非電中性的區域性電場，那些局部不規則的電場，使得晶體在隨著溫度變化時，它的相變呈現所謂的“瀰散（diffuse)”現象，而不像一般典型的鐵電材料在相變點附近呈現陡峭的變化。弛豫鐵電材料在相變點附近的物理性質，和正離子的有序程度有關，而有序的程度和長晶的方法、正離子所帶的價數、回溫過程的處理方法（例如是淬火或是退火）

等條件有關。

弛豫鐵電晶體和典型的鐵電晶體鈦酸鉛（$PbTiO_3$：簡稱 PT）相比，其介電常數在相變點顯示出較大範圍緩慢的變化，且和外加場的頻率有關，也就是所謂的“瀰散相變”。Pb（$Mg_{1/3}Nb_{2/3}$）O_3（PMN）是一種早期最引人注目的弛豫鐵電材料，它在 280 K 時，介電常數呈現了“瀰散相變”，而鈦酸鉛單晶，它的鐵電相變是發生在 $T_c = 760$ K，是一個典型尖銳變化的一階鐵電相變。因爲 PMN 和 PT 的單晶混合物（PMN-PT）具有很大的室溫壓電常數，且介電損耗低（≦0.01），[1] 此外，它的晶體生長方法也比 $PbZrO_3$-$PbTiO_3$（PZT）簡單，因此，PMN-PT 型材料已成爲高效能轉換器（transducer）及感應器（sensor）相關應用的明日之星。目前據我們所知，對於 PMN-PT 型材料的介電常數現象的研究還十分缺乏，因此引起了我們去測量 PMN-33%PT 介電常數的動機，經由對介電常數的量測，使我們瞭解此類材料對外加電場的物理反應特性。

實驗步驟

本實驗是用平面板狀的樣品，它的大小是 7.28（長）×6.23（寬）×1.00（厚）mm^3。在實驗過程中，晶體的表面塗上銀膠當作電極，施加的電場是沿著晶體的 [001] 方向，降溫過程和升溫過程的速率是每分鐘 1.5 K，我們使用 Wayne-Kerr Precision Analyzer model PMA3260A（可產生 20 Hz 到 3 MHz 頻率的交流電場）來量測樣品的電容及電阻，量測的電場大小爲 1 V/cm。溫控設備是使用 Janis model CCS-450 封閉循環冷卻器（設計的使用溫度範圍是從 10-475 K）和 Lakeshore model 340 溫度控制器。此實驗的流程如圖 1 所示，在實驗過程，我們量測樣品的電容及電阻值，再由下列方程式 (1)，(2) 和 (3)，求得介電常數的實數部分和虛數部分（ε_c' 和 ε_c''）。

$$C^* = C - i\frac{G}{w} = \frac{\partial Q}{\partial V} = \frac{\partial(A\varepsilon_0\varepsilon^* E)}{\partial V} = \frac{\varepsilon_0 A}{d}\varepsilon^* = C_0\varepsilon^* \tag{1}$$

$$\varepsilon^* = \varepsilon'_{ij} - i\varepsilon_{ij}''; C_0 = \frac{\varepsilon_0 A}{d} \tag{2}$$

$$\varepsilon_{ij}^*(f,T) = \frac{C_{ij}(f,T)}{C_0}; \varepsilon_{ij}''(f,T) = \frac{G_{ij}(f,T)}{2\pi f C_0} \tag{3}$$

這裡，C =電容；G =電導（電阻的倒數）；C_0 =幾何電容；f =量測的頻率。

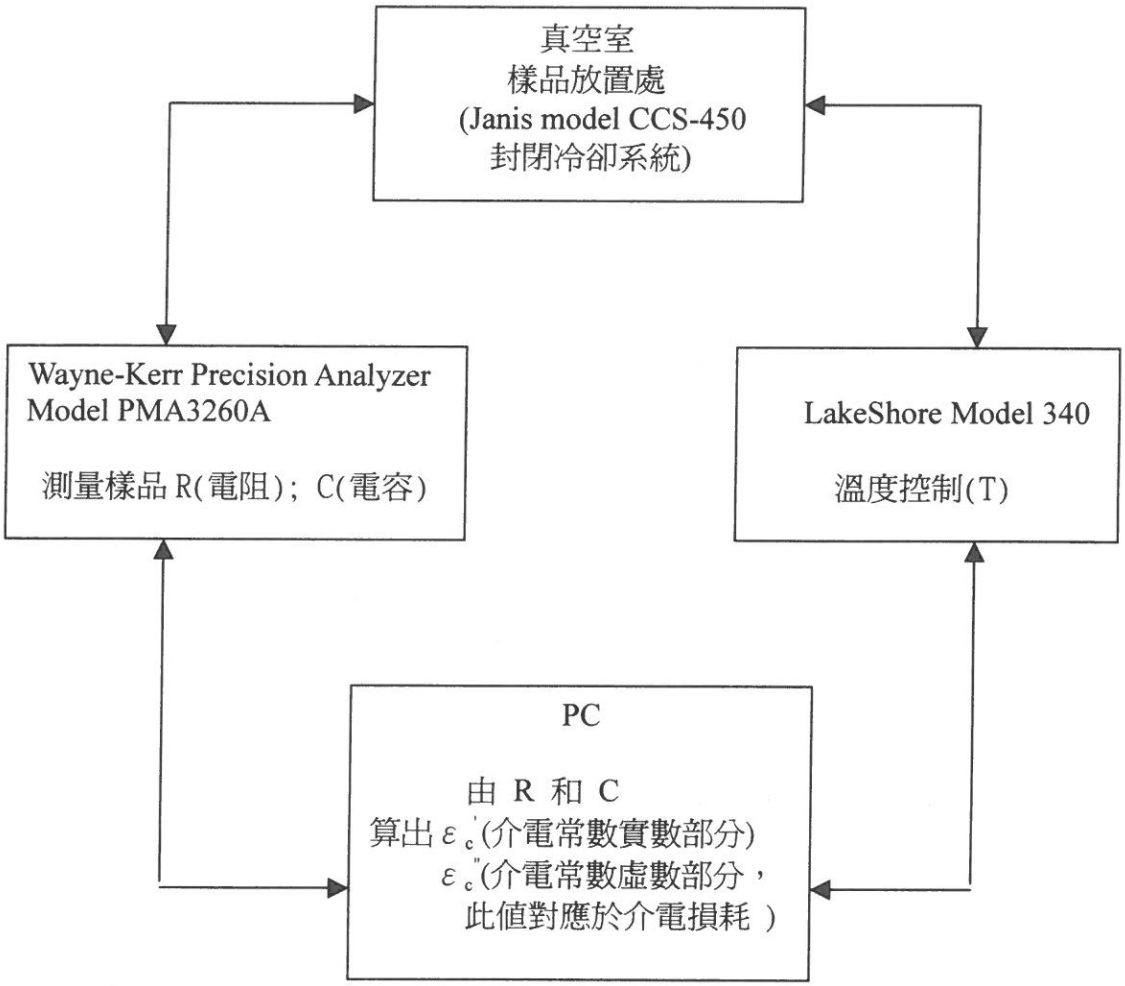


圖 1　介電常數實驗流程示意圖。

結果和討論

圖 2(a)和 2(b)展示 PMN-33%PT 在降溫過程（300 Hz-2.5 MHz），量測的介電常數實部和虛部（ε_c'和ε_c''）隨著溫度變化的情形，下標 *C* 是指沿著晶體晶格的［001］方向。圖 2 中的插圖爲溫度範圍 250 K-450 K 之 ε_c'和ε_c''放大信號對溫度的變化圖形，使它能看的更明顯。和典型的弛豫鐵電單晶 PMN 相比，[2] PMN-33%PT 在 440 K 附近的相變，呈現比較窄的頻率色散現象，而且相變溫度也比PMN ($T_C \approx 210$ K)

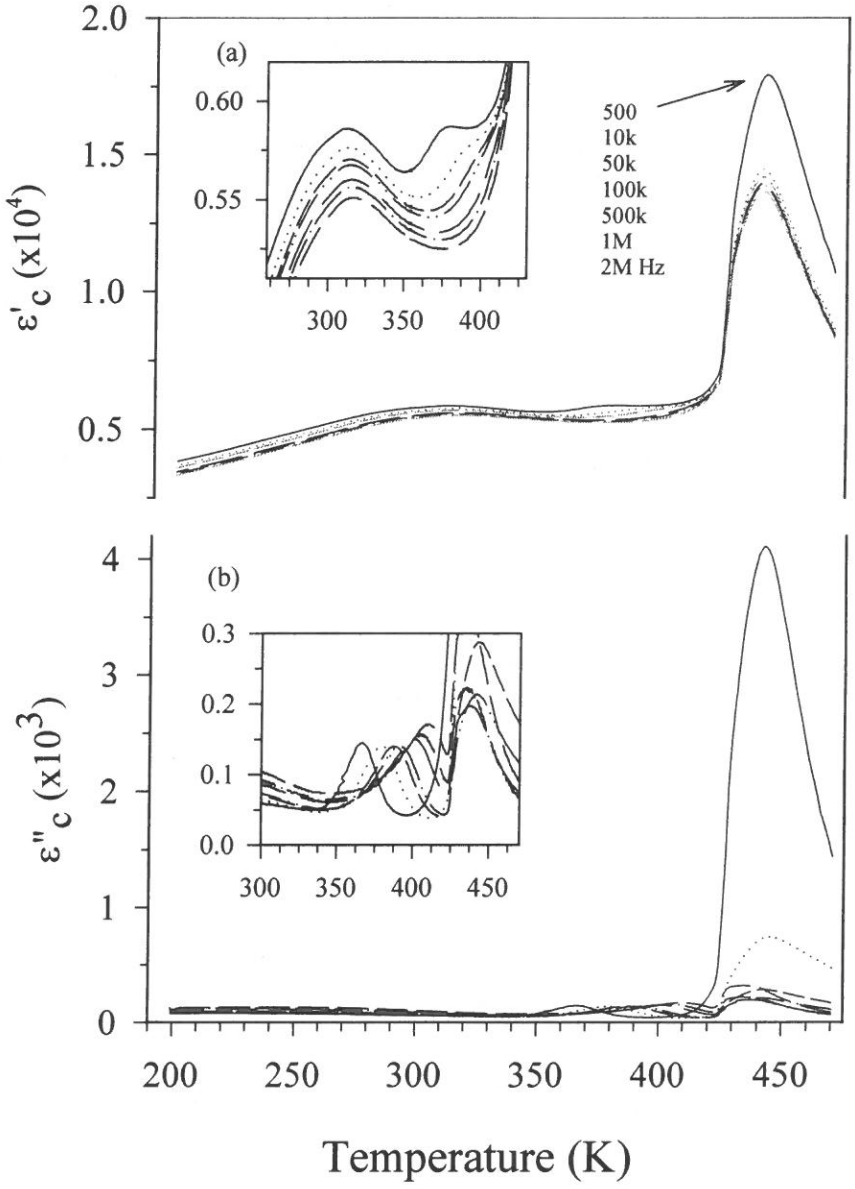


圖 2 PMN-33%PT 在降溫過程中（每分鐘降 1.5 K），由交流電電場頻率從 500 Hz 到 2.5 M Hz，所測得的介電常數對溫度變化的關係圖。(a) 實數部分的介電常數值 ϵ_c' (b) 虛數部分的介電常數值 ϵ_c''。(a) 和 (b) 中的插圖是部分數據的放大。

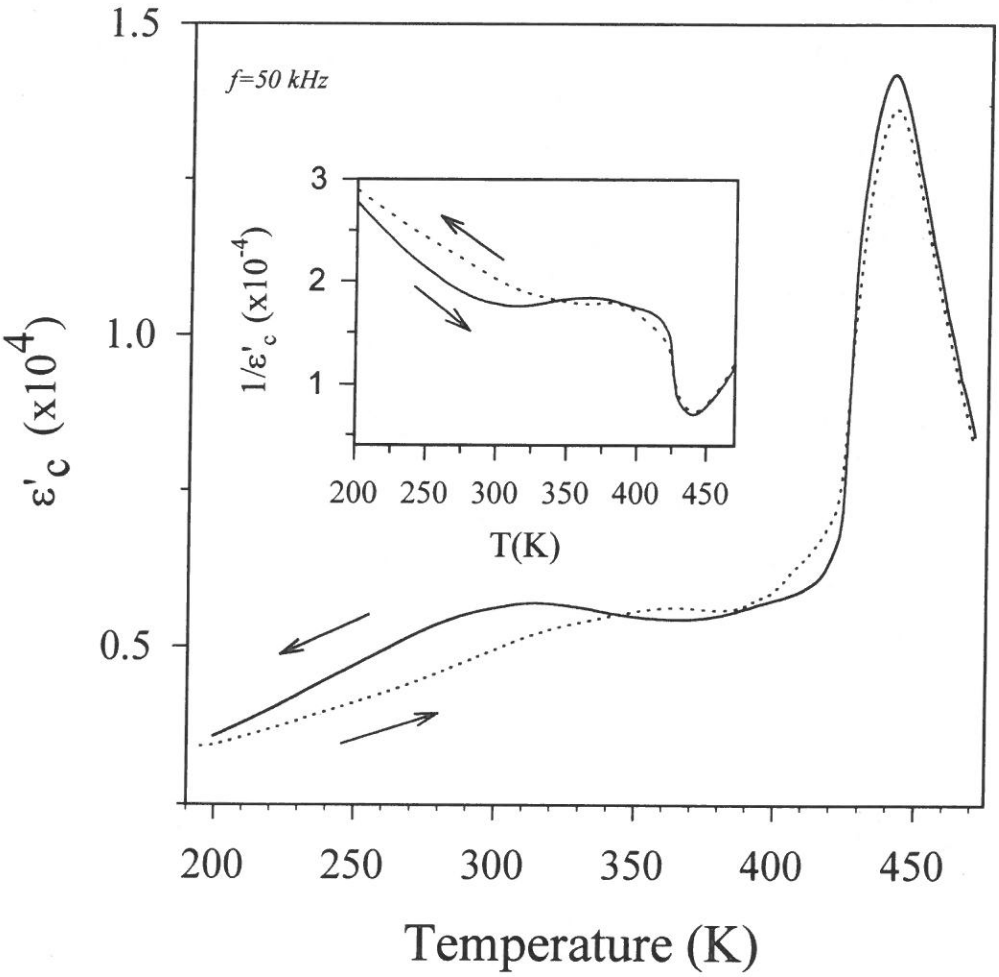


圖 3　由交流電電場頻率 $f = 50$ kHz，所測得的介電常數實部 ε_c' (f, T) 的熱滯曲線圖。插圖部分是 ε_c' (f, T) 的倒數和溫度的關係圖。

高，這是因爲它含有鈦酸鉛的成分。在溫度低於 360 K 時，ε_c'和ε_c''都表現出不規則的頻率色散，圖 3 展現頻率 $f = 50$ kHz 在升溫和降溫過程，ε_c'隨著溫度的變化，一個明顯的熱滯曲線可以在 340 K 到 200 K 附近觀察到，圖 3 的插圖是 ε_c'的倒數和溫度的關係圖，從其中我們可以看出有一個典型的一階鐵電相變（first-order ferroelectric phase transition）發生在 430 K 附近。[3]

PMN-33%PT 在 430 K 附近時的介電異常現象（圖 2-3）的成因是什麼呢？實數

部分的介電常數 ε_c'在 430 K 附近變的很陡，伴隨的頻率色散也很窄，在一般鐵電材料相變中，在相變點 T_C 附近，其介電常數常會出現劇烈的改變，因爲當溫度低於 T_C 時，材料中的電偶極會沿著晶體的極化方向排列，因此我們斷言 PMN-33%PT 單晶在 $T_C \approx 430$ K 時，經歷了一階鐵電相變。[3]

在低溫區域（340 K 以下），介電常數的實數部分 ε_c'，呈現出緩和的改變伴隨著顯著的頻率色散，相對應的介電常數的虛數部分也表現出相關的頻率色散，我們稱這種現象叫“瀰散相變”。這是由於材料中的結構無序造成的。這裡的結構無序，很可能是因爲材料中，同時存在兩種不同的結構，也就是 tetragonal 和 trigonal，因爲此兩種不同的對稱結構，在材料中彼此競爭，導致材料中無法有長程有序（long-range order）的存在，這種無序現象常出現在混合的材料系統中。[1-2] 可以想像爲在材料中有許多宏觀小，微觀大的小區域（cluster），每個小區內的結構和成份是均勻的，但不同小區中原子的排列結構對稱和成分有差異。使得每個小區對於不同頻率的外加電場反應不同，以致不同小區對不同頻率電場的相變溫度也不同，而宏觀物理量（例如介電常數）是許多小區反應的疊加結果。而在溫度 340 K 附近有一個明顯的熱滯曲線，這種現象表示這裡的“瀰散相變”爲一階相變，因爲介電常數的熱滯曲線說明，在這個溫度範圍，系統是處於半穩態（metastable）。[3] 一般用 dP/dT 不連續來區別一階和二階相變的方法，不能用在“瀰散相變”的情形之下。[3] 這裡 dP/dT 指的是電偶極化隨溫度的變化率。

圖 2(a)和圖 2(b)的插圖部分展現出在 410 K 以下，介電常數的實數部分 $\varepsilon_c'(f,T)$ 和虛數部分 $\varepsilon_c''(f,T)$ 均存在著頻率色散，那些色散現象確認了“鬆弛過程”是在 410 K 以下才發生的。由圖 4 所示，我們發現在“鬆弛過程”產生的溫度區間遵循 Vogel-Fulcher 指數方程，[4]

$$f = f_0 e^{E_a/k_B(T_a - T_0)} \tag{4}$$

其中 f 是量測的電流頻率，f_0 是 attempt frequency，E_a 是在短程無序系統中，使不同小區域重新轉向的平均活躍能，k_B 是 Boltzmann 常數，T_0 是 Vogel-Fulcher 溫度，T_g 是介電常數的虛數部分 ε_c'' 達到最大值的溫度。對於 PMN-33%PT，最後得到的結果是，$E_a \approx 199$ meV (2313 K)，$f_0 \approx 7.09 \times 10^{13}$ Hz，$T_0 \approx 277$ K。這些參數有什麼意義？第一，attempt frequency 落在一般晶格振動的頻率範圍之內。第二，E_a 是在部分無序的系統中，使不同小區域重新轉向的平均活躍能。和 PZN-9.5%PT 相比，PMN-33%PT的E_a值比較大，[4]這表示如果要使PMN-33%PT晶體中的小區域重新轉

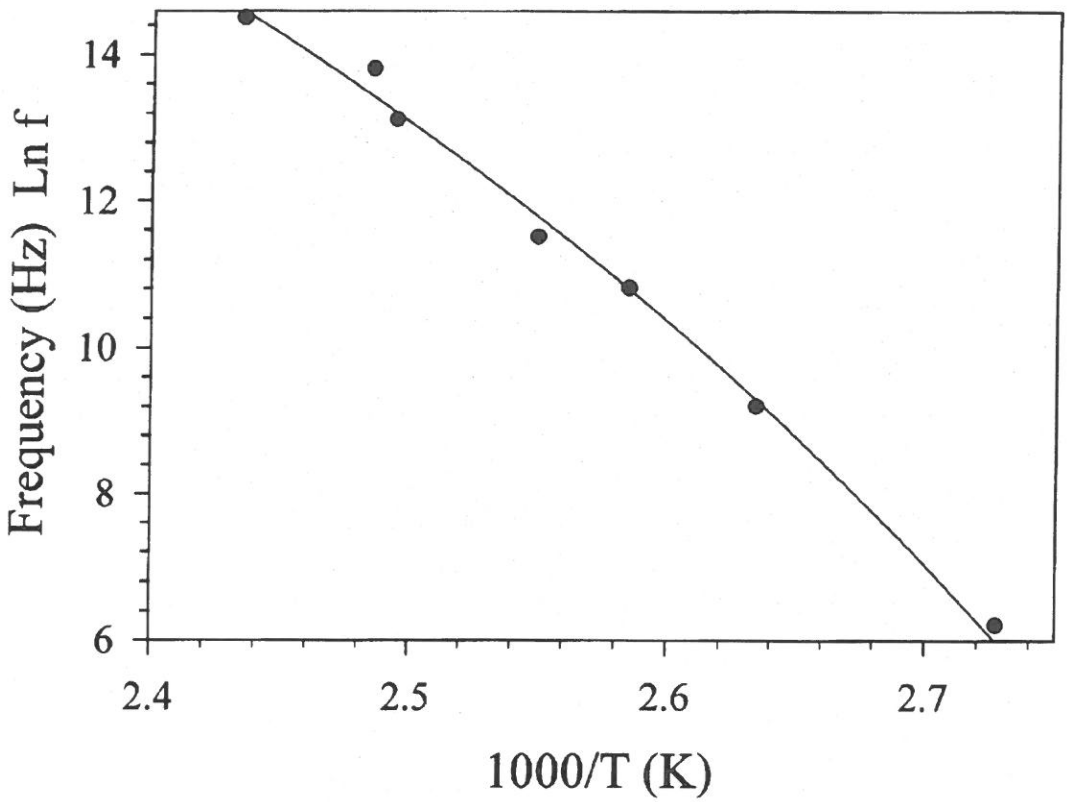


圖 4 ln (f) 和 1000/T 的關係圖。f 是測量的交流電頻率，T 是 ε_c'' 的值爲最大時所處的溫度 (<360 K)。圖中的實線部分是用文章中提到的參數代入方程式 (4) 中所畫出來的。

向，會比使 PZN-9.5%PT 晶體中的小區域重新轉向，須要更多的能量。因此對於溫度的影響，PMN-33%PT 比 PZN-9.5%PT 相對的穩定。第三，T_0 是用外推法算出在此溫度以下所有的小區域重新轉向都會被“凍結 (freezing)”。在很多無序的系統中，Vogel-Fulcher 方程比 Arrhenius 方程，更吻合實驗數據的事實，顯示出當溫度下降時，典型的電偶極重新轉向，是由很多的電偶極集體 (electric dipolar cluster) 參與的。

結　　論

由 PMN-33%PT 的變溫介電常數 (ε_c' 和 ε_c'') 量測，證實了在 430 K 附近和 340

K 附近，分別有典型的鐵電相變及“瀰散相變”。我們由介電常數 $\varepsilon_c'(f,T)$ 的倒數在 410 K 附近時所表現的行爲，和 340 K 附近介電常數的熱滯曲線行爲，確認了這兩種相變是屬於一階相變。由介電常數 $\varepsilon_c''(f,T)$ 和不同頻率的關係，證實了在 360 K 以下，PMN-33%PT 有一個電偶極化的鬆弛過程。經由理論的 fitting，我們同時也計算了鬆弛過程的活躍能，Vogel-Fulcher temperature ，attempt frequency。

誌　　謝

在這裡我們特別感謝輔仁大學中西文化研究中心的研究補助（補助編號：0201-一般-1999-1.0-0140）。

參 考 論 文

(1) C. -S. Tu, F. -C. Chao, C. -H. Yeh, and C. -L. Tsai, Phys. Rev. B 60, 6348 (1999).
(2) C. -S. Tu, V. Hugo Schmidt, and I. G. Siny, J. of Appl. Phys. 78 (9), 5665 (1995).
(3) M. E. Lines and A. M. Glass, Principles and Applications of Ferroelectrics and Related Materials (Oxford Uinversity Press, London, 1977).
(4) E. Courtens, Phys. Rev. Lett. 52, 69 (1983).

88年10月 6日　收稿
88年11月24日　修正
88年12月 2日　接受

$(PbMg_{1/3}Nb_{2/3}O_3)_{0.67}(PbTiO_3)_{0.33}$

F. -C. Chao, W. -S. Su, Chi-Shun Tu

Department of Physics

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

ABSTRACT

Dielectric permittivities along the [001] direction have been measured as a function of temperature for relaxor-ferroelectric single-crystal $(PbMg_{1/3}Nb_{2/3}O_3)_{0.67}$ $(PbTiO_3)_{0.33}$. A sharp ferroelectric phase transition was observed near 430 K. As the temperature decreases, a diffuse phase transition was detected near 340 K. The nature of the thermal hysteresis for the dielectric permittivity confirms that these transitions (near 340 K and 460 K) are both first order. The frequenay-dependent dielectric data $\varepsilon''(f,T)$ prove the existence of an electric dipolar relaxation process below 360K. The activation energy, the Vogel-Fulcher temperature, and attempt frequency corresponding to this relaxation process are also calculated.

Key Words: 介電常數 (dielectric constant), 弛豫鐵電體 (relaxor ferroelectric materials), 瀰散相變 (diffuse phase transition)

Asymptotic Distribution for Estimator of Process Incapability Index C''_{PP} for Second Order Stationary Processes

Sy-Mien Chen
Department of Mathematics
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.

Yu-Sheng Hsu
Department of Mathematics
National Central University
Chungli, Taiwan 320, R.O.C.

Abstract

A major concern about using the process capability index is the assumption of the indepedent and normally distributed properties of the process data. Under such assumptions, Greenwich and Jahr-Schaffrath (1995) introduced an incapability index C_{pp} which provides an uncontaminated separation between information concerning process accuracy and process precision. However, the index C_{pp} inconsistently measures process capability in many cases and thus reflects process potential and performance inaccurately. Chen (1998) proposed an adjusted incapability index, called C''_{pp}, to handle processes with both symmetric and asymmetric tolerances for a normal process. However, in practice, process data are often time dependent and nonnormal. In this paper, the central limit theorem of the estimator $\hat{C}''_{pp}$ for second order stationary processes are derived.

Key Words: process capability index

Introduction

A key issue during a certification stage is the long-term capability of the supplier's processes, which includes the consistency of the process, the ability to deal with process disturbances, and the determination of the closeness to target. For a high-quality product, it is desirable that a very small percentage of items lied outside of the specification limits. Process capability (incapability) index has been introduced as a tool to aid in the assessment of process performance.

Greenwich and Jahr-Schaffrath (1995) introduced the incapability index under normality:

$$C_{pp} = \frac{(\mu - T)^2}{[\frac{1}{3}min\{U - T,\ T - L\}]^2} + \frac{\sigma^2}{[\frac{1}{3}min\{U - T,\ T - L\}]^2},$$

where μ is the process mean, σ^2 is the process variance, T is the target value, L and U are the lower and upper specification limits, respectively. This index provides an uncontaminated separation between information concerning the process accuracy and process precision. Greenwich and Jahr-Schaffrath (1995) discuss confidence intervals for C_{pp} based on a small and large samples. When both the mean μ and the variance σ^2 of the measurements are unknown, Greenwich and Jahr-Schaffrath (1995) proposed an unbiased and consistent estimator $\hat{C}_{pp}$ of C_{pp} under normality for small samples as the following:

$$\hat{C}_{pp} = \sum_{i=1}^{n}(X_i - T)^2/nD^2,$$

where $D = \frac{1}{3}\ min\{U - T,\ T - L\}$. For large samples, by assuming that the fourth moment exits, a natural estimator defined by

$$\hat{C}_{pp} = \frac{(\overline{X} - T)^2}{D^2} + \frac{S_n^2}{D^2},$$

is asymptotically normally distributed with mean C_{pp} and variance σ_{pp}^2/n, where $\overline{X} = \Sigma_{i=1}^{n}X_i/n$, $S_n^2 = \Sigma_{i=1}^{n}(X_i - \overline{X})^2/n$ and $\sigma_{pp}^2 = \frac{4(\mu - T)^2\sigma^2}{D^4} + \frac{4\mu_3(\mu - T)}{D^4} + \frac{\mu_4 - \sigma^4}{D^4}$, μ_3

$= E[(X-\mu)^3]$ and $\mu_4 = E[(X-\mu)^4]$. Chen (1998) showed that such a natural estimator is UMVUE under normality and the rth moment of the estimator was derived.

However, it is found that the index C_{pp} inconsistently measures process capability in many cases and thus reflects process potential and performance inaccurately. Chen (1998) proposed an adjusted incapability index, called C''_{pp}, to handle processes with both symmetric and asymmetric tolerances for a normal process.

$$C''_{pp} = \frac{[max\{(\mu-T)d/(U-T),\ (T-\mu)d/(T-L)\}]^2+\sigma^2}{[\frac{1}{3}min\{U-T, T-L\}]^2}$$

$$= \frac{1}{D^2}\{[max\{(\mu-T)d/(U-T),\ (T-\mu)d/(T-L)\}]^2+\sigma^2\}$$

where $d = \frac{U-L}{2}$ is the half length of specification interval. This index is written as the sum of the inaccuracy and imprecision of the process and achieves sensitivity for departures of the process mean μ from the target value T.

Chen (1998) proposed a natural estimator $\hat{C}''_{pp}$ of this incapability index where sample mean and sample variance (MLE) are used when both the mean μ and the variance σ^2 of the measurement are unknown. i.e.

$$\hat{C}'_{pp} = \frac{1}{D^2}\{[max\{(\overline{X}-T)d/(U-T),\ (T-\overline{X})d/(T-L)\}]^2+S_n^2\}$$

where $\overline{X} = \Sigma_{i=1}^n X_i/n$ and $S_n^2 = \Sigma_{i=1}^n (X_i-\overline{X})^2/n$ be the MLEs of the mean and the variance, respectively. In his paper, Chen derived the rth moment of $\hat{C}''_{pp}$. The mean squared error are also calculated to explore the behavior of the estimator. It has been shown that in general this natural estimator is biased. But for the case when the production tolerance is symmetric, the estimator $\hat{C}''_{pp}$ is an UMVUE of C''_{pp}.

Chen and Hsu (1999) discussed some large sample properties of $\hat{C}''_{pp}$ for i.i.d. processes. In this paper, we discuss the limiting distribution of an estimator $\hat{C}''_{pp}$ of the incapability index C''_{pp} for second order stationary processes under some mild conditions. The results allow one to find approximate $100(1-\alpha)\%$ confidence intervals.

Central Limit Theorem for the Random Vector $(\overline{X}, S_n^2)$.

Let $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, where $Z_t's$ are i.i.d. with mean 0 and variance σ^2, denote a discrete second order stationary process with mean μ and autocovariance function

$$\gamma(h) = Cov(X_t, X_{t+h}) = \sigma^2 \sum_{a=-\infty}^{\infty} \psi_a \psi_{a+h}$$

which rely on the lag h only. Let $X_1, \cdots, X_n$ be a sample of size n obtained from the stationary process $\{X_t\}$. In this section, we will derive the central limit theorem for the random vector $(\overline{X}, S_n^2)$ that is useful in section 3.

Theorem 1: Let $\{X_t\}$ denote a stationary process as defined above, such that $\sum_{i=-\infty}^{\infty} \psi_i$, $\sum_{j=-\infty}^{\infty} \psi_j^2$, $\sum_{j=-\infty}^{\infty} \psi_j^3$, $\sum_{j=-\infty}^{\infty} \psi_j^4$, $\sum_{i,j=-\infty}^{\infty} \psi_i \psi_j$, $\sum_{i,j=-\infty}^{\infty} \psi_i \psi_j^2$, $\sum_{i,j=-\infty}^{\infty} \psi_i^2 \psi_j^2$. are all finite, and $EZ_t^3 = \epsilon\sigma^3 < \infty$, $EZ_t^4 = \eta\sigma^4 < \infty$. Define

$$Y_t = \lambda_1(X_t - \mu) + \lambda_2(X_t - \mu)^2,$$

where λ_1 and λ_2 are two arbitrary constants not equal to 0 simultaneously, then

$$\sqrt{n}(\overline{Y} - \lambda_2\gamma(0)) \xrightarrow{L} N(0, \sum_{h=-\infty}^{\infty} \gamma_Y(h)),$$

where

$$\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

and

$$\gamma_Y(h) = \lambda_1^2\gamma(h) + \lambda_1\lambda_2\epsilon\sigma^3\sum_a \psi_a \psi_{a+h}^2 + \lambda_1\lambda_2\epsilon\sigma^3\sum_a \psi_a\psi_{a+h}^2 + \lambda_2^2(\eta - 3)\sigma^4\sum_a \psi_a^2\psi_{a+h}^2 + 2\lambda_2^2\gamma^2(h).$$

Proof:

It is known from page 220 of Brockwell and Davis (1987) that

$$E(X_i - \mu)^2(X_{i+h} - \mu)^2 = (\eta - 3)\sigma^4\sum_a \psi_a^2\psi_{a+h}^2 + \gamma^2(0) + 2\gamma^2(h).$$

By a similar computation, we obtain

$$Cov[(X_t - \mu)^2, (X_{t+h} - \mu)^2] = (\eta - 3)\sigma^4\sum_a \psi_a^2\psi_{a+h}^2 + 2\gamma^2(h).$$

Simlarly,

$$Cov[(X_t - \mu), (X_{t+h} - \mu)^2] = \epsilon\sigma^3 \sum_a \psi_a \psi_{a+h}^2, \tag{1}$$

and

$$Cov[(X_t - \mu)^2, (X_{t+h} - \mu)] = \epsilon\sigma^3 \sum_a \psi_a^2 \psi_{a+h}.$$

Let $\gamma_Y(h)$ be the covariance of Y_t and Y_{t+h}. Then

$$\begin{aligned}
&\gamma_Y(h) \\
&= Cov(Y_t, Y_{t+h}) \\
&= Cov[\lambda_1(X_t - \mu) + \lambda_2(X_t - \mu)^2, \lambda_1(X_{t+h} - \mu) + \lambda_2(X_{t+h} - \mu)^2] \\
&= \lambda_1^2 Cov(X_t - \mu, X_{t+h} - \mu) + \lambda_1\lambda_2 Cov[(X_t - \mu), (X_{t+h} - \mu)^2] + \lambda_1\lambda_2 Cov \\
&\quad [(X_t - \mu)^2, (X_{t+h} - \mu)] + \lambda_2^2 Cov[(X_t - \mu)^2, (X_{t+h} - \mu)^2] \\
&= \lambda_1^2\gamma(h) + \lambda_1\lambda_2\epsilon\sigma^3 \sum_a \psi_a \psi_{a+h}^2 + \lambda_1\lambda_2\epsilon\sigma^3 \sum_a \psi_a^2\psi_{a+h} + \lambda_2^2(\eta - 3)\sigma^4 \sum_a \psi_a^2\psi_{a+h}^2 \\
&\quad + 2\lambda_2^2\gamma^2(h).
\end{aligned}$$

The following arguments borrow the techniques used in the proof of Theorem 7.1.2 in Brockwell and Davis (1987).

Define

$$\begin{aligned}
X_{tm} &= \mu + \sum_{j=-m}^{m} \psi_j Z_{t-j} \\
Y_{tm} &= \lambda_1(X_{tm} - \mu) + \lambda_2(X_{tm} - \mu)^2 \\
\overline{Y}_{nm} &= \frac{1}{n}\sum_{t=1}^{n} Y_{tm} \\
\overline{Y} &= \frac{1}{n}\sum_{t=1}^{n} Y_t.
\end{aligned}$$

Then by theorem 6.4.2. of Brockwell and Davis (1987), $\sqrt{n}(\overline{Y}_{nm} - \lambda_2\sigma^2)$ is asymptotically normal with mean 0 and variance v_m^2, such that $\lim_{m\to\infty} v_m = \sum_{h=-\infty}^{\infty} \gamma_Y(h)$.

It is sufficient to prove $\lim_{m\to\infty} \lim_{n\to\infty} Var\sqrt{n}(\overline{Y} - \overline{Y}_{nm}) = 0$, since then $\sqrt{n}(\overline{Y} - \lambda_2\sigma^2)$ $\to N(0, \sum_{h=-\infty}^{\infty} \gamma_Y(h))$ as desired. Clearly

$$\overline{Y} - \overline{Y}_{nm}$$

$$= \frac{1}{n}\sum_{t=1}^{n}(Y_t - Y_{tm})$$

$$= \frac{1}{n}\sum_{t=1}^{n}[\lambda_1(X_t - X_{tm}) + \lambda_2(X_t - X_{tm})(X_t + X_{tm} - 2\mu)]$$

$$= \frac{1}{n}\sum_{t=1}^{n}\{\lambda_1\sum_{|j|>m}\psi_j Z_{t-j} + \lambda_2(\sum_{|j|>m}\psi_j Z_{t-j})[\sum_{-\infty}^{\infty}\psi_j Z_{t-j} + \sum_{|j|<m}\psi_j Z_{t-j}]\}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\{\lambda_1\sum_{|j|>m}\psi_j Z_{t-j} + \lambda_2(\sum_{|j|>m}\psi_j Z_{t-j})^2 + 2\lambda_2(\sum_{|j|>m}\psi_j Z_{t-j})(\sum_{|j|<m}\psi_j Z_{t-j})\}$$

$$= \frac{1}{n}\sum_{t=1}^{n}A(t,m)$$

where

$$A(t,m) = \lambda_1\sum_{|j|>m}\psi_j Z_{t-j} + \lambda_2(\sum_{|j|>m}\psi_j Z_{t-j})^2 + 2\lambda_2(\sum_{|j|>m}\psi_j Z_{t-j})(\sum_{|j|<m}\psi_j Z_{t-j}).$$

Since $\Sigma\psi_j^2$, $\Sigma\psi_j^3$, $\Sigma\psi_j^4$ are finite, and $VarA(t,m)$ is a linear combination of terms like $\sum_{|j|>m}\psi_j^2$, $(\sum_{|i|>m}\psi_i^2)(\sum_{|j|>m}\psi_j^2)$, $\sum_{|j|>m}\psi_j^3$, $\sum_{|j|>m}\psi_j^4$, and $(\sum_{|i|<m}\psi_i^2)(\sum_{|j|>m}\psi_j^2)$, etc., we have

$$\lim_{m\to\infty} VarA(t,m) = 0$$

for $t = 1, 2, \cdots, n$. Similarly, since $\sum_{i,j=-\infty}^{\infty}\psi_i\psi_j$, $\sum_{i,j=-\infty}^{\infty}\psi_i\psi_j^2$ and $\sum_{i,j=-\infty}^{\infty}\psi_i^2\psi_j^2$ are finite, and $\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(A(i,m),\ A(j,m))$ is a linear combination of terms like $\sum_{\substack{a-b=1-i\\ |a|>m,|b|>m}}^{n-i}\psi_a\psi_b$, $\sum_{\substack{a-b=1-i\\ |a|>m,|b|>m}}^{n-i}\psi_a\psi_b^2$, $\sum_{\substack{a-b=1-i\\ |a|>m,|b|>m}}^{n-i}\psi_a^2\psi_b^2$, $\sum_{\substack{a-b=1-i\\ |a|<m,|b|>m}}^{n-i}\psi_a^2\psi_b^2$, $\sum_{|b|>m}\psi_b^3$, etc., for $i = 1, 2, \cdots, n$, we have $\lim_{n\to\infty}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(A(i,m),\ A(j,m))$ is finite and

$$\lim_{m\to\infty}\lim_{n\to\infty}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(A(i,m),A(j,m)) = 0, \text{ for } i = 1, 2, \cdots, n.$$

Therefore,

$$\lim_{m\to\infty}\lim_{n\to\infty} Var\sqrt{n}(\overline{Y} - \overline{Y}_{nm})$$

$$= \lim_{m\to\infty}\lim_{n\to\infty} Var\frac{1}{\sqrt{n}}\sum_{t=1}^{n}A(t,m)$$

$$= \lim_{m\to\infty}[Var\,A(1,m) + \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(A(i,m),A(j,m))] = 0,$$

and the proof is completed. □

Theorem 2: Under the same assumptions as theorem 1,

$$\sqrt{n}\{(\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu),\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^2)-(0,\gamma(0))\}\xrightarrow{L}N((0,0),\Sigma), \tag{2}$$

where Σ is given by

$$\begin{pmatrix}\sum_{h=-\infty}^{\infty}\gamma(h) & \epsilon\sigma^3\sum_a\psi_a^3\\ \epsilon\sigma^3\sum_a\psi_a^3 & (\eta-3)\gamma^2(0)+2\sum_{h=-\infty}^{\infty}\gamma^2(h)\end{pmatrix}. \tag{3}$$

Proof: By the following observation

$$Cov[(X_t-\mu),(X_t-\mu)^2]=\epsilon\sigma^3\sum_a\psi_a^3$$

obtained from (1) (let $h=0$), theorems 1, and the standard Cramer and Wold (1936) argument. □

Theorem 3: Under the same assumptions as theorem 1,

$$\sqrt{n}[(\frac{1}{n}\sum_{t=1}^{n}X_t,\frac{1}{n}\sum_{t=1}^{n}(X_t-\overline{X})^2)-(\mu,\gamma(0))]\xrightarrow{L}N((0,0),\Sigma), \tag{4}$$

where Σ is given by (3).

Proof: Equation (2) can be rewritten as

$$\sqrt{n}[(\frac{1}{n}\sum_{t=1}^{n}X_t,\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^2)-(\mu,\gamma(0))]\xrightarrow{L}N((0,0),\Sigma), \tag{5}$$

and (5) implies (4) in view of (2). □

Limiting Distribution for $\hat{C}''_{pp}$

Let $\{X_t\}$ denote a discrete second order stationary process with common mean μ, autocovariance function $\gamma(h)$ and higher moments as defined in section 2. Then the common variance is $\gamma(0)$.

The process incapability index

$$C''_{pp}=\frac{1}{D^2}\{[max\{(\mu-T)d/(U-T),(T-\mu)d/(T-L)\}]^2+\gamma(0)\}$$

can be rewritten as

$$C''_{pp} = \begin{cases} \frac{1}{D^2}\{[(T-\mu)d/(T-L)]^2 + \gamma(0)\} & \text{if } L < \mu < T < U \\ \frac{\sigma^2}{D^2} & \text{if } L < \mu = T < U \\ \frac{1}{D^2}\{[(-T+\mu)d/(U-T)]^2 + \gamma(0)\} & \text{if } L < T < \mu < U. \end{cases}$$

From section 2, $\overline{X}$ and S_n^2 are consistent estimators of μ and $\gamma(0)$, respectively. When both the mean μ and the variance $\gamma(0)$ of the measurement are unknown, let's consider the following estimator:

$$\hat{C}''_{pp} = \frac{1}{D^2}\{[max\{(\overline{X}-T)d/(U-T),\ (T-\overline{X})d/(T-L)\}]^2 + S_n^2\}.$$

which is equivalent to the following expression:

$$\hat{C}''_{pp} = \begin{cases} \frac{1}{D^2}\{[(T-\overline{X})d/(T-L)]^2 + S_n^2\} & \text{if } L < \mu < T < U \\ \frac{1}{D^2}\{[max\{(\overline{X}-T)d/(U-T),\ (T-\overline{X})d/(T-L)\}]^2 + S_n^2\} & \text{if } L < \mu = T < U \\ \frac{1}{D^2}\{[(-T+\overline{X})d/(U-T)]^2 + S_n^2\} & \text{if } L < T < \mu < U. \end{cases}$$

Theorem 4: Let $\{X_t\}$ denote a discrete second order stationary process which satisfying the assumptions in Theorem 1. Then

$$\sqrt{n}(\hat{C}''_{pp} - C''_{pp}) \xrightarrow{L} N(0, \mathbf{B}\Sigma\mathbf{B}')$$

where Σ is given in (3), and

$$\mathbf{B} = \begin{cases} (\frac{-2(T-\mu)d^2}{D^2(T-L)^2}, \frac{1}{D^2}) & \text{if } L < T < \mu < U. \\ (\frac{2(-T+\mu)d^2}{D^2(U-T)^2}, \frac{1}{D^2}) & \text{if } L < \mu \leq T < U \end{cases}$$

Proof:

Define

$$g_1(a,b) = \frac{1}{D^2}\{[max\{(a-T)d/(U-T),\ (T-a)d/(T-L)\}]^2 + b\}$$

$$= \begin{cases} \frac{1}{D^2}\{[(T-a)d/(T-L)]^2 + b\} & \text{if } L < a \leq T < U \\ \frac{1}{D^2}\{[(-T+a)d/(U-T)]^2 + b\} & \text{if } L < T < a < U. \end{cases}$$

Then $g_1(a,b)$ is a real valued function and is differentiable for all $a \in (L,U)$, and $b > 0$ with

$$\frac{\partial g_1(a,b)}{\partial a} = \begin{cases} \frac{1}{D^2}\{[-2\frac{\{(T-a)d^2}{(T-L)^2}]\} & \text{if } L < a \leq T < U \\ \frac{1}{D^2}\{[2\frac{(-T+a)d^2}{(U-T)^2}]\} & \text{if } L < T < a < U. \end{cases}$$

and $\frac{\partial g_1(a,b)}{\partial b} = \frac{1}{D^2}$ for $a \in (L,U)$.

It is clear that $C''_{pp} = g_1(\mu, \gamma(0))$ and $\hat{C}''_{pp} = g_1(\overline{X}, S_n^2)$.

Define $\mathbf{B} = (\frac{\partial g_1(a,b)}{\partial a}, \frac{\partial g_1(a,b)}{\partial b})\Big|_{(a,b)=(\mu,\gamma(0))}$.

$$= \begin{cases} (\frac{-2(T-\mu)d^2}{D^2(T-L)^2}, \frac{1}{D^2}) & \text{if } L < \mu \leq T < U \\ (\frac{2(-T+\mu)d^2}{D^2(U-T)^2}, \frac{1}{D^2}) & \text{if } L < T < \mu < U. \end{cases}$$

then $\mathbf{B} \neq (0, 0)$. Hence, by Theorem on page 24 and page 72 in Serfling (1980), $\sqrt{n}(g_1(\overline{X}, S_n^2) - g_1(\mu, \gamma(0))) \xrightarrow{L} N(0, \mathbf{B}\Sigma\mathbf{B}')$, where Σ is given in (3). Therefore,

$$\sqrt{n}(\hat{C}''_{pp} - C''_{pp}) \xrightarrow{L} N(0, \mathbf{B}\Sigma\mathbf{B}'). \square$$

It was shown by Chen (1998) that in general the estimator $\hat{C}''_{pp}$ is biased for a normal process. From Theorem 3 above, we see that the estimator $\hat{C}''_{pp}$ is asymptotically normally distributed with mean C''_{pp} under some mild conditions, i.e. the estimator $\hat{C}''_{pp}$ is asymptotically unbiased and consistent. And with the asymptotic variance $\sigma^2_{C''_{pp}}/n$, where

$$\sigma^2_{C''_{pp}} = \frac{4}{D^4}\{\frac{(T-\mu)^2 d^4 \gamma(0)}{G^4} - \frac{(T-\mu)d^2\mu_3}{G^2} + \frac{\mu_4 - \gamma(0)^2}{4}\},$$

$$G = \begin{cases} T - L \text{ if } L < \mu \leq T < U \\ U - T \text{ if } L < T < \mu < U. \end{cases}$$

Except what we observed above, the limiting distribution of $\hat{C}''_{pp}$ also provides a way to find the probability of events and to find approximate confidence interval. i.e. an approximate $100(1-\alpha)\%$ confidence interval for C''_{pp} is $(\hat{C}''_{pp} - \frac{(\mathbf{B}\Sigma\mathbf{B}')^{1/2}}{\sqrt{n}} z_{1-\alpha/2}, \hat{C}''_{pp} + \frac{(\mathbf{B}\Sigma\mathbf{B}')^{1/2}}{\sqrt{n}} z_{1-\alpha/2})$.

ACKNOWLEDGMENTS

The authors are grateful to the referees' suggestions that have greatly improved the quality of this paper.

Bibliography

(1) Brockwell P. J. and Davis R. A. (1987). "Time Series: Theory and Methods". Springer Verlag, New York.

(2) Chen, K. S (1998) "Incapability Index With Asymmetric Tolerances" *Statistica Sinica*, 8, p253-p262.

(3) Chen, S. M. and Hsu, Y. S. (1999) "Large Sample Properties of the Estimator of the Process Incapability Index C''_{pp}", Submitted to Metrika.

(4) Cramer, H. and Wold, H. (1936). "Some theorems on distributionfunctions," *J. London Math. Soc.*, 11, 290-295.

(5) Greenwich, M. and Jahr-Schaffrath, B. L. (1995). "A Process Incapability Index," *International Journal of Quality and Reliability Management*, 12 (4), 58-71.

(6) Serfling, R. J. (1980). "Approximation Theorems of Mathematical Statistics," John Wiley and Sons, New York, 1-125.

received September 1, 1999
revised November 19, 1999
accepted December 6, 1999

製程能力指標估計式 $\hat{C}''_{pp}$ 在貳階穩態過程之近似分佈

陳 思 勉
輔仁大學數學系

許 玉 生
中央大學數學系

摘 要

對於獨立常態過程資料，Chen（1998）提出製程能力指標 C''_{pp}。相較於其它製程能力指標，此製程能力指標非但可將製程準確度與製程精確度分開表之，而且適用於對稱容忍型及非對稱容忍型問題。然而在實務上有相當多的非獨立且非常態型資料。本文將探討當資料來自於貳階穩態過程時製程能力指標 估計式 C''_{pp} 之近似分佈。

關鍵詞： 製程能力指標

A Novel Label Distribution Protocol for Label-Switching based IP Networks

Huey-Ing Liu
Department of Electronic Engineering
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.
Chun-Te Wu
Department of Computer Science
Chung Hua University
Hsinchu, Taiwan 300, R.O.C.

Abstract

Utilizing hardware-based switching to improve the cost, performance and scalability of Layer 3 routing, referred to as Layer 3 switching, is one of the most prospective techniques to efficiently transfer large amounts of diverse traffic over Internet or Intranet. One of Layer 3 switching implementations, referred to as label switching, is to attach short, fixed length labels to the forwarded packets. Switching entities perform table lookups based on these simple labels to determine where packets should be forwarded. This paper presents a loop-free label distribution protocol for label-switching based IP networks. Multiple paths are provided to enhance network connectivity and flexibility. A novel loop prevention algorithm, which employs a port-id list to prevent loops, is proposed. The proposed scheme applies a constant processing time and a shorter message length. Furthermore, no global unique identifier is required for edge routers and switching nodes.

Key Words: Layer 3 Switching, Label Switching, IP, Edge Router

INTRODUCTION

Statistics demonstrate that both the number of hosts and traffic on Internet are increasing exponentially. Traditional routers cannot keep up with the explosive growth of Internet and Intranet traffic. Several new technologies for routing IP traffic such as wire-speed routing, layer 3 switching and layer 4 switching are emerging [1~7]. By exploiting recent advances in high-speed switch hardware, layer 3 and layer 4 switching efficiently accelerate packet-forwarding speed. Label switching is one of layer 3 switching implementations, where a short, fixed length, locally significant identifier is augmented in front of IP header to guide the route of IP packets. The identifier is generally referred to as label. The label of each packet is used as an index for a switching table lookup to determine where data should be forwarded.

In traditional IP packet routing, as a packet traverses the network, each router extracts all the information relevant to routing from the IP header. In the most common case, the routing information is usually the destination IP address. Using this destination address as a matching pattern, a time consuming longest-prefix matching algorithm is then performed to find the packet's next hop. Sometimes the routing information may be contained in other fields, a complex header analysis is then required. All these should be repeated at each router through which the packet travels. Even in the simplest case, since the longest-prefix matching is really complicated and usually implemented by software, routing IP packet is time consuming. Recently, several fast IP table lookup algorithms are proposed to accelerate the lookup speed [8~11]. Wire-speed routers, which route IP packets at wire speed, are designed to speedup IP packet forwarding [7]. For wire-speed router, complicated and expansive hardware design is required to handle the complex routing processing. However, fixed-length label significantly simplifies the hardware design and wire-speed packet switching could be easily achieved via label switching. Furthermore, no routing information analysis for each data packet is required at label switching nodes. However, a special router, referred to as edge router, at the edge of label switching networks is necessitated to perform label and complete IP routing information mapping. The edge router is a gateway node between label switching domain and

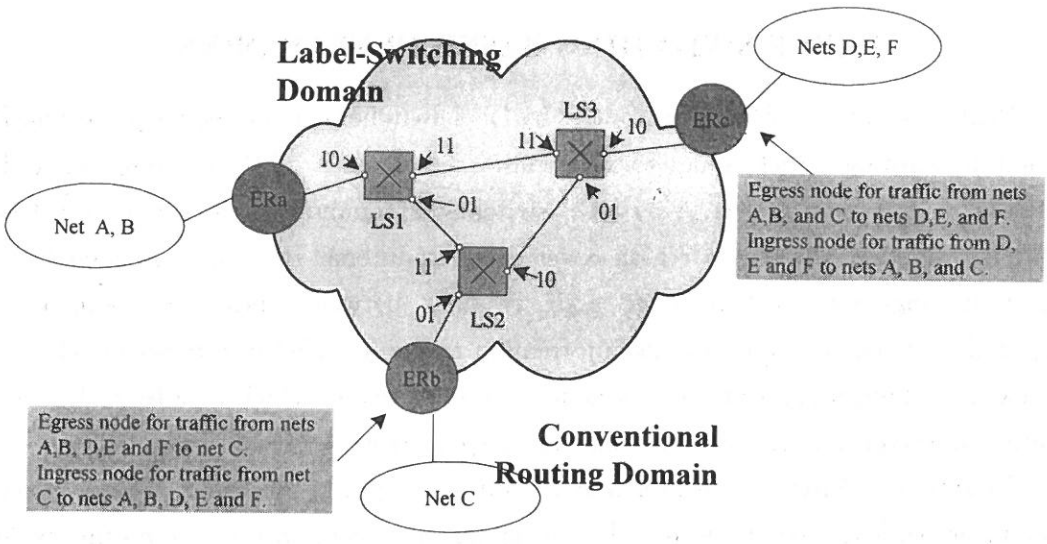


Fig. 1. A typical label switching based IP network.

legacy routing domain as shown in Fig. 1. Where the label-switching domain is a set of contiguous nodes, referred to as label switches, which operate label switching routing and forwarding. A set of contiguous traditional routers forms a legacy routing domain. Edge router has to handle the insertion and deletion of label.

For label switching based IP networks, a label distribution protocol to establish the routing database for each label switch is proposed in this paper. The proposed Loop Free Label Distribution Protocol (LFLDP) provides multiple paths to enhance the network connectivity and flexibility. A Novel Loop Prevention Algorithm (NLPA), which employs a port-id list to prevent loops, is proposed. The NLPA applies a constant processing time and a shorter message length. Furthermore, no global unique identifier is required for edge routers and label switches.

The rest of this paper is organized as follows. Section 2 describes the system configuration of a label switching based IP network. The details of LFLDP are presented in Section 3. Section 4 gives the proposed loop prevention algorithm and its implementation. Finally, some concluding remarks are made in Section 5.

LABLE SWITCHING BASED IP NETWORKS

Replacing core traditional IP routers by label switches could effectively speedup IP packet forwarding. Such a network is considered to be a label switching based IP network. The region governed by label switches is commonly referred to as a label switching domain whereas a region covered by traditional IP routers is named as a legacy routing domain. An edge node residing between these two domains is responsible of the label and routing information mapping, label insertion and deletion. A label switching ingress and egress node is an edge node, which handles traffic as it enters and leaves a label-switching domain, respectively.

Traditional routers use destination IP address to guide the route. Thus, in each router, network identifier and next hop mappings are maintained in the routing table. Each router would attempt to find a network identifier in its routing table, which has a longest-prefix matching with the destination IP address of the incoming packet's IP header. We call such a routing a topology-based routing. However, several new applications especially multimedia applications drive the requirements of flow-based or stream-based routing. A flow is a single instance of an application to application flow of data whereas a stream stands for an aggregation of one or more flows. In such a case, a flow (stream) descriptor instead of network identifier is used to guide the route. Different from topology-based routing, since no stream or flow descriptor is resided in IP header and all intermediate routing nodes have no idea of each specific flow, a connection set up is required for each flow- (stream-) based routing. For a label switching based IP network, since a label could be stand for a specific network- or flow- (stream-) identifier, both topology- and stream-based routing could be supported. The flow- (stream-) based routing could be provided via integrating with existing control protocols especially the quality of service (QOS) guaranteed routing protocols, such as RSVP [12 ~ 14]. The label could be piggyback as part of the confirmation in set-up phase. In our opinion, a best-effort packet that is it has no special QOS guaranteed requirements it could be forwarded via topology-based switching. However, those packets with QOS guaranteed requirements need flow-based routing. A flow set up is required to establish the route and reserve resources

along the routing path. Thus, the control and distribution of flow-to-label mappings of flow-based routing could be hooked with existing QOS guaranteed routing protocols. However, for topology-based routing, since no route set-up procedure is proceeded previously, it is not suitable to integrate the label distribution of topology-based routing with a QOS guaranteed routing protocol. Thus, in this paper, a topology-based label distribution protocol is proposed.

THE LFLDP

The LFLDP employs a flooding-styled label distribution and consists of two main portions: route maintenance and loop prevention. The former one is used to establish topology-based routings for all connected IP networks. In order to keep consistency with network status changing, each edge router would periodically advertise its currently active routes. In the mean while, the loop prevention algorithm is invoked to prevent establishing loop-routings. Before detailing the operation of the LFLDP, the packet formats of LFLDP are introduced firstly.

1. LFLDP packet format

Similar to ICMP, UDP, and TCP, LFLDP packets are carried as the payload of IP datagrams. Fig. 2 shows the common header of LFLDP. The LFLDP common header contains four fields. The first byte indicates the LFLDP version number. The second byte identifies the specific packet type. LFLDP uses two different types of packets, listed in Table 1. Other than the common LFLDP header, each packet type has its own unique format and would be introduced in the following subsection. The packet length field indicates the size of the packet in bytes. The next field uniquely identifies the router that originated the packet; this value is one of the router's IP addresses.

Table 1. LFLDP packet type.

1	Route-Advertisement
2	Label-Canceling

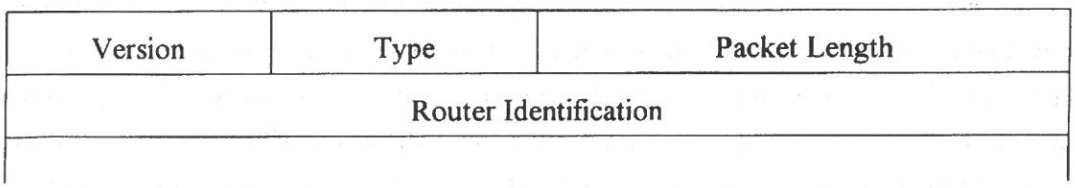


Fig. 2. The common header of LFLDP.

(1) **Route-Advertisement Message:**

LFLDP establishes multiple Topology-based routings for each connected IP networks. Initially, each edge router generates a Route-Advertisement packet to establish routings for all its locally connected networks. The Route-Advertisement packet is then flooded on all interfaces in label switching domain. Once a label switch receives a Route-Advertisement Message, it updates its routing database and distributes that information to the rest of the network. To do this, it floods the Route-Advertisement Message on all interfaces except the incoming one. Furthermore, in order to keep consistency with network changing, the edge router would periodically broadcast its routing information to refresh routings for all locally connected networks.

After the common header, the Route-Advertisement packet in Fig. 3 contains one 32-bit word with the number of individual route advertisements in the packet. It is then followed by the advertisements themselves. Each route-advertisement packet consists of five fields. The first field is the route identifier (RID) of this advertisement. The RID is identified via two sub-fields: network IP address and network mask. The second field specifies the label associate with this route. The route dead interval (RDI) tells how long it takes to declare a route unavailable. If the route has not been advertised in this number of seconds, it is considered dead.

The metric specifies the aggregate route status. It consists of two sub-fields: metric type (MT) and value. LFLDP uses three different types of metric, listed in Table 2. Hop count, naturally, indicates the number of hops from the originating node to current node. Delay specifies the total delay time form the originating node to current one. Link cost indicates the aggregate cost associated with this route. In order

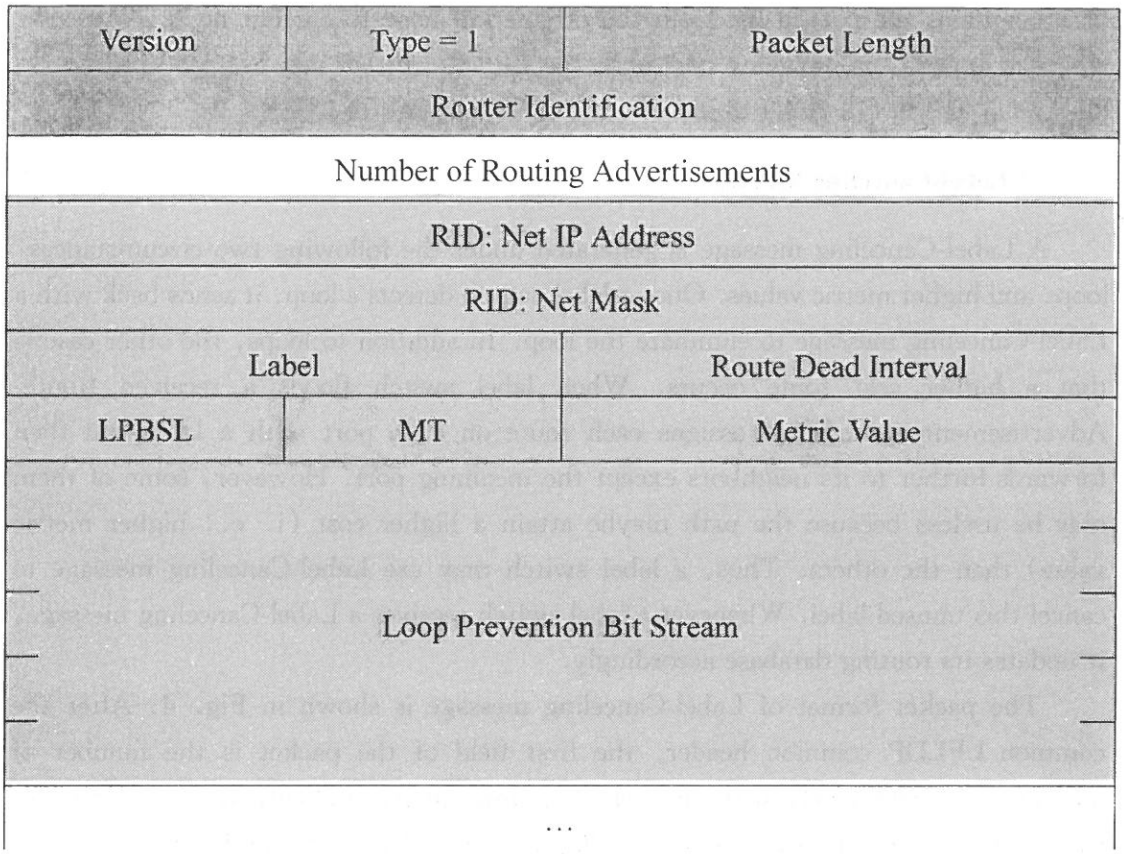


Fig. 3. The packet format of Route-Advertisement message.

Table 2. Metric type.

1	Hop Count
2	Delay
3	Link Cost

to enhance connectivity and flexibility, LFLDP provides multiple paths for each route. The metric value helps to calculate the optimal multi-path for each route.

As mentioned before, LFLDP employs port-id list to prevent loops. For a route, each intermediate node along the route appends its incoming port-id, which is represented by a bit string, to the loop prevention bit stream (LPBS) field. Thus, the

LPBS contains the port-id list from the originating node to current node. To realize where to append the incoming port-id to the LPBS, LFLDP uses the loop prevention bit stream length (LPBSL) to indicate the current length of LPBS.

(2) **Label-Canceling Message**:

A Label-Canceling message is generated under the following two circumstances: loops and higher metric values. Once a label switch detects a loop, it sends back with a Label-Canceling message to eliminate the loop. In addition to loops, the other case is that a higher cost route occurs. When label switch floods a received Route-Advertisement message, it assigns each route on each port with a label and then forwards further to its neighbors except the incoming port. However, some of them may be useless because the path maybe attain a higher cost (i. e., higher metric value) than the others. Thus, a label switch may use Label-Canceling message to cancel this unused label. Whenever a label switch receives a Label-Canceling message, it updates its routing database accordingly.

The packet format of Label-Canceling message is shown in Fig. 4. After the common LFLDP common header, the first field of the packet is the number of individual canceled-labels in the packet. It is immediately followed by canceled-labels themselves. Each canceled-label is specified via two fields: RID and label.

2. Routing databases

In this subsection, the routing databases within edge routers (ER) and label switches (LS) are introduced.

LS Label-Switching Database (LSLSD): At each label switch, the incoming label (IN-LB): of an arriving packet is used as indices in the LSLSD to uniquely identify the outgoing label (OUT-LB) and the outgoing port (OUT-PORT). In order to speedup the data switching, label switch maintains distributed LSLSD. That is, for each port, each label switch keeps an individual LSLSD. Each entry in LSLSD is corresponded to a topology-based route destined to an IP network. The routing information, such as route-id, route dead interval, ···etc., is recorded in LS routing database.

Version	Type = 2	Packet Length
Router Identification		
Number of Label-Cancelings		
RID: Net IP Address		
RID: Net Mask		
Label		Reserved
...		

Fig. 4. The packet format of Route-Canceling message.

LS Routing Database (LSRD): A label switch uses a LSRD to record the routing information for each connected network. A LSRD contains the following information:

- RID: The RID identifies the destination of the route and is indicated via network IP address and network mask. For instance, a route destined to network A as shown in Fig. 1 denotes as (A, M_A), where MA is the net mask of network A.
- Metric: Three different metric values are provided: Hop Count (HC), Delay (D) and Link Cost (LC).
- LPBS: The LPBS specifies the port-id list from the originating node to current node. We assume that the maximum port number of each visited label switch is 32 ($32=2^5$) ports and the maximum length of a route is 32 hops. Thus, the LPBS field is 20 ($32 \times 5/8 = 20$) byte. If the current LPBS is shorter than 20 bytes, the remaining bits are set to be 0.
- LPBSL: The LPBSL specifies the length of LPBS.
- RDI: The RDI tells how long the route needs to be refreshed again.
- Timer: The Timer records how long the time has been elapsed since last refreshing.
- OUT-PORT: The outgoing port of this route. This information also appears in LSLSD. The label used by OUT-PORT is denoted as the OUT-LB.
- LB_i: The LBi records the label of port i. This information also appears in LSLSD.
- Counter: The Counter contains the number of useful labels excluding the label used

in the incoming port (i. e., OUT-LB). Notably, the incoming port of a Route-Advertisement message is the outgoing port (i. e., OUT-PORT) of the route (s) in this message. While establishing a route, the route-advertisement message is flooded on all ports except the incoming one and a label is assigned for each port. However, some of them are useless since its neighbor may choose a different path or the route contains a loop. In such a case, its neighbor would feedback a Label-Canceling message to cancel the label.

Notably, since LFLDP provides multiple paths to improve connectivity and flexibility, each label switch would keep at most m ($=2$ or 3) records with the same RID.

ER Route-to-Label Database (ERRLD): Edge router is a gateway between legacy routing domain and label-switching domain. Thus, in addition to the original routing table, each edge router also maintains an ER Route-to-Label Database to handle packets entering into or leaving from label switching domain. An ERRLD contains the following information: RID, RDI, Timer, OUT-PORT, LB. Fig. 5 demonstrates how these databases help data forwarding.

3. Protocol descriptions

Initially, each edge router assigns a label for each of locally connected networks. Then, it creates a Route-Advertisement message with all Rote-to-Label mappings and floods this message into label-switching domain. For example, the edge router ERa, as shown in Fig. 1, generates a Route-Advertisement message with two route advertisements for networks A and B and then sends it to LS1. Wile receiving a Route-Advertisement message, the operations of label switches and edge routers are different; thus, their operations are discussed individually as follows.

Label Switch's Operations

Two types of messages may be received and discussed as follows.

(1) Route-Advertisement Message

Once a label switch receives a Route-Advertisement Message, for each route advertisement, it makes the following checks.

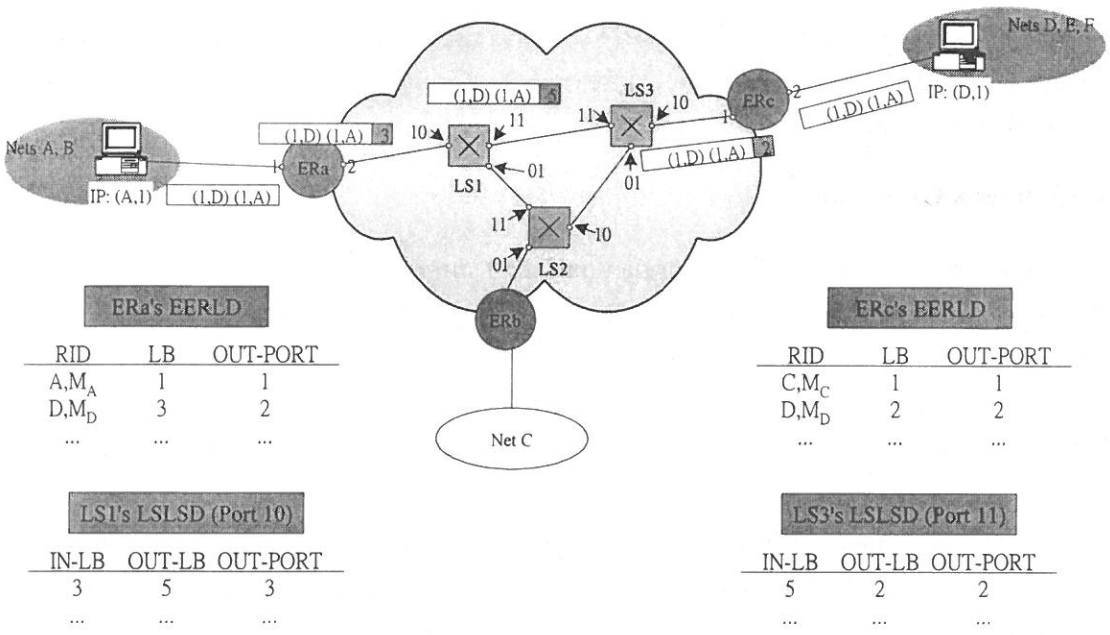


Fig. 5. An illustrative example of data forwarding in a label-switching based IP network.

1. The label switch first checks whether it has the route or not. If yes, it resets the timer of the route.
2. Otherwise, it compares the metric value, chooses a better one and then if necessary, cancels the unused route via sending back a Label-Canceling message.
3. If the new incoming route passes the second checking, it invokes loop-prevention algorithm to verify whether there is a loop or not. If yes, it sends back a Label-Canceling message to cancel the loop; otherwise, it inserts the new incoming route into its LSRD and LSLSD.

The label switch then floods the routes, which have survived after the above checks, on all interfaces except the incoming one. Notably, for each connected network, the originating edge router would choose a metric type and all the successors should follow the same metric type.

(2) Label-Canceling Message

When a label switch receives a Label-Canceling message, it first cancels the

unused label (s) in LSLSD and LSRD. Then, update the corresponding counter of LSRD. If the counter becomes zero, it further issues a Label-Canceling message on the OUT-PORT to cancel the OUT-LB of the corresponding route.

Edge Router's Operations

Assume each edge router contains only one interface in label-switching domain. Suppose an edge router receives a Route-Advertisement message, for each route advertisement, it first checks whether it contains the route or not. If yes, it updates the timer of this route in its ERRLD. Since the edge router holds only one interface in label-switching domain, no loop-prevention algorithm is invoked. Thus, it simply checks the metric value and reserves the best m paths. If necessary, it feeds back a Label-Canceling message to cancel all unused labels. To adapt to topology change, all edge router periodically distributes a Route-Advertisement message for locally connected networks. While receiving a Label-Canceling message, an edge router simply updates the ERRLD. No Label-Canceling message is sent by an edge router since it is the edge node of label-switching domain.

The operations of label switches and edge routers are formally depicted in Figs. 6 and 7, respectively. A node may be in one of the following four states:

IDLE: A node with nothing to do, it is in IDLE state.

UPDATE-ROUTE: While a node updates its routing database according to the received route advertisements, it is in UPDATE-ROUTE state.

FLOODING: After updating routing database (s), a node floods the updated routes on all interfaces except the incoming one. Then, it is in FLOODING state.

CANCEL-LABEL: CANCEL state stands for a node is in canceling unused routes and updating the corresponding databases.

CANCELING and CANCELING-C: If necessary, while a node is in sending a Label-Canceling message, it is in CANCELING or CANCELING-C state. For distinguishing, two different states are defined since the following operations are distinct.

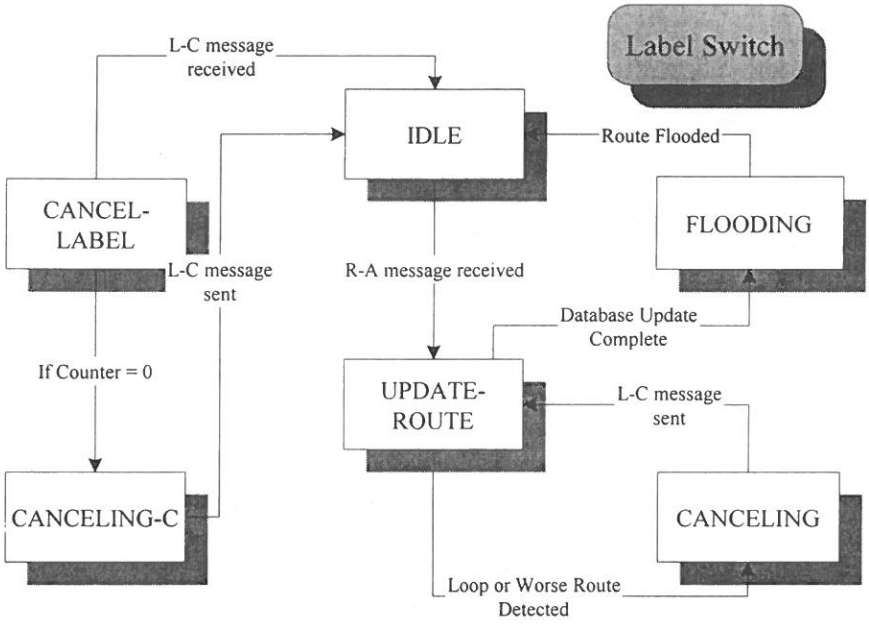


Fig. 6. The state transition diagram of label switch.

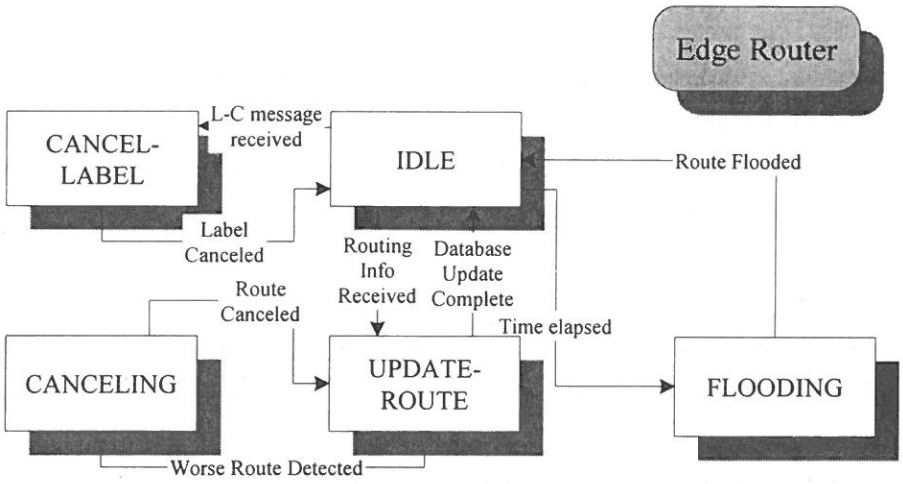


Fig. 7. The state transition diagram of edge router.

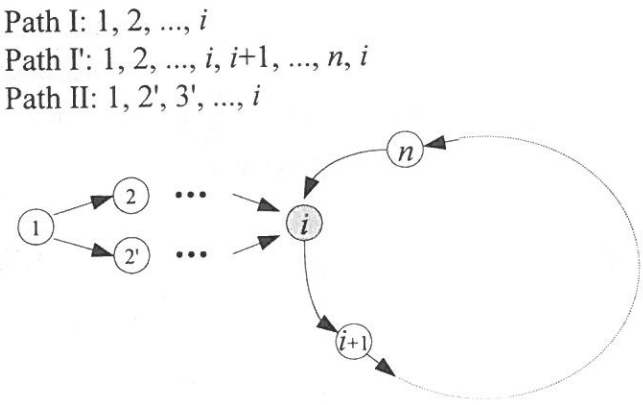


Fig. 8. An illustrated example of loop path.

THE NLPA

1. Operation principles

Obviously, suppose a routing algorithm supports only one path and metric values are used to select the optimal path, a loop path would not be picked. Without loss of generality, consider an example as shown in Fig. 8. While Path I' reaches node i, node i compares path I with Path I'. Path I is selected since the metric value of Path I must be better than or equal to that of Path I'. Although LFLDP use metric value to select a better path, a loop path may be chosen because LFLDP provides multiple paths in its routing database. Assume LFLDP supports two paths. Let the metric value of a path P denote as MV (P). Consider the example in Fig. 8 again, suppose MV (Path I) $<=$ MV (Path I') $<=$ MV (Path II), node i would select Path I and Path I' where Path I' contains a loop.

In order to eliminate such a circumstance, LFLDP uses port-id lists to prevention loops. In LFLDP, each route is associated with a port-id list containing the *incoming* port ids from the beginning node to current nod. For example, consider a route as shown in Fig. 1, which is destined to network A, originating from ERa, traversing LS1, and arriving LS2, the resulting port-id list (i. e., LPBS) becomes 1011. Such

an incoming port-id list has a useful feature as shown in Lemma 1.

Lemma 1. *Let G (V, E) be a graph in which each node (label switch) runs LPLDP. Given two routes R1 and R2 established via LPLDP. Both R1 and R2 originate from node o and end at node e. Then, R1 = R2 if and only if the LPBS of R1 is the same as that of R2.*

Proof: Obviously, if R1 = R2, i. e., all the visited nodes and ports of R1 and R2 are the same, then the LPBS of R1 is the same as that of R2.

Without loss of generality, assume R1 = (*o*, ···, *n*, *e*) and R2 = (o, ···, n', e). Suppose that the LPBS of R1 is the same as that of R2. Since R1 and R2 end at the same node e, R1 and R2 must have the same incoming port in node N. Otherwise, the LPBS of R1 would be different from that of R2. The incoming ports of R1 and R2 in node e are the same, thus, the previous nodes of R1 (i. e., *n*) and R2 (i. e., *n'*) are the same (i. e., $n = n'$). Similarly, since LPBS of R1 is equal to that of R2, the incoming ports of R1 and R2 in node *n* must be the same. Following the same procedure, we conclude that R1 = R2. This completes the proof.

Observing Path I and Path I' in Fig. 8, since they have the same previous path from node 1 to node i, the port-id list of Path I is the same as the prefix of the port-id list of Path I'. Thus, once Path I' reaches node *i* again, path I's LPBS could be used as a matching pattern to check whether Path I' contains a loop or not. Following the same idea, the loop prevention algorithm of LFLDP is proceeded as follows. For an incoming route (identified by the incoming RID, denoted as IN-RID), current traversed label switch would append its incoming port id to the LPBS of IN-RID. The resulting LPBS denotes as IN-LPBS. Using the LPBS (denoted as RD-LPBS), which has RID = IN-RID in current label switch's LSRD, as a matching pattern to match the prefix of IN-LBPS. If matched, a loop is detected and a Route-Canceling message is then generated and sent back to the previous node.

Notably, LFLDP first uses metric value to screen route and the loop prevention algorithm is then proceeded to prevent loops. Furthermore, since each intermediate node would verify whether an incoming route has loop or not, LFLDP would find the loop whenever it is first generated. A loop is generated while a route reaches a node that has been traversed before. Thus, if a node *n* receives a route R containing a loop,

then R must have a form began with a loop-free path P and followed by a loop L. Note that routes R and P have the same originating and ending nodes (the ending node is node n). And they have the same RID. Consequently, we have the following lemma.

Lemma 2. *Let G (V, E) be a graph in which each node (label switch) runs LPLDP. For a node $n \in V$, node n* receives an incoming route R. If route R contains a loop, R must have a form began with a loop-free path P and followed by a loop L. Furthermore, the ending point of P is node n.

Lemma 3. *Let G (V, E) be a graph in which each node (label switch) runs LPLDP. For a node $n \in V$, node n* receives an incoming route R, which has passed the metric-value screening. Route R contains a loop if and only if there is a route R' in node n's LSRD. Where the LPBS of R' is the prefix of the LPBS of R and the RID of R' is the same as that of R.

Proof: Suppose R contains a loop, by Lemma 2, we have that R must have a form began with a loop-free path P and followed by a loop L. Furthermore, routes R and P have the same originating and ending nodes (the ending node is node *n*). Consequently, they have the same RID. If R has survived after the metric-value screening, P must be also in node n's LSRD because P' metric value is better than or equal to that of R and P arrives earlier. Thus, let P to be R' and the LPBS of P (i.e., R') is the same as the prefix of R's LPBS. Then, once R reaches node n, node *n* may use the LPBS of R' to verify whether R has a loop or not.

In the other hand, if the loop prevention algorithm finds a route R', which matches the prefix of R's LPBS. By definition, R and R' have the same RID. Thus, R and R' have the same originating and ending nodes. By applying Lemma 1, we conclude that route R' overlaps the previous path of R since the LPBS of R' matches the prefix of R's LPBS. Therefore, route R must have traversed node *n* before because the ending point of R' is node n. Furthermore, route R now backs to node *n* again and this generates a loop. This completes the proof.

Theorem 1. *A route established by LFLDP is loop-free.*

Proof. LFLDP first uses metric value to screen route and the loop prevention algorithm is then proceeded to prevent loops. There are two phase to screen a loop route. The first one is metric-value screening. The second one is loop prevention

algorithm. If a route contains a loop and owns a worse metric value, it would be sieved in the first phase. By Lemma 3, each loop route passing the first screening would be sieved by the loop prevention algorithm. Thus, all routes established via LFLDP are loop-free.

2. Implementation of NLPA

The proposed loop prevention algorithm using LPBS matching to verify loops could be simply implemented by hardware. The hardware implementation of the proposed loop prevention algorithm is depicted in Fig. 9. Owing to the remaining bits of LPBS are set to be 0, for distinguishing, the port id starts from 1 instead of 0. After concatenating the incoming port id, the IN-LPBS first masks with a bit string, which has a number of DB-LPBSL 1s and followed by 0s. The result is then compared with DB-LPBS. If equal, there is a loop; otherwise, the route is loop free.

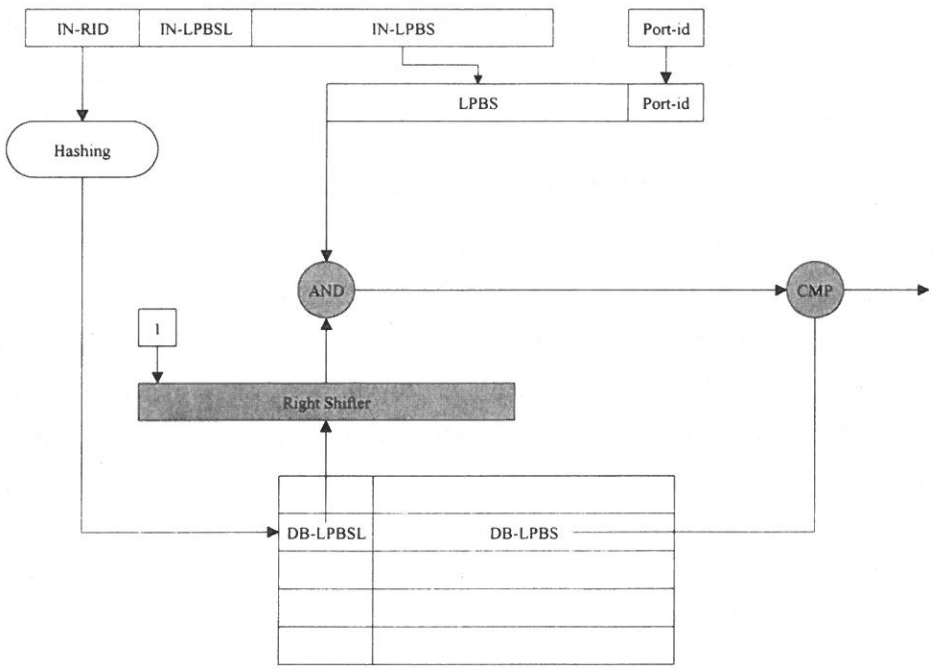


Fig. 9. The hardware implementation of NLPA.

3. Performance of loop prevention algorithm

The proposed loop prevention algorithm furnishes the following advantages.

- **Shorter message length.** MPLS employs node-id list to prevent loops [1]. For IP networks, node-id is commonly the IP address of the node. Suppose a route traverses 32 hops, the length of the node-id list becomes $32 \times 4 = 128$ and $16 \times 32 = 512$ bytes for IPv4 and IPv6, respectively. These are much longer than the length of LPBS (20 bytes).
- **No global unique id is required.** Node-id list requires a global unique id for each node. However, the proposed loop prevention algorithm using port-id list does not need a global unique id.
- **Constant processing time.** The proposed loop prevention algorithm is quite simple and could be implemented via hardware. The processing time of the proposed loop prevention algorithm is proportional to the number of entries in LSRD and it is always a small constant ($m = 2$ or 3). Thus, it takes only constant processing time to prevent loops. For node-id list scheme, the processing time is proportional to the number of nodes traversed.

CONCLUDING REMARKS

In this paper, a loop free label distribution protocol-LFLDP for label switching based IP networks has been proposed. In order to enhance network connectivity and flexibility, LFLDP provides multiple paths. We have also proposed a novel loop prevention scheme -NLPA, which employs a port-id list to prevent loops. The NLPA applies a constant processing time and a shorter message length. Furthermore, no global unique identifier is required for edge routers and switching nodes. The LFLDP provides only unicast routing; however, more and more applications and services require multicast routing. Thus, integrating multicast routing with LFLDP is going to investigate in the future study.

REFERENCES

(1) Arun Viswanathan and Ross Callon, "Multiprotocol Label Switching Architecture," Internet Draft <draft-ietf-mpls-arch-02.txt>, July 1998.

(2) R. Callon, P. Doolan, N. Feldman, A. Fredette, G. Swallow and A. Viswanathan, "A Framework for Multiprotocol Label Switching," work in progress, Internet Draft < draft-ietf-mpls-framework-02.txt >, November 1997.

(3) A. Viswanathan, N. Feldman, R. Boivie and R. Woundy, "ARIS: Aggregate Route-Based IP Switching," work in progress, Internet Draft < draft-viswanathan-aris-overview-00.txt>, March 1997.

(4) N. Feldman and A. Viswanathan, "ARIS Specification," work in progress, Internet Draft <draft-feldman-aris-spec-00.txt>, March 1997.

(5) Rekhter, Davie, Katz, Rosen, Swallow and Farinacci, "Tag Switching Architecture-Overview," work in progress, Internet Draft < draft-rekhter-tagswitch-arch-00.txt>, January, 1997.

(6) Doolan, Davie, Katz, Rekhter and Rosen, "Tag distribution Protocol," work in progress, Internet Draft <draft-doolan-tdp-spec-01.txt>, May, 1997.

(7) C. Partridge et al., "A Fifty Gigabit Per Second IP Router," To appear in IEEE/ACM Trans. On Networking.

(8) Butler Lampson, V Srinivasan and George Varghese, "IP Lookups using Multiway and Multicolumn Search," *IEEE INFOCOM'98*, March 1998.

(9) Pankaj Gupta, Steven Lin and Nick McKeown, "Routing Lookups in Hardware at Memory Access Speeds" *IEEE INFOCOM'98*, March 1998.

(10) A. Brodnik, S. Carlsson, M. Degermark and S. Pink, "Small Forwarding Tables for Fast Routing Lookups," *ACM SIGCOMM'97*, pp. 3-14, Cannes, France.

(11) M. Waldvoggel, G. Varghese, J. Turner and B. Plattner, "Scalable Hih-Speed IP Routing Lookups," *ACM SIGCOMM'97*, pp. 25-36, Cannes, France.

(12) Yakov Rekhter, Eric Rosen, Arun Viswanathan, Vijay Srinivasan and Steven Blake, "Use of Label Switching With RSVP," Internet Draft <draft-ietf-mpls-rsvp-00.txt>, March 1998.

(13) Braden, R. et al. "Resource ReSerVation Protocol (RSVP) —Version 1 Functional Specification," RFC 2205, Sep. 1997.

(14) Zappala, D. "RSRR: A Routing Interface For RSVP," S Internet Draft < draft-ietf-rsvp-routing-01.txt>, Nov. 1996.

received October 28, 1999
revised November 26, 1999
accepted December 8, 1999

標籤交換爲基礎之 IP 網路上標籤分配通訊協定

劉 惠 英
輔仁大學電子系

吳 俊 德
中華大學電機系

摘　要

由於網際網路的蓬勃發展，使得網路上傳輸的資料量大量的增加，解決網路頻寬成爲一個相當重要的課題，其中利用硬體交換的方式來提昇第三成路徑選擇的效能是一項極具前景的技術，此項技術被稱爲「第三成交換」，而「標籤交換」是第三成交換技術中一項重要的實作方式，其做法是在每筆封包的標頭加上一個固定長度的標籤作爲封包交換的依據，基於這樣的一個網路架構－標籤交換的 IP 網路，我們提出一個無迴路的標籤交換通訊協定，該方法提供多成選擇的路徑以提昇網路的連結性與彈性，並提出一個迴路避免的演算法，此演算法有下列三項優點：（1）所需的資料長度較短，(2）交換器或邊際路徑器不需要唯一的識別碼，(3）同時僅需常數處理時間。

關鍵詞： 第三層交換器，標籤交換，網際網路通訊協定，邊際路徑器