

FU JEN STUDIES

SCIENCE AND ENGINEERING

NO. 36, DEC. 2002

CONTENTS

	Page
Experiencing leisure: The case of Chinese university students by <i>Luo Lu and Chia-Hsin Hu</i> ...	1
High-Performance Class E Tuned Power Amplifiers for Wireless Communications.....by <i>Steve Hung-Lung Tu</i> ...	23
Design and Implementation of a Two-Dimensional DCT/IDCT Processor for Video Compression Systemsby <i>Shyue-Kung Lu and Chung-Yang Chen</i> ...	49
Supercapacitor Application in the Battery Energy System of the Electric Scooters.....by <i>Jeremy C. C. Chang and Yuang-Shung Lee</i> ...	73
An Evolving Bidding Strategy for Intelligent Agents in Multiple Auctionsby <i>Jong Yih Kuo</i> ...	93
Low Power Programming Design for Recharge Cycle Extension of a Mobile Computer.....by <i>Ying-Wen Bai and Chou-Su Chow</i> ...	113
A Practical Receipt-Free e-Voting Scheme Based on Smart Cardsby <i>Wei-Chi Ku and Chun-Ming Ho</i> ...	137
A Simulation Study of Confidence Intervals of Process Capability Index C_{pk} via Miss Rate.....by <i>Sy-Mien Chen and Hsiao-Wen Huang</i> ...	159
Executable Test Sequence for the Protocol Control and Data Portionby <i>Wen-Huei Chen</i> ...	179
Abstracts of Papers by Faculty Members of the College of Science and Engineering that appeared in the 2001~2002 Academic Year.....	201

輔仁學誌—理工類

中華民國九十一年十二月

第三十六期

目 錄

	頁次
休閒的主觀體驗—以台灣大學生為例	陸洛 胡家欣 ... 1
無線通訊所用之高性能 E 級功率放大器	杜弘隆 ... 23
使用於影像壓縮系統之二維離散餘弦正/反轉換器之架構設計與實現	呂學坤 陳忠陽 ... 49
超高電容器在電動機車電池能源系統中的應用	張佳祺 李永勳 ... 73
應用智慧型代理人漸進式拍賣策略於多重拍賣網站	郭忠義 ... 93
利用低功率程式設計延長行動電腦再充電周期	白英文 邱國書 ... 113
以智慧卡為基礎之無收據型實用電子投票系統	顧維祺 何俊明 ... 137
製程能力指標 C_{pk} 之信賴區間之擬研究	陳思勉 黃筱雯 ... 159
通訊協定控制與資料部分可執行測試序列產生方法之研究 ...	陳文輝 ... 179
91 學年度理工學院專任教師對外發表之論文摘要 201

Experiencing leisure: The case of Chinese university students

Luo Lu*

Department of Psychology

Fu-Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Chia-Hsin Hu

Graduate Institute of Behavioral Sciences

Kaohsiung Medical University

Kaohsiung, Taiwan 807, R.O.C.

Abstract

One hundred and fifty undergraduate students took part in group discussions regarding their subjective experiences of leisure. A combined qualitative and quantitative analysis revealed that (1) functionality, autonomy, and contrast with work were major components for defining leisure; (2) relaxation, enjoying life, self-growth, filling the time, social interaction, and health promotion were motivations for leisure; (3) biopsychological factors, activity itself, interpersonal factors, time, economy, and facility were both facilitators of and barriers to leisure; (4) relaxation, self-growth, enjoying life, social interaction, health promotion, filling the time, and costs were the actual the consequences of leisure.

Key words: meaning of leisure, leisure motivation, facilitators of leisure, barriers to leisure, consequences of leisure

*Corresponding author. Tel.: +886-2-29031111 ext. 3812; fax: +886-2-29010171
E-mail address: luolu@mails.fju.edu.tw (Dr. Luo Lu)

Experiencing leisure: The case of Chinese university students

Leisure is a very important topic, for many reasons. For instance, many people find their leisure more satisfying than their work; leisure can be a major source of pleasure and sense of achievement; we are having more leisure time than ever in the human history; however, many people who have a lot of spare time fail to find satisfying forms of leisure. Leisure has not really been accepted as a recognized field of psychological research, though there has been work on certain types of leisure, such as sports, musical activities, religious activities and watching TV. Research on leisure with Chinese people is even more in the rarity. Perhaps until very recently, leisure has never been accorded a significant status in the life of Chinese people either in a traditional subsistence economy or during the stage of economic take-off. However, things are changing. With relative material abundance and shrinking working hours, the vast population of the Chinese people is paying increasing interest in high quality and wide variety of leisure activities. Thus, understanding the subjective experiences of leisure will not only help to answer some interesting issues for psychology but also contribute ultimately to better leisure policies and practices. In the present paper, we will rely on qualitative data collected from Chinese university students in Taiwan to inform us on perceived meaning of leisure, personal motivations, facilitators of and barriers to leisure, as well as received effects of leisure. We concede that this effort is just a start, but a significant start in the bare field of the Chinese psychological study of leisure.

The meaning of leisure

In the Western literature, there has been a lot of learned discussion about the definition of leisure. Leisure has often been defined as time left over from work, and understood as activities contrasted with work. However, leisure is obviously not the entirety of free time, as Neulinger (1981) found that people regarded only one-third of their free time as leisure. Furthermore, unemployed people don't experience their free time as leisure, rather as "embarrassment" (Glyptis, 1989). The difference

between leisure and work is very subtle. For some people, the line between leisure and work is impossible to draw, for instance academics and artists, and some businessmen who carry on the public relation activities off-time. It seems that the only reliable difference is that leisure is not paid.

Recognizing the subjective element of leisure, other scholars have attempted to define leisure as a unique personal state of mind, or an evaluation of an activity (deGrazia, 1962). Emphasizing perceived freedom and internal motivation, Iso-Ahola (1976, 1997) asserted that leisure can only been subjectively defined and personally recognized. Encompassing the above discussion, Argyle and Lu (1992) proposed a comprehensive definition of leisure as “what people choose to do in their spare time, for its own sake, because it is enjoyable, or felt to be intrinsically worthwhile, but not for any external reward” (p.5). However, there are unresolved issues lingering still. For instance, do people really have the freedom to choose leisure activities and actually engage in them? In other words, are there facilitators of and barriers to leisure? Which aspect of leisure is most emphasized in people’s perception of its meaning? Our present study was designed to offer some answers to these questions.

Leisure motivation

This is an important aspect of the study of leisure because it can partially explain why people engage in leisure activities, or particular types of leisure activities. Examining leisure motivation can also inform the psychological study of motivation in general, especially intrinsic motivations. In a study with Australian adults, Kabanoff (1982) listed ten leisure motivations: autonomy, relaxation, family activity, escape from routine, social interaction, stimulation, skill utilization, health, self-esteem, challenge/competition, and leadership/social power.

For more serious forms of leisure, such as rock climbing and sky diving, where carousal efforts, long-term devotion and physical extremity are often involved, intrinsic motivations are most prominent. Csikzentmihalyi (1975) interviewed 173 individuals who engaged in serious leisure activities, and found that the reasons were:

enjoyment of the experience and use of skills; the activity itself—the pattern, the action, the world it provides; development of personal skills. Csikszentmihalyi regarded these as the main sources of intrinsic motivation, though other reasons were mentioned too—friendship and relaxation, risk and chance, problem solving, competition and creativity. For devoted viewers, watching “soap opera” is arguably a serious leisure. Livingstone (1988) found that reported motivations of committed viewers were quite varied: entertainment and escapism, realism, characters as extension of real-world social networks, as an educational medium, as part of daily life, and emotional experience. In a rare study of the idolatry phenomenon, Ju and Lu (2000) found that young Chinese fans of popular music engaged in idolatry behavior for reasons of self-presentation, social learning, vicarious satisfaction, emotional expression, and nostalgic memory.

To sum, social motivation, physical excitement, and social learning seem to be the most important elements of intrinsic motivation for those who engage in serious committed leisure. However, in mundane and trivial daily leisure activities such as watching TV (non-serious viewers of particular programs) and singing in KTV, simple relaxation and casual interaction may be more prominent motivations for a much wider section of the population. It may be more so for the Chinese people, as they are yet to fully recognize the value of leisure (Chen, 1989), and many have no committed forms of leisure (Chen, 1997). Thus, we hoped to delineate general leisure motivation of the Chinese students with their own accounts.

Facilitators of and barriers to leisure

Facilitators of leisure are usually conceived as the opposite of barriers to it and most extant research focus on the latter. Kay and Jackson (1991) noted money and time as the two most significant constraints of leisure. Later Jackson (1993) identified six dimensions of barriers among a large community sample: social isolation, accessibility, personal reasons, costs, time, and facilities. Furthermore, Alexandris and Carroll (1997) found that perceived barriers to leisure were indeed related to reduced engagement in leisure.

There has been relatively more research on this topic with the Chinese population. For instance, Chang (1998) found that university students reported intrapersonal, interpersonal as well as structural barriers to leisure. Li (1997) noted family responsibility, lack of facilities, and lack of companions as the three major barriers to leisure for urban women.

It seems that various factors may be perceived as barriers to leisure, including money, time, resources and accessibility, ability, conflicts with family and work. However, the relative importance of these factors may vary for different people, and it would be interesting to reveal the specific list for young students. It would also be interesting to empirically test the implicit assumption that facilitators of leisure are the opposite of barriers and to uncover possible distinct facilitators or barriers.

Consequences of leisure

If leisure were what people truly and freely choose to do, then it would be very odd if they didn't enjoy it. However, if we allow people to define the "leisure" as they will, then it is possible to enquire whether it does in fact produce positive effects, or even negative effects at time.

Leisure satisfaction and happiness are perhaps the most direct indicators of leisure effects. Andrew and Withey's (1976) early US survey showed that most people (43%) were pleased or delighted with their leisure, and only a small minority (8.5%) was dissatisfied. However, different leisure activities may generate different levels of satisfaction. Lu and Argyle (1994) found that people reported greater leisure satisfaction and happiness when they had a serious, committed, and constructive leisure activity. They experienced their leisure as more stressful, challenging and absorbing, but more under control. In contrast, less serious leisure activities, such as TV watching, have been found to produce less positive effects. Lu and Argyle (1993) noted that people who watch a lot of TV were bored more often, had lower leisure satisfaction, and less happy. Nonetheless, the same study also found that regular "soap opera" watchers were more satisfied with their leisure and happier, perhaps because this activity has some elements of a serious leisure.

In addition to satisfaction and happiness, leisure has been found beneficial for mental health, perhaps through its stress buffering effects (Coleman & Iso-Ahola, 1993). Social integration, companionship, and social support are most pronounced effects of leisure (Argyle & Lu, 1992) and beneficial to mental health. Needless to say, some forms of leisure, such as sports, can greatly enhance physical health (Wannamethee, Shaper & Macfarlane, 1993). In addition, physical and mental health has a mutually effecting relationship (Lu & Hsieh, 1997).

Thus far, existing leisure research has focused almost exclusively on positive effects and found short-term benefits including positive mood, physical fitness and immediate satisfaction, as well as long-term effects of happiness, mental health, physical health, and social integration. In the present study, we aimed to find out which benefits were actually perceived by students, and whether they perceived less pleasing even negative consequences of leisure.

Method

Participants

One hundred and fifty undergraduate students, age 19 to 25, participated in the study. The students were in their second to fourth year, enrolling on a health psychology and a social psychology course at a medical university in Taiwan. These students freely formed 26 groups for class discussion (3-8 in a group).

Procedure

To reflect the exploratory nature of the present study, a qualitative approach was adopted. Data were collected through group discussions that took place in fall 1999. Participants were instructed to freely and thoroughly discuss four open-ended questions: (1). What is the meaning of leisure to you? Please give a definition of leisure. (2) Why do you engage in leisure activities? (3) What factors facilitates your leisure participation, and what factors prevents you from leisure participation? (4)

What are the consequences of leisure participation? Each group handed in a point-to-point record, which formed the basis for later data analysis.

All the transcribed verbatim were first coded using thematic analysis to create master schemes of categorization for each of the four issues discussed. This allows the emergence of concepts, constructs, and typologies directly rooted in data, reflecting the essence of grounded theory practice (Strauss & Corbin, 1997). Frequency counts were then made for each subcategory in the master schemes, following the usual practice of content analysis. However, similar responses by different participants were counted only once. The combination of thematic analysis and content analysis enabled the development of new organizing schemes firmly supported by data as well as the presentation of relative prominence of subjective experiences. In other words, both the qualitative and quantitative aspects of the leisure experiences were explored for this group of students.

The thematic analysis were mainly conducted by the two researchers, but a regular research group composed of scholars of psychological, sociological, and social work background provided insightful and constructive inputs in interpreting the data, clarifying the constructs, and elaborate the schemes. The results of this series of analyses were then communicated back to the participants at a later class session. The researchers explained their construction of the master themes, presented results of the frequency calculation, elaborated on the major findings, and answered any queries. This communication exercise served as a debriefing act as well as a validation practice. Participants were invited to actively engage as co-researchers too. They largely agreed with researchers' interpretations of and conclusions drawn from the data, and enthusiastically engaged in the discussion. This procedure of data analysis hence ensured the "trustworthiness" of the study, through the intersubjectivity between the researchers and their academic peers, as well as between the researchers and participants.

Results and Discussion

Results are presented below and grouped into categories and subcategories representing the main themes emerged for each issue. Numbers in brackets are frequency counts for each category and subcategory.

The meaning of leisure

Usually a definition of leisure included several aspects. We discerned three major components: functionality, autonomy, and contrast with work. The diagram in Figure 1 shows the frequencies these components were mentioned when participants attempted to define leisure.

Category 1: Functionality (30) Leisure was defined in terms of its positive function to the individual. For instance, “leisure is to make myself happy”, “leisure is a kind of liberation”, and “leisure is to get rest and loose up”.

Category 2: Autonomy (25) Leisure was defined as things one can freely choose to do. For instance, “leisure is time I can freely use”, “leisure is things I want to do”, and “leisure is no constraint and totally free”.

Category 3: Contrast with work (18) Leisure was defined as the opposite of work. For instance, “leisure is time after work”, “leisure is things I do which are unrelated to work”, and “leisure is completely different from work”.

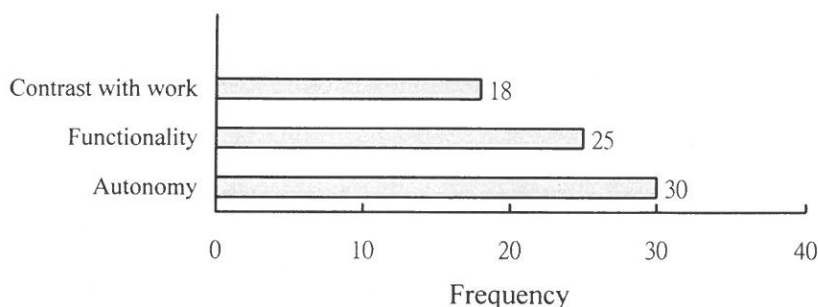


Figure 1. Frequency for components in leisure definition.

From Figure 1, we can see that functionality and autonomy were the most often emphasized components of leisure. The emphasis on functionality is in consonant with the view that people engage in leisure “because it is enjoyable, or felt to be intrinsically worthwhile” (Argyle & Lu, 1992, p.5). Leisure is indeed pursued for its positive value and generally regarded as pleasurable activities. The emphasis on autonomy is consistent with views of leading scholars in the field (Csikszentmihalyi, 1975; Iso-Ahola, 1976, 1997), which stress the importance of perceived freedom in leisure. Leisure is not only pleasurable but also under personal control, and is a manifestation of human agency. Leisure pursuits have been suggested as a way of self-expression (Little, 1983; Ju & Lu, 2000), and have implications for identity development (Argyle & Lu, 1992). The exercise of autonomy in leisure has also been suggested as a counter dose for stress, which often threatens personal control (Coleman & Iso-Ahola, 1993). However, the exact mechanism of applying autonomy to leisure construction and factors enhancing or hampering autonomy need to be delineated more systematically in the future.

Also consistent with learned views in the field, leisure is often held in contrast with work. May be for our young students, the line between leisure and work is relatively easy to draw, and even consciously maintained. Some students volunteered in the feedback discussion that they used leisure as a safe heaven to escape from academic pressure. It is also a common practice in the campus to celebrate the end of exams with some frantic leisure activities. We thus confirmed that at least for some people leisure is conceived of as something different from work or even an escape from work.

However, as stated early, our students typically defined leisure with a combination of two or even three components. For example, one participant stated that “leisure is no constraint and totally free, it’s a kind of liberation”. It is obvious that any one of the elements is not sufficient to encompass the leisure experience. In other words, different scholarly views in the field all have some of the truth in reality, but none is encompassing the whole truth yet. With this valuable insight into lay people’s conception of leisure, we should modify Argyle and Lu’s (1992) definition

to make it even more comprehensive: “leisure is what people choose to do in their spare time, to escape or to be different from work, for its own sake, because it is enjoyable, or felt to be intrinsically worthwhile, but not for any external reward”.

Leisure motivation

Most students gave several leisure motivations. We discerned six major motivations: relaxation, enjoying life, self-growth, filling the time, social interaction, and health promotion. The diagram in Figure 2 shows the frequencies these motivations were mentioned by participants.

Category 1: Relaxation (61) Leisure was pursued for its value in relaxation. More actively, some students engage in leisure activities to build up resources to cope with stress. For instance, one student claimed that his work efficiency was improved after playing sports. Less actively, students engage in leisure activities simply to get rest and get the mind off.

Category 2: Enjoying life (13) Leisure was actively pursued for its hedonic value. For instance, “I engage in leisure to seek happiness”, and “I take part in leisure to lead an interesting life”.

Category 3: Self-growth (10) Leisure was actively used to acquire new knowledge, learn new skills, and achieve self-growth. For instance, “leisure is to accumulate new knowledge”, and “leisure is to develop a wider interest”.

Category 4: Filling the time (10) Leisure was used to handle the vacuum of tasks and responsibilities. For instance, “I engage in leisure because there is nothing else to do”, and “I take part in leisure to kill time”.

Category 5: Social interaction (9) Leisure was actively used to promote social integration and communication. For instance, “I engage in leisure to maintain my social relationships”, and “I take part in leisure to make new friends”.

Category 6: Health promotion (5) Leisure was actively used to enhance or maintain health. These leisure activities are usually sports and exercise. For instance, “leisure is to build up physical strength”, and “leisure is to help me lose weight”.

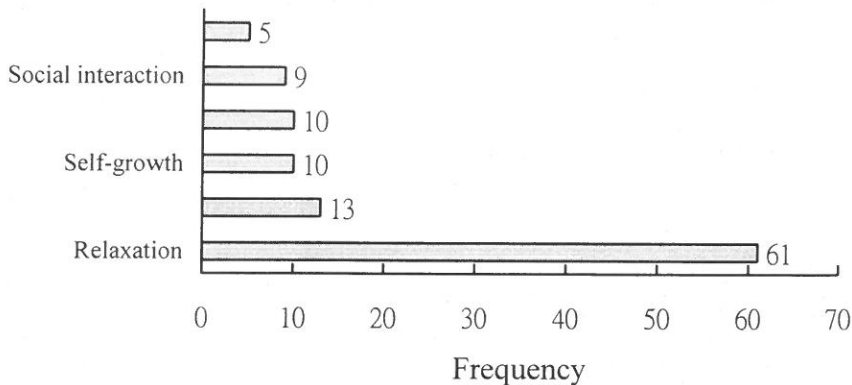


Figure 2. Frequency for leisure motivation.

Two of our findings here should be noted. First, relaxation is the most prominent leisure motivation for our students. Second, filling time is the only one not on Kabanoff's (1982) list of ten leisure motivations. These two features may be embedded in a common cultural milieu. Seeking relaxation is considered a minor motivation for people with committed leisure (Csikzentmihalyi, 1975). However, if people engage in leisure simply to kill time and banish boredom, they most likely don't have a serious committed leisure activity and often indulge in more passive forms of leisure such as watching a lot of TV, which end up producing less satisfactory results (Lu & Argyle, 1993). Empirical research with the Chinese people has shown that the actual participation in leisure is very low (Chen, 1997; Tu, 1998), not to mention serious leisure. Perhaps as we expected, for the Chinese people who are still in an early stage of forming a constructive culture of leisure (Chen, 1989), seeking simple relaxation is no less important than other deeper and more seasoned intrinsic motivations as emphasized in the Western culture. In particular, relaxation as a leisure motivation may be related to the conception of leisure as a counter dose to, or escape from academic stress among students.

Facilitators of and barriers to leisure

Facilitators of and barriers to leisure were conceived as opposites to each other, and were encompassed by a single set of categories. These factors were: biopsychological factors, activity itself, interpersonal factors, time, economy, and facility. The diagrams in Figure 3 and 4 show the frequencies these factors were mentioned by participants. For easy comparison, facilitators of and barriers are presented together below.

Category 1: Biopsychological factors (29/11) Subjective biological and psychological states can either facilitate or inhibit leisure participation. For instance, good mood, end-of-exams high, longing for a long time, and optimal physical state were generally regarded as facilitators. On the other hand, for instance, stress, impending exams, certain personality traits and lack of ability were regarded as barriers.

Category 2: Activity itself (27/11) Positively or negatively perceived characteristics inherent in a particular leisure activity might become facilitating or inhibiting factors to participation. On the positive side, for instance, if a certain activity can open up new horizon, offer a sense of mastery and achievement, or improve body image, these may facilitate participation. On the negative side, for instance, lack of interest, and extensive requirement of physical strength may become inhibiting factors.

Category 3: Interpersonal factors (26/18) Positively or negatively perceived social influences may become facilitating or inhibiting factors to participation. On the positive side, for instance, invitation from friends and approval of parents may facilitate participation. On the negative side, for instance, lack of companions, being pushed by friends, and undesirable persons on the scene may become inhibiting factors.

Category 4: Time (20/20) Availability or lack of time may become facilitating or inhibiting factors to participation. On the positive side, for instance, having free time, and being on holidays may facilitate participation. On the negative side, for instance, lack of free time, and having too much work to do may become inhibiting factors.

Category 5: Economy (13/23) Personal economic circumstances as well as monetary costs of leisure activities may become facilitating or inhibiting factors to participation. On the positive side, for instance, having extra money to spend, and budget activities may facilitate participation. On the negative side, for instance, lack of extra money, and expansive activities may become inhibiting factors.

Category 6: Facility (7/22) Availability or lack of necessary facilities and supporting logistics for leisure may become facilitating or inhibiting factors to participation. On the positive side, for instance, convenient venue, comfortable space, and easy transportation may facilitate participation. On the negative side, for instance, lack of venue, inconvenient transportation, and insecure environment may become inhibiting factors.

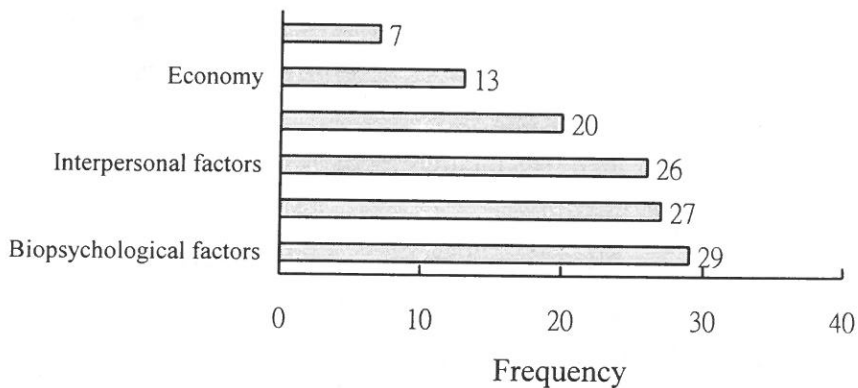


Figure 3. Frequency for facilitators of leisure.

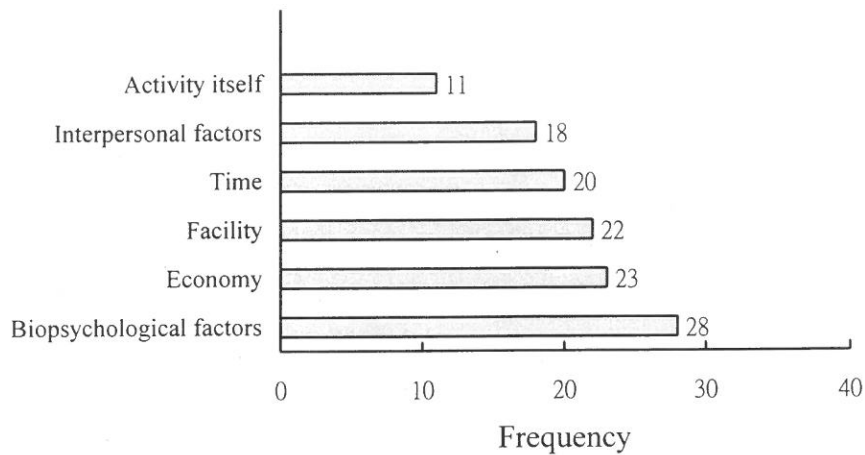


Figure 4. Frequency for barriers to leisure.

Our list of 6 barriers to leisure was almost identical to Kay and Jackson's (1991) list, and can be read as a detailed account of Chang's (1998) list of intrapersonal, interpersonal and structural factors. However, conflicts with family responsibility were not prevalent for students for obvious reasons.

Our analysis has confirmed that facilitators of leisure are in general the opposite of its barriers. However, the relative importance of the same factor as a facilitator or a barrier varies. For facilitators, biopsychological factors, activity itself, and interpersonal factors were the top three most prominent factors. For barriers on the other hand, activity itself was the least prominent factor while the remaining five were almost indistinguishable in their importance. Basis for this disparity needs further exploration.

Consequences of leisure

Most of the consequences mentioned by students could fit into the same categorization scheme for leisure motivations. One distinct feature though was negative consequences, which we termed "costs". We thus discerned seven major consequences of leisure: relaxation, self-growth, enjoying life, social interaction,

health promotion, filling the time, and costs. The diagram in Figure 5 shows the frequencies these consequences were mentioned by participants.

Category 1: Relaxation (43) Leisure was perceived to lead to relaxation, for instance, releasing stress and recovering strength.

Category 2: Self-growth (31) Leisure was perceived to lead to self-growth, for instance, acquiring new knowledge, and facilitating academic work.

Category 3: Enjoying life (18) Leisure was perceived to lead to hedonistic enjoyment of life, for instance, inflated happiness, and a more interesting life.

Category 4: Social interaction (15) Leisure was perceived to lead to improved social integration, for instance, maintaining relationships, promoting friendships, and increasing social participation.

Category 5: Health promotion (10) Leisure was perceived to lead to the enhancement of health, for instance, enjoying the pleasure of sweating, and promoting health.

Category 6: Filling the time (5) Leisure was perceived to lead to better structuring of free time, for instance, filling up free time, and getting rid of boredom.

Category 7: Costs (5) Leisure was perceived to have various negative consequences, for instance, getting even more tired, and reduction in sleep time.

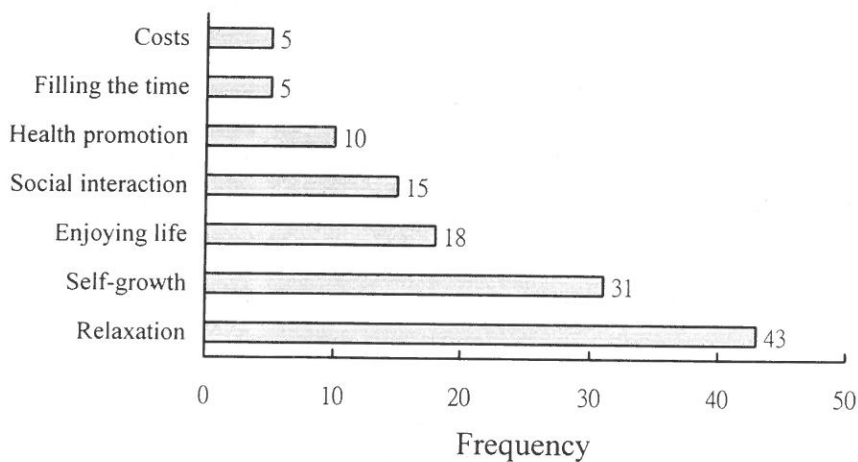


Figure 5. Frequency for consequences of leisure

It seems that regardless whether people have a serious, committed leisure activity or not, most people were pleased with their leisure, as found by Andrew and Withey's (1976) and Lu and Argyle (1993, 1994). The overwhelming perception of positive results of leisure is understandable, after all leisure is done out of free choice largely, in free time, and usually under high expectations to generate a pleasurable state of mind. Reading Figure 5, we can see that leisure indeed was perceived to have short-term benefits including positive mood, physical fitness and better structure of time, as well as long-term effects of happiness, health, educational benefits, and social integration. The use of leisure to structure time may be particularly useful for certain sections of the population, such as housewives, the elderly and unemployed. As we have seen earlier, free time does not equate meaningful leisure (Neulinger, 1981; Glyptis, 1989). However, helping those people to plan leisure constructively can not only provide a more manageable time structure, but also bring about a whole array of positive effects such as social integration, health promotion, self-growth, and happiness.

However, unlike the rosy picture painted by the existing research, our present study noted that students did perceived some less pleasing even negative consequences of leisure. We suspect that these costs of leisure may encompass many aspects, for instance, physical (overuse of physical energy), economic (budget pressure), work (competing priority), social (interpersonal friction) aspects. Few early studies did found that watching TV put people in a drowsy, weak, and passive psychological state (Csikzentmihalyi & Kubey, 1981), and heavy TV watchers were bored, dissatisfied, and unhappy (Lu & Argyle, 1993). It is imperative then to more systematically explore the negative sides of leisure, in order to rectify theoretical blindness as well as to inform leisure management practices.

Conclusion

This study probed into the subjective experiences of leisure, in particular the conceived meaning of leisure, reported leisure motivation, perceived facilitators of

and barriers to leisure, as well as actual consequences of leisure. Some of our findings were consonant with the views and results in the extant literature, while others extended or challenged the established accounts about leisure. Whether these distinct findings are due to sample characteristics (e.g. students), social characteristics (e.g. under-developed culture of leisure), or cultural characteristics (e.g. Chinese views of leisure) need detailed examination in the future. Although our effort is just a start, we hope that data we collected will contribute to the development of a Chinese psychology of leisure.

Authors' Note

The first author writes the present paper, while empirical data were drawn from part of those in the second author's Master thesis, which was completed under the supervision of the first author.

References

- (1) Alexandris, K. & Carroll, B. (1997) Perception of constraints and strength of motivation: Their relationship to recreational sport participation in Greece. Journal of Leisure Research, 29, 279-299.
- (2) Andrew, F.M. & Withey, S.B. (1976) Social indicators of well-being. New York: Plenum.
- (3) Argyle, M. & Lu, L. (1992) New directions in the psychology of leisure. The New Psychologist, 1, 3-11.
- (4) Chang, Y.L. (1998) Relationships among intrinsic leisure motivation, barriers to leisure, leisure boredom, and self integration for university students. A Master thesis at Institute of Guidance, Kaohsiung Normal University
- (5) Chen, C.Y. (1989) Work and leisure: An examination of leisure theories and research from the industrial psychological perspective. Taipei: Shu-Hsiun Books.
- (6) Chen, M.L. (1997) An examination of leisure attitudes, leisure involvement and related factors among adults in Chia-Yi area from an adult development perspective. Chia-Yi: Institute of Adult Education, Chong-Cheng University.
- (7) Coleman, D. & Iso-Ahola, S.E. (1993) Stress, social support, and health: the role of social support and self-determination. Journal of leisure research, 25, 111-128.
- (8) Csikszentmihalyi, M. (1975) Beyond boredom and anxiety. San Francisco, fl: Jossey-Bass.
- (9) Csikszentmihalyi, M. & Kubey, R. (1981) Television and the rest of life. Public Opinion Quarterly, 45, 317-328.
- (10) De Grazia, S. (1962) Of time, work, and leisure. New York: Twentieth Century Fund.
- (11) Glyptis, S. (1989) Leisure and unemployment. Milton Keynes, UK: Open University Press.
- (12) Iso-Ahola, S.E. (1976) On the theoretical link between personality and leisure. Psychological Reports, 39, 3-10.

- (13) Iso-Ahola, S.E. (1997) Leisure-related social support and self-determination as buffer of stress-illness relationship. Journal of Leisure Research, 28, 169-187.
- (14) Jackson, E.L. (1993) Recognizing patterns of leisure constraints: results from alternative analyses. Journal of Leisure Research, 25, 129-149.
- (15) Ju, L.S. & Lu, L. (2000) An exploratory study of attitudes and behaviors of idolatry among fans of popular songs. Research in Applied Psychology, 8, 171-208.
- (16) Kabanoff, B. (1982) Occupational and sex differences in leisure needs and leisure satisfaction. Journal of Occupational Behavior, 3, 233-245.
- (17) Kay, T. & Jackson, G. (1991) Leisure despite constraint: The impact of leisure constraints on leisure participation. Journal of Leisure Research, 23, 301-13.
- (18) Li, S.H. (1997) Leisure types and barriers to leisure among urban women. Research on Outdoor Recreation, 10, 43-68.
- (19) Little, B. (1983) Personal projects: A rationale and method of investigation. Environment and Behaviour, 15, 273-309.
- (20) Livingstone, S.M. (1988) Why people watch soap opera: an analysis of the explanations of British viewers. European Journal of Communication, 3, 55-80.
- (21) Lu, L. & Argyle, M. (1993) TV watching, soap opera and happiness. Kaohsiung Journal of Medical Sciences, 9, 350-360.
- (22) Lu, L. & Argyle, M. (1994). Leisure satisfaction and happiness as a function of leisure activity. Kaohsiung Journal of Medical Sciences, 10, 89-96.
- (23) Lu, L. & Hsieh, Y.H. (1997) Demographic variables, control, stress, support and health among the elderly. Journal of Health Psychology, 2, 97-106.
- (24) Neulinger, J. (1981) The psychology of leisure (2nd ed.). Springfield, Ill: Charles C. Thomas.
- (25) Strauss, A. & Corbin, J. (1997) Basics of qualitative research: Grounded theory procedures and techniques. London: Sage.
- (26) Tu, H.W. (1998) Relationships among personality traits, social relationships, and happiness among junior middle school students. A Master thesis at Institute of Education, Kaohsiung Normal University.

- (27) Wannamethee, G., Shaper, A.G. & Macfarlane, P.W. (1993) Heart rate, physical activity, and mortality from cancer and other noncardiovascular disease. American Journal of Epidemiology, 137, 735-48.

received September 29, 2002
revised October 12, 2002
accepted November 2, 2002

休閒的主觀體驗—以台灣大學生為例

陸洛

輔仁大學心理系

胡家欣

高雄醫學大學行為科學研究所

摘 要

150 位大學生參加了有關其主觀休閒體驗的團體討論，以質、量並重的分析方法發現：（1）「功能性」、「自主性」、「與工作的對比」是休閒定義的三大主要成分；（2）放鬆減壓、享受生活、自我成長、填補空閒、社會互動、及促進健康是休閒的主要動機；（3）生心因素、活動本身、人際因素、時間因素、經濟因素、及設施因素既可能是休閒活動的促進因子，也可能是阻礙因子；（4）放鬆減壓、自我成長、享受生活、社會互動、促進健康、填補空閒、及活動負效則為休閒的後果。

關鍵詞：休閒定義、休閒動機、休閒促進因子、休閒阻礙因子、休閒後果

High-Performance Class E Tuned Power Amplifiers for Wireless Communications

Steve Hung-Lung Tu*

Department of Electronic Engineering

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

This paper introduces several configurations of Class E power amplifiers in CMOS technologies. Each configuration, however, alleviates some problems in the design of Class E power amplifiers. The two-stage Class E power amplifier reveals the design technique for the driving stage, which provides more efficient driving signal in terms of Class E operation. The complementary configuration takes advantage of the symmetrical circuit topology which allows much lower total harmonic distortion (THD) in the output signal. Exploration of the relationship between the switching device and the loaded quality factor of the load network depicted in this paper provides a useful guide in designing the load network as well as the switching device.

Key Words: Class E, tuned power amplifier, loaded quality factor, complementary.

*Corresponding author. Tel.: +886-2-29031111 ext. 2427; fax: +886-2-29042638
E-mail address: stu@ee.fju.edu.tw

1.Introduction

The personal communications industry has seen high-speed growth over the past several years. Combining computational and communications facilities in a portable unit has become the trend for future personal communications systems (PCS). Massive computations such as data compression algorithms are needed to reduce the large amount of communication data. Meanwhile, sophisticated modulation schemes are also used in portable communications to extend bandwidth. The power needed to drive these functions can be easily prohibitive for portable operations. Therefore, the goal of low power design is not only desirable but also necessary.

In the wake of the progress of CMOS technologies, low power design for digital signal processing (DSP) is feasible due to the low power consumption of small feature-sized CMOS devices. However, the analog interface for the portable applications is still the design bottleneck since analog signals should be transmitted by the use of high power to allow communication for long distance. Moreover, because of spectrum congestion, personal communications systems must operate at carrier frequencies above 1GHz and the resulting complexity and difficulties in implementation imply that simplifying the circuitry or relaxing the required analog performance should be paramount.

Traditionally, for gigahertz-band RF circuitry, discrete Gallium Arsenide (GaAs) transistors have dominated these design. However, significant area and large power consumption in driving high-speed analog signals across board interconnects make GaAs technology not the best option for wireless communications. CMOS technologies, by contrast, seem to be more promising since they have seen a breakthrough in both performance and size through the use of device scaling. Furthermore, silicon CMOS technologies will continue to scale rapidly owing to the demand for high-performance and denser digital circuits such as microprocessors and memory chips.

A challenging functional block in designing such a wireless communication

transceiver is the power amplifier due to the trade-offs between supply voltage, output power, power efficiency, and distortion. Meanwhile, the basic requirement for the power amplifier is the ability to work at low supply voltages as well as high operating frequencies. As a result, they have deleterious effects on the power efficiency. Theoretically, Class C, D, and E nonlinear power amplifiers all have 100% power efficiency[1]. However, in practice, few of them can achieve this target particularly when implemented as integrated circuits in CMOS technologies. In addition, with the progress of the CMOS device fabrication technology and on-chip passive component requirements, accurate and suitable modeling is becoming highly important.

The emphasis throughout this paper mainly focuses on both analyses and design aspects of Class E power amplifiers in CMOS technologies for wireless communications. With a thorough mathematical treatment for each configuration of the Class E power amplifier, we can, therefore, give a more useful guide in the design efforts.

2. General principles of Class E operation

An ideal Class E amplifier configuration is shown in Fig. 1, which consists of a single supply voltage V_{DD} , an RF choke inductor, a switch (generally using bipolar transistors or FET's) with a parallel capacitor, a resonant circuit, and a load R_L .

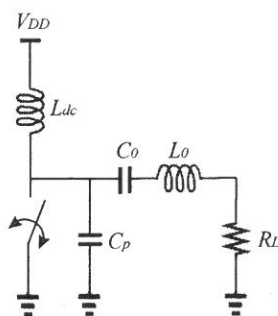


Fig.1. Ideal Class E amplifier configuration

The switch is turned on and off periodically at the input frequency. L_0 - C_0 resonates at the input frequency and only passes a sinusoidal current to the load R_L .

C_p ensures that in the time when the switch is turned off the voltage across the switch still keeps relatively low until after the drain current has reduced to zero. The switch usually uses active devices such as silicon bipolar transistors or FET's. In practice, in order to make the switch near ideal and reduce the on-resistance, the transistor is designed with a large gate width.

Maximum output power can be obtained if the duty ratio of the input frequency is made approximately 50 percent [1]. Since the operating frequencies are different for the on state and the off state of the active device, the load network may include filters to suppress harmonics of the output signal. The well-known Class E switching conditions [2] include:

1). *Voltage return to zero at the switch turnon*: This ensures that the voltage of the switch and the current flowing through it cannot happen simultaneously, and thereby the power dissipation in the switch is zero.

2). *Zero voltage slope at the switch turnon*: Although the former point can be satisfied with proper circuit design. The condition of slight mistuning of the amplifier may happen. This point can prevent severe power loss at the transient point.

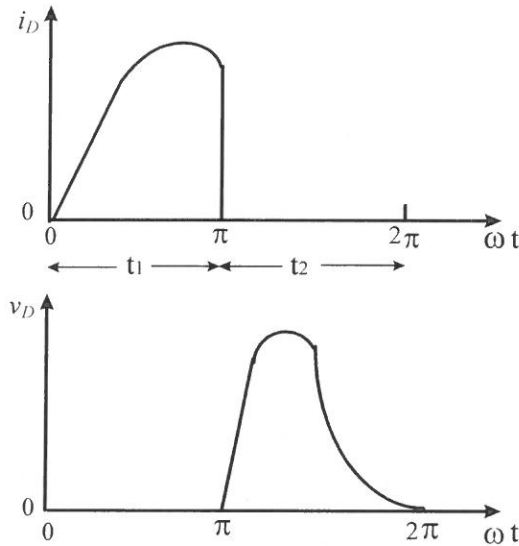


Fig. 2. Ideal Class E voltage and current waveforms.

Based on the Class E switching conditions, the ideal Class E voltage and current waveforms can be illustrated in Fig. 2. t_1 is the period of the switch closed and t_2 is the period of the switch open. During t_1 period, v_D is zero whereas during t_2 period the current i_D is zero. Because of the characteristic of nonoverlap of the current and the voltage, the power dissipation of the switch is zero.

3.Circuit parameters on power efficiency

Since the Class E power amplifier was proposed in 1975 by N. Sokal[2], the design involves much empirical knowledge. M. Kazimierczuk[3]-[7] and F. Raab[1], [8]-[10] did a number of insight analyses concerned with the topic and addressed many useful relationships between the circuit parameters in terms of output power, power efficiency and harmonic distortion. More recently, the Class E power amplifier has been demonstrated in the application of wireless communications at gigahertz band with an advanced GaAs technology[11]. However, few of these recently published papers presented design equations or guidelines.

The conventional assumption of the infinite choke inductance needs to be revised at high operating frequency, which it, in turn, leads to different analytical methods and results. Optimum performance under the assumption of finite choke inductance was described approximately by the following equation[12],

$$L_{dc} \approx \frac{0.7436}{\omega^2 C_p} \quad (1)$$

where ω is the operating frequency, L_{dc} and C_p ¹ are the choke inductance and the shunt capacitance, respectively as shown in Fig.1. L_0 and C_0 act as an ideal resonator operating at the frequency of ω . Notice that when it operates at much higher frequencies with a limited L_{dc} value, the C_p value can only be the nonlinear parasitic drain-to-substrate capacitance of the MOS switching device. The analysis

¹ The capacitance is the sum of the drain-to-substrate parasitic capacitance and the shunt capacitance.

for the effect can be found in [13] which reveals that the peak voltage across the transistor increases and the normalized power capability decreases. This can result in the junction breakdown and the effect is even worse for deep-submicron CMOS technologies.

The power loss of the Class E power amplifier is relevant to the switching behavior of the device. Since the current flowing through the device is negligible at the transition of turn-on, the power loss at this period is much smaller than the transition of turn-off. Therefore, the power efficiency can be investigated from the switch-off process of the device.

M. Kazimierczuk discussed the effects of the collector current fall time of the transistor on power efficiency assuming the current is linear decays after the transistor is switched off[4]. Blanchard and Yuan extend the analysis by assuming the current is an exponential decay during the fall time[14]. Based on these important assumptions, Tu and Toumazou[15] further investigated the power efficiency at the operating frequency ω in terms of the loaded quality factor which is defined as[2]

$$Q_L = \frac{\omega L_0}{R_L} \quad (2)$$

The investigation in[16] illustrated that no configuration of linear circuit elements can provide zero voltage and zero current at both transitions of the switch if nonzero output power is to be delivered to a load. Hence, a jump of current or voltage must be tolerated at the switch turn-off or turn-on. As a result, the output current/voltage signals shall present a certain degree of distortion. The choice of the loaded quality factor Q_L is a compromise between several factors including low harmonic output (high Q_L), low power losses (low Q_L), low changes of the amplifier parameters with frequency (low Q_L), and narrow bandwidth (high Q_L)[4]. Obviously, the chief conflict exists between low harmonic output and low power losses.

However, with the advantages of the state-of-the-art CMOS technologies, the transition behaviors of the switching devices may alleviate the problem. Two different feature-sized transistor HSPICE models are employed to compare the

power efficiency in terms of the loaded quality factor. The simulation results are shown in Fig.3. The power efficiency decreases with an increase of Q_L . For the 0.6- μm model, since it has lower decay angle[14], the power efficiency is also higher. Comparing the results to the 0.8- μm model, it shows much slower decrease with an increase of Q_L .

The investigation indicated that given the continued advances in scaling of silicon MOS technologies, it is evident that high power efficiency CMOS power amplifier is possible in the near future as nanoelectronic technologies become commercially available.

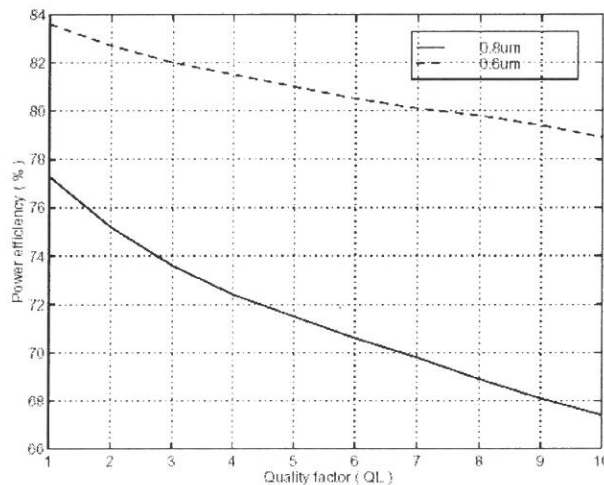


Fig.3. Power efficiency versus Q_L for two different feature-sized models

4. Design of driving stage for Class E operation

A practical problem emerged when we design a Class E power amplifier - a driving stage is actually required in order to turn on and off the switching devices of the final stage efficiently. One method to meet this requirement is to reduce the input load for the precedent functional block, mixer, and the other is to establish a more efficient driving signal to reduce the switching loss. Here we may focus on discussing the latter issue.

4.1. Effect of v_{GS} transition time on the power loss

We firstly investigate the operation of the MOSFET in triode region and in saturation region. Since the voltage across the MOSFET, v_{DS} , is partly controlled by v_{GS} and partly controlled by the load network, then, in order to simplify the following analysis we can make an assumption that v_{DS} is linear. Other more accurate analysis such as assuming v_{DS} is a parabolic waveform[17] can be accounted for and could be the topic of a future research. In Fig. 4, since there are no closed relations between input signal and the voltage of the drain-source of the MOSFET, we can assume the angular time, θ_1 , makes the MOSFET change operating mode from triode region to saturation region, and the MOSFET is OFF after the angular time θ_2 . For the MOSFET operating in strong inversion region, we can consider

$$i_D \approx \beta(v_{GS} - V_T)v_{DS} \quad (3)$$

and
$$i_D = \frac{\beta}{2}(v_{GS} - V_T)^2 \quad (4)$$

in the saturation region. Where β is the device transconductance parameter, V_T is the threshold voltage. Here, we also assume that the falling rate of the input signal is U_{GS} , so we have two relations

$$v_{GS}(\theta) = U_{GS}\theta + V_H \quad (5)$$

and
$$v_{DS}(\theta) = \theta \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right) \quad (6)$$

where V_H is the high value of the input signal and $V_{HT} = V_H - V_T$. Hence, the power loss in the triode region is given as

$$\begin{aligned} P_{triode} &= \frac{1}{2\pi} \int_0^{\theta_1} \beta(U_{GS}\theta + V_H) \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right) \theta^2 \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right) d\theta \\ &= \frac{\beta}{2\pi} \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right)^2 \left[\frac{U_{GS}}{4} \theta_1^4 + \left(\frac{V_{HT}}{3} \right) \theta_1^3 \right] \end{aligned} \quad (7)$$

and the power loss in the saturation region is

$$\begin{aligned}
 P_{sat} &= \frac{1}{2\pi} \int_{\theta_1}^{\theta_2} \theta \left[\frac{\beta}{2} (v_{GS} - V_T)^2 \right] \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right) d\theta \\
 &= \frac{\beta}{4\pi} \left(U_{GS} + \frac{V_{HT}}{\theta_1} \right) \left\{ \frac{1}{12} \left(\frac{V_{HT}^4}{U_{GS}^2} \right) - \frac{1}{4} U_{GS}^2 \theta_1^4 - \frac{2}{3} V_{HT} U_{GS} \theta_1^3 - \frac{1}{2} V_{HT}^2 \theta_1^2 \right\} \quad (8)
 \end{aligned}$$

Since the MOSFET is “OFF” after θ_2 , there is no current flowing through this device and hence, no power loss.

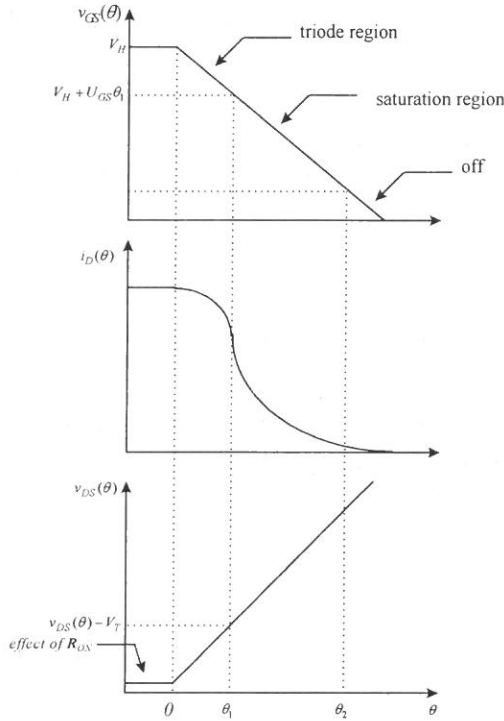


Fig. 4. Waveforms of the MOSFET during falling angular time of the input signal v_{GS} .

We can employ Fourier expansion for the input signal shown in Fig. 5. Assuming the period of the input signal is 2π , we can thus represent its Fourier series as

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \quad (9)$$

$$\text{where } a_0 = \frac{1}{2\pi} \left(\frac{m_1}{2} x_1^2 - \frac{m_2}{2} x_2^2 + \frac{m_2}{2} x_3^2 - y_1 x_1 - b_2 x_2 + b_2 x_3 + y_1 x_2 \right) \quad (10)$$

$$a_n = \frac{1}{\pi} \left[\frac{1}{n} (m_1 x_1 - y_1) \sin nx_1 + \frac{m_1}{n^2} \cos nx_1 + \frac{1}{n} (y_1 - m_2 x_2 - b_2) \sin nx_2 - \frac{m_2}{n^2} \cos nx_2 + \frac{1}{n} (m_2 x_3 + b_2) \sin nx_3 + \frac{m_2}{n^2} \cos nx_3 - \frac{m_1}{n^2} \right] \quad (11)$$

$$b_n = \frac{1}{\pi} \left[\frac{m_1}{n^2} \sin nx_1 + \frac{1}{n} (-m_1 x_1 + y_1) \cos nx_1 - \frac{m_2}{n^2} \sin nx_2 + \frac{1}{n} (-y_1 + m_2 x_2 + b_2) \cos nx_2 + \frac{m_2}{n^2} \sin nx_3 - \frac{1}{n} (m_2 x_3 + b_2) \cos nx_3 \right] \quad (12)$$

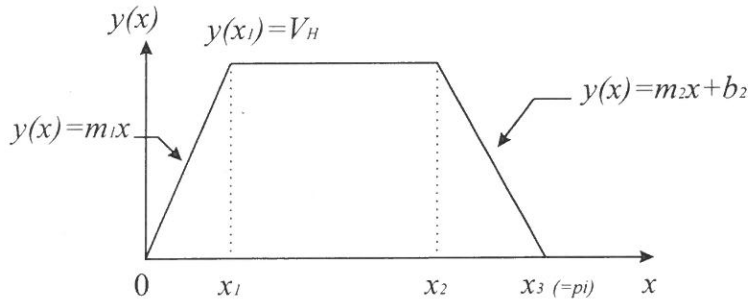


Fig. 5. Waveform of input signal used in Fourier expansion

Substituting the circuit parameters into these equations, gives

$$a_0 = \frac{1}{2\pi} \left(\pi V_H + \frac{V_H^2}{U_{GS}} \right) \quad (13)$$

$$a_n = 0 \quad n = 1, 3, 5, \dots$$

$$= \frac{-2}{\pi} \frac{U_{GS}}{n^2} \left(\cos \frac{nV_H}{U_{GS}} - 1 \right) \quad n = 2, 4, 6, \dots$$

(14)

$$b_n = \frac{2}{\pi} \frac{U_{GS}}{n^2} \sin \left(\frac{nV_H}{U_{GS}} \right) \quad n = 1, 3, 5, \dots$$

$$= 0 \quad n = 2, 4, 6, \dots$$

(15)

where $m_1 = -U_{GS}$, $y_1 = V_H$, $x_1 = \frac{V_H}{-U_{GS}}$, $x_2 = \pi + \frac{V_H}{U_{GS}}$, $x_3 = \pi$, $m_2 = U_{GS}$,
 $b_2 = -\pi U_{GS}$

Hence,

$$f(x) = \frac{1}{2\pi} \left(\pi V_H + \frac{V_H^2}{U_{GS}} \right) + \sum_{n=1,3,5,\dots} \left[\frac{2 U_{GS}}{\pi n^2} \sin \left(\frac{n V_H}{U_{GS}} \right) \right] \sin nx + \sum_{n=2,4,6,\dots} \left[\frac{-2 U_{GS}}{\pi n^2} \left(\cos \frac{n V_H}{U_{GS}} - 1 \right) \right] \cos nx \quad (16)$$

The power loss of the MOSFET is affected by many factors. The factors of the greatest interest are the input signal level, the falling rate of the input signal, and the angular time which makes the MOSFET change operation mode. Common parameter values used in the analysis are: $W=4000\mu\text{m}$, $L=0.6\mu\text{m}$, $V_T=0.68\text{V}$, $V_H=2\text{V}$, $\beta=2.1617 \text{ A}\cdot\text{V}^{-2}$.

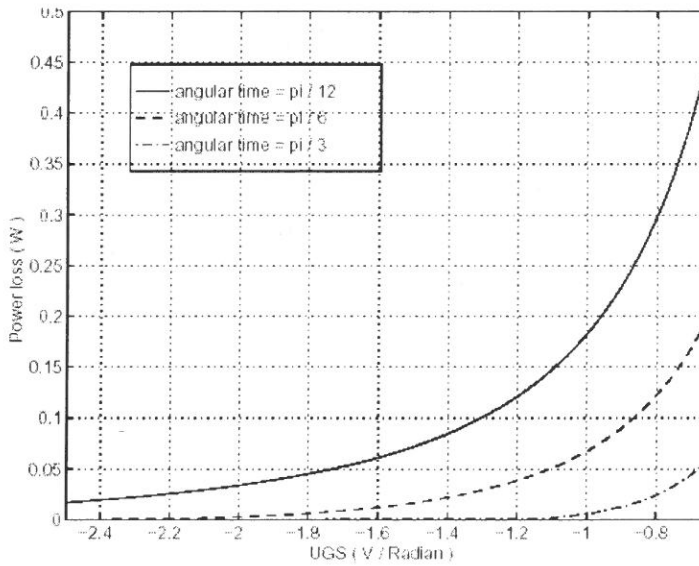


Fig. 6. Power loss versus falling rate of input signal.

4.1.1. Dependence on Input Signal Falling Rate and Angular Time θ_1

Fig.6 shows the falling rate of the input signal versus power loss (The value of U_{GS} is negative for the polarity.). For a higher falling rate (The absolute value of U_{GS} is higher), the power loss is much smaller than a lower falling rate. Clearly,

large power loss can be seen for the absolute value of a falling rate under 1 V/R .

Fig.6 also shows the effect of different angular time (θ_1) on the power loss. The optimum performance for Class E operation is the origin of v_{DS} should be higher than θ_2 as depicted in Fig. 4. Since the voltage and the current waveforms of the MOSFET do not overlap under this condition, the power loss in the MOSFET is zero. Therefore, for higher θ_1 angular time, the power loss becomes smaller.

4.1.2. Relations Between the Falling Rate and the Fourier Expansion Components

Fig.7(a) and (b) show the relationships between the falling rate and the percentages of each components of the Fourier series. For a higher falling rate, the percentages of the odd-order harmonics become higher whereas the percentages become lower for even-order harmonics. On the other hand, for a lower falling rate, the second harmonic frequency has much higher percentages and in this case, the power loss is even more serious. From equation (16), if we let $U_{GS} \rightarrow -\infty$, we can obtain a square wave which contains $\sin nx$ terms only where n is an odd number. According to the analysis, the power loss is the smallest.

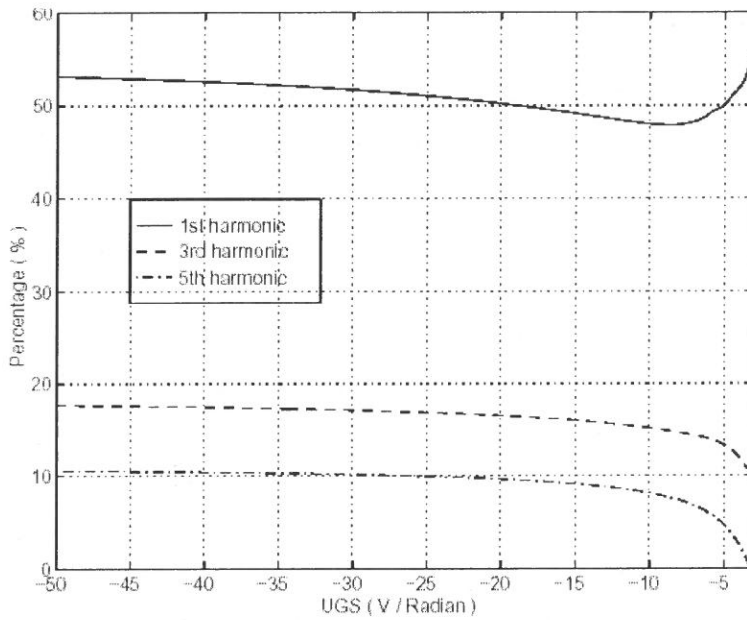


Fig. 7(a). Percentages of the odd-order terms of the spectrum versus falling rate of v_{GS} .

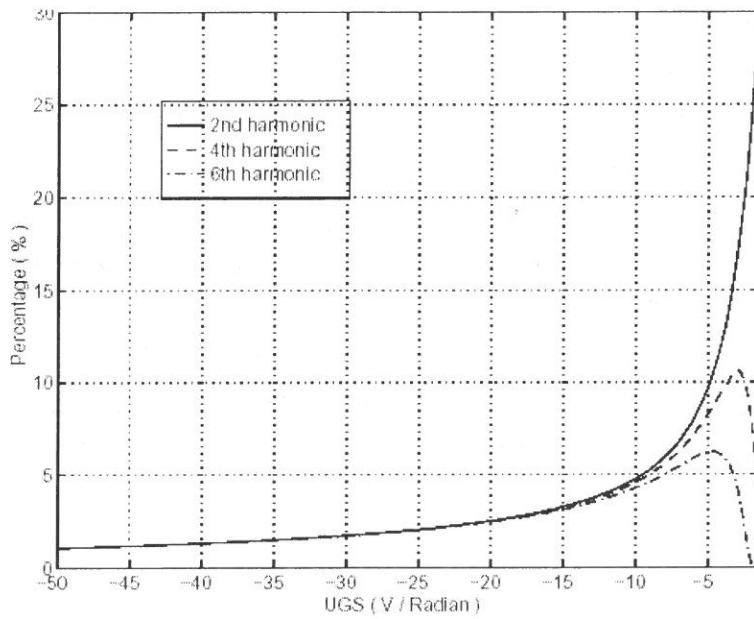


Fig. 7(b). Percentages of the even-order terms of the spectrum versus falling rate of v_{GS} .

4.2. Pre-amplifier design

In order to obtain higher output power and prevent overly loading the previous stage, *eg.* a mixer or whatever block might precede the PA, a multi-stage architecture is required in power amplifier design.

Theoretically, a square-wave should be employed to act as the driving signal for the Class E power amplifier due to its short transition time as analysed in the previous section, whereas in high frequency operation, a square-wave is difficult to generate due to its inherent much higher frequency components. This is especially true for the CMOS technology due to its significant parasitics. A two-stage design was proposed by T. Sowlati *et al.*[11] to achieve this target in GaAs technology. Notice that the low substrate loss and much smaller parasitics of GaAs MESFET technology make this target more feasible.

An attempt has been launched in CMOS technology to employ a high frequency filter to remove the second harmonic frequency of the driving signal thereby making it more like a square wave[18]. Since more components are required to achieve the filtering function, the design is more susceptible to destroy the benefit of less switching power loss. However, vigorous efforts to reduce the power loss in the passive components such as on-chip spiral inductors[19]-[20] or bondwire inductors[21] have prompted another look at this situation. Also by taking advantages of multi-stage design, the final stage can be over-driven by the large driving signal, which it is equivalent to employ a square wave as the input driving signal[22]. A typical design of a two-stage architecture is shown in Fig. 8 in which the biasing inductors are employed to perform the level shifting function since the threshold voltage is positive for the NMOS device. Notice that these inductors should have extremely low parasitic resistance in order to reduce the metal power loss.

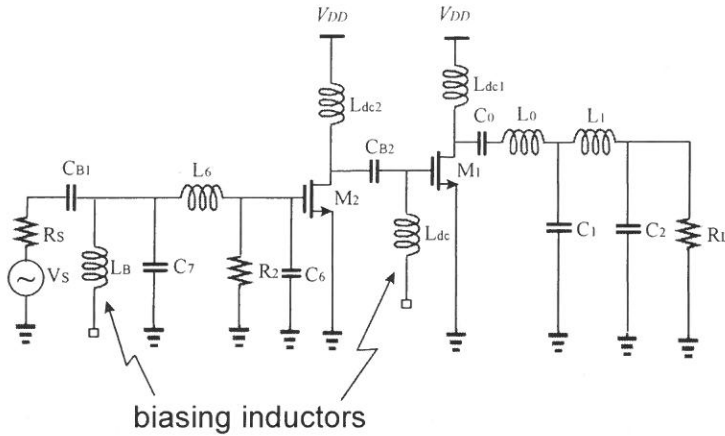


Fig. 8. An example of driving stage design.

The output matching circuit ($L_1 C_1 C_2$) transfers the 50Ω standard load to the optimum load, of 9Ω for 2.5V design[11]. ($L_0 C_0$) resonates at the 1.8GHz operating frequency. The parallel capacitance is absorbed into the junction capacitance of M1 at such high operating frequency. L_{dc1} acts as the RF choke for the final stage. C_{B2}, C_{B2} are the coupling capacitors. (C_6, L_6, C_7, R_2) forms the input matching circuit for the source resistance R_S .

Fig.9 and Fig.10 show the HSPICE simulated results of output power and power efficiency versus the variation of the supply voltage V_{dd} respectively. The quadratic relationship of output power and supply voltage has been verified.

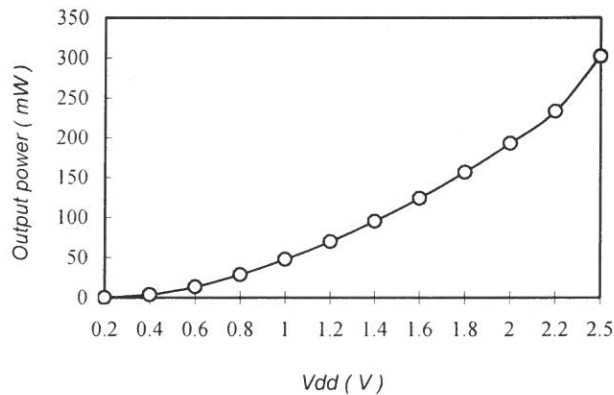


Fig. 9. Output power versus supply voltage

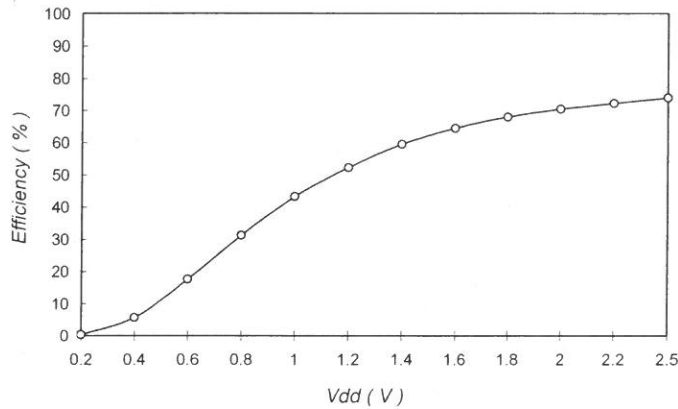


Fig. 10. Power efficiency versus supply voltage

5. Low-distortion Class E configurations

With the proposition of multi-standard transceivers and the maturity of the new frequency band in the range of 1.8-GHz, dual-band operation has rapidly become the mainstream of personal communications. Combining the design of a 900-MHz and 1.8-GHz CMOS transmitter for dual-band applications is, therefore, highly desirable. However, a critical design issue for the dual-band standard is that the second harmonic distortion of GSM 900-MHz signals in the power amplifier (PA) may deteriorate the generation of DCS 1.8-GHz signals since PA's are conventionally single-ended[23].

5.1. Differential Class E power amplifiers

While many RF circuits adapt single-ended topology due to off-chip component connection considerations, differential counterpart exhibits smaller harmonic distortion since even-order harmonics vanish if the system has odd symmetry such as fully differential topologies[24].

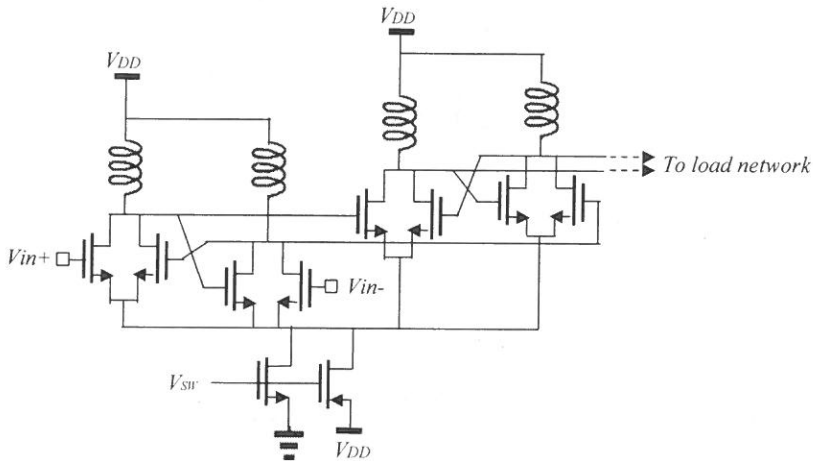


Fig.11. The differential Class E power amplifier.

A fully differential configuration for Class E power amplifier has been proposed as shown in Fig. 11[25], which it can alleviate the problem of substrate coupling since current is being discharged to ground twice per cycle. However, the input driving requirement for the switching devices may result in more complicated differential driving stage, which it may require off-chip components such as a balun to convert the input signal to differential form.

5.2. Complementary Class E power amplifiers

In this section, we investigate a different configuration power amplifier based on a highly symmetrical circuit topology as shown in Fig. 12[26]. With this approach, we may achieve lower harmonic distortion.

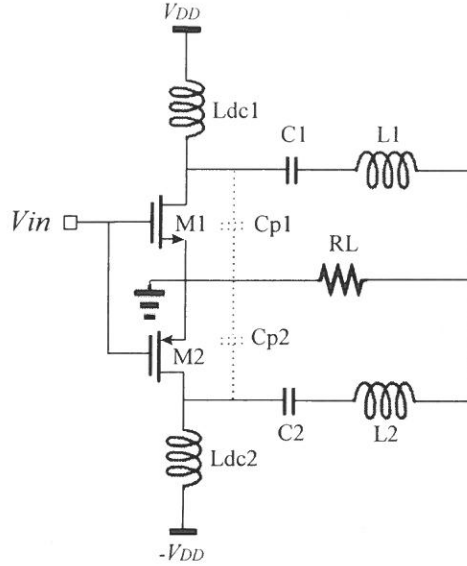


Fig. 12. The complementary power amplifier

Since the transistor acts as a switch rather than an amplifier, large-gate-width transistors are necessary in order to approach the ideal switch. However, large-gate-width transistors lead to large parasitic junction capacitances. For Class E power amplifiers, since the capacitance of the resonant circuits are different in switch-on and switch-off states, this leads to harmonic distortion of the output signal. Basically, a solution to this problem is to use a high-Q inductor or a bandpass filter[10]. For integrated implementation, however, these approaches do not only consume large chip area but present difficulties in obtaining high-Q inductors, especially using CMOS technology.

Intuitively, the power amplifier is more symmetrical than the conventional architecture. Thus, the inherent harmonic distortion problem may be alleviated. Fig. 13(a) and (b) show the equivalent circuits when M1 on, M2 off and M1 off, M2 on, respectively. According to KVL, KCL and boundary conditions, we can find the optimum operating conditions.

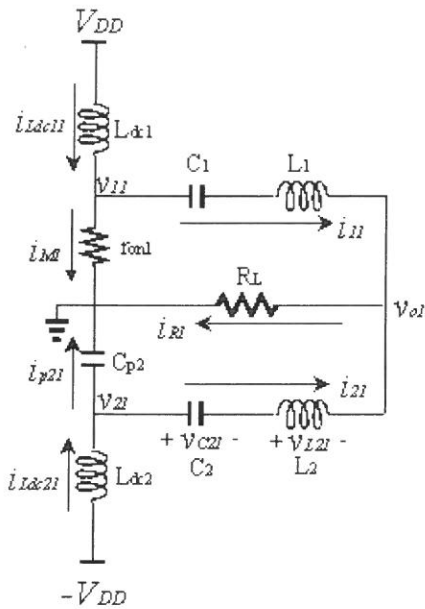


Fig.13(a). Equivalent circuit when M1 on, M2 off.

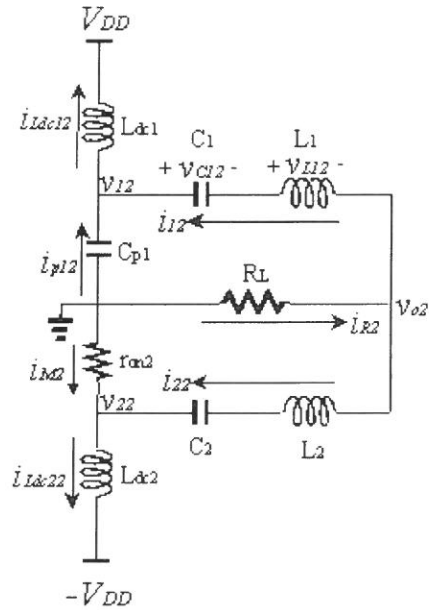


Fig. 13(b). Equivalent circuit when M1 off, M2 on.

We also compared the performance with the conventional single-ended structure by HSPICE simulations. Fig. 14 shows the simulated results of the power efficiency as a function of loaded quality factor based on the optimum load² for different supply voltages ranging from 1.0V to 2.6V which correspond to a loaded quality factor from 2.94 to 1.37, respectively. Both architectures maintain high power efficiency for a wide range of supply voltages.

² The definition of optimum load is the output load value which leads to maximum output power.

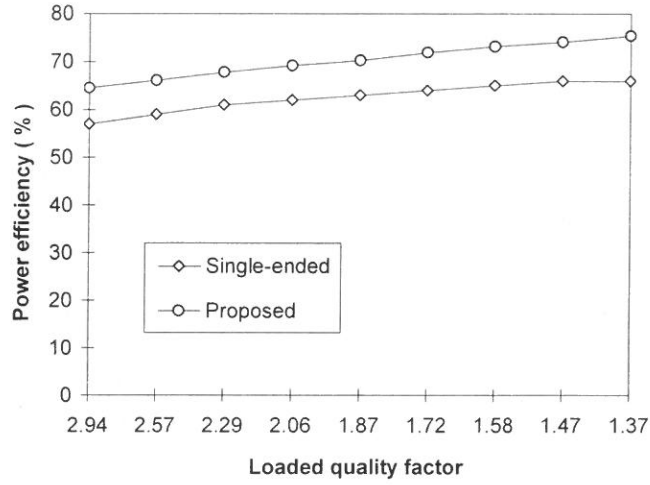


Fig. 14. Simulated results of power efficiency as a function of loaded quality factor

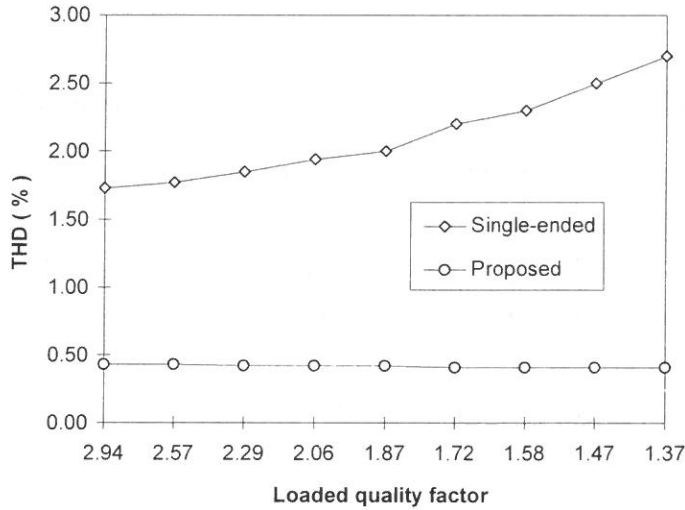


Fig. 15. Simulated results of THD as a function of loaded quality factor

Fig. 15 illustrates the comparisons of THD's as a function of loaded quality factor. The proposed architecture demonstrates much lower THD's since the balanced configuration inherently has a function of the second harmonic cancellation. Moreover, the proposed architecture alleviates the conflict between the distortion and the loaded quality factor since the THD's are independent of the loaded quality factor

as shown in the figure. Table 1 depicts the characteristic of low second harmonic of the proposed power amplifier.

Supply voltage (volts)	2nd harmonic (dBc)		3rd harmonic (dBc)		4th harmonic (dBc)		5th harmonic (dBc)	
	conventional	proposed	conventional	proposed	conventional	proposed	conventional	proposed
1.0 V	-36.2	-53.5	-43.1	-48.9	-57.0	-64.0	-53.4	-62.6
1.2 V	-36.0	-53.4	-43.0	-49.0	-58.1	-63.9	-54.2	-62.6
1.4 V	-35.4	-53.0	-43.1	-49.3	-59.6	-64.0	-55.3	-62.8
1.6 V	-34.9	-52.6	-43.1	-49.5	-60.5	-64.2	-56.4	-62.8
1.8 V	-34.5	-52.8	-43.1	-49.7	-60.2	-63.8	-57.5	-62.8
2.0 V	-33.7	-52.5	-43.2	-50.0	-60.0	-63.9	-58.6	-63.0
2.2 V	-33.1	-52.3	-43.1	-50.2	-59.2	-63.9	-59.0	-63.1
2.4 V	-32.5	-52.2	-42.9	-50.3	-58.4	-63.7	-59.4	-63.2
2.6 V	-31.8	-52.0	-42.9	-50.1	-58.2	-63.7	-59.7	-63.4

Tab.1 Comparisons of the harmonics of the output signal for the two different architectures based on the optimum loads of different supply voltages.

6. Conclusion

We have investigated the fundamental principles of Class E power amplifiers. The potential advantages of Class E operation leading to high efficiency and low supply voltage have also been examined. The summary is as follows:

1. For fundamental Class E power amplifiers, several conflicts exist between output power, power efficiency, and harmonic distortion. Although ideal analyses reveal general principles and results, great discrepancy exists between the theoretical analyses and practical results for giga-hertz operating frequencies due to large parasitic capacitors in MOSFET's.
2. The loaded quality factor Q_L of the load network of the single-ended Class E power amplifier is an important index for the trade-off between harmonic distortion and power efficiency.
3. Small feature-sized transistors or high-speed transistors should be employed as the

switch in Class E power amplifiers to reduce the power loss in the switch and thereby alleviate the conflict between harmonic distortion and power efficiency.

4. To reduce the power loss in the switching devices, shaping the current and voltage waveforms should be employed. In this paper, we reveal the considerations for the pre-amplifier stage design.
5. Since the resonant frequencies of the load network are different at the switch-on and switch-off states for single-ended Class E power amplifiers, symmetrical circuit topologies such as differential or complementary Class E architectures can be employed to reduce distortion.

Acknowledgment

The author is very grateful to the Fu Jen Catholic University for the financial support of this research project.

References

- (1) F.H.Raab, "Idealized operation of the Class E tuned power amplifier," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 725-735, Dec 1977.
- (2) N.O. Sokal and A.D. Sokal, "Class E, a new class of high efficiency tuned single-ended switching power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. SC-10, pp. 168-176, June 1975.
- (3) M.K.Kazimierczuk and K.Puczek, "Exact analysis of Class E tuned power amplifier at any Q and switch duty cycle," *IEEE Trans. Circuits and Systems*, vol. 34, pp. 149-158, 1987.
- (4) M.K.Kazimierczuk, "Effects of the collector current fall time on the Class E tuned power amplifier," *IEEE J. Solid-State Circuits*, vol. SC-18, pp. 181-193, 1983.
- (5) M.K.Kazimierczuk, "Collector amplitude modulation of the Class E tuned power amplifier," *IEEE Trans. Circuits and Systems*, vol. 31, pp. 543-549, 1984.

- (6) M.K.Kazimierczuk, "Class E tuned power amplifier with nonsinusoidal output voltage," *IEEE J. Solid-State Circuits*, vol.SC-21, pp.575-581, 1986.
- (7) M.K. Kazimierczuk and D. Czarkowski, *Resonant Power Converters*, Wiley, 1995, chapter 7.
- (8) F.H. Raab, "Effects of circuit variations on the Class E tuned power amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp.239-247, 1978.
- (9) F.H. Raab and N.O. Sokal, "Transistor power losses in the Class E tuned power amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp.912-916, Dec. 1978.
- (10) N.O. Sokal and F.H. Raab, "Harmonic output of Class E RF power amplifier and load coupling network design," *IEEE Journal of Solid-State Circuits*, vol. SC-12, no.1, pp. 86-88, Feb. 1977.
- (11) T. Sowlati *et al.*, "Low voltage, high efficiency GaAs Class E power amplifier for wireless transmitters," *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 1074-1080, Oct. 1995.
- (12) C.-H. Li and Y.-O. Yam, "Maximum frequency and optimum performance of class E power amplifiers," *IEE Proc. - Circuits Devices Syst.*, Vol. 141, No. 3, June 1994.
- (13) M.J. Chudobiak, "The use of parasitic nonlinear capacitors in Class E amplifiers," *IEEE Trans. Circuits Syst.-I*, vol.41, pp. 941-944, Dec. 1994.
- (14) J.A. Blanchard and J.S. Yuan, "Effects of collector current exponential decay on power efficiency for Class E tuned power amplifier," *IEEE Trans. Circuits Syst.-I*, vol.41, pp.69-72, Jan. 1994.
- (15) S. H.-L. Tu and C. Toumazou, "Effect of the loaded quality factor on power efficiency for CMOS Class E RF tuned power amplifiers," *IEEE Transactions on Circuits and Systems - Part I Fundamental Theory and Applications*, pp.628-634, May 1999.
- (16) B. Molnar, "Basic limitations on waveforms achievable in single-ended switching-mode tuned (Class E) power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 19, pp. 144-146, Feb. 1984.

- (17) H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980.
- (18) S. H.-L. Tu and C. Toumazou, "Highly efficient CMOS Class E power amplifier for wireless communications," *IEEE ISCAS* vol. 3, pp.530-533, June 1998.
- (19) J. Chang, A. A. Abidi, M. Gaitan, "Large suspended inductors on silicon and their use in a 2- μ m CMOS RF amplifier," *IEEE Electron Device Letters*, vol. 14, no. 5 May 1993.
- (20) C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," *IEEE Journal of Solid-State Circuits*, vol. 33, no.5, pp. 743-752, May 1998.
- (21) J. Craninckx and M. S. J. Steyaert, "A 1.8-GHz CMOS low-phase-noise voltage-controlled oscillator with prescaler," *IEEE Journal of Solid-State Circuits*, vol. 30, no.12, pp. 1474-1482, Dec. 1995.
- (22) D. Su and W. McFarland, "A 2.5-V, 1-W monolithic CMOS RF power amplifier," *IEEE Custom Integrated Circuits Conference*, pp. 189-192, 1997.
- (23) B. Razavi, "A 900-MHz/1.8-GHz CMOS transmitter for dual-band applications," *IEEE Journal of Solid-State Circuits*. vol. SC-34, no. 5, pp. 573-579, May 1999.
- (24) B. Razavi, *RF Microelectronics*, Prentice Hall, 1997.
- (25) K. Tsai and P. Gray, "A 1.9-GHz, 1-W CMOS Class-E power amplifier for wireless communications," *IEEE Journal of Solid-State Circuits*, vol.34, pp.962-970, July 1999.
- (26) S. H.-L. Tu and C. Toumazou, "Low-distortion CMOS complementary Class E RF tuned power amplifiers," *IEEE Transactions on Circuits and Systems - Part I Fundamental Theory and Applications*, vol. 47, no. 5 pp.774-779, May 2000.

received September 14, 2002
revised October 9, 2002
accepted November 6, 2002

無線通訊所用之高性能 E 級功率放大器

杜弘隆*

輔仁大學電子工程系

摘 要

本論文介紹幾種金氧半（CMOS）E 級功率放大器的配置。每種配置都能減少此 E 級功率放大器設計上的某些問題。兩級式 E 級功率放大器探討了前置放大器的設計技巧，此技巧提供了以 E 級操作而言最有效率的前置推動信號。而互補式架構利用對稱的電路結構造成輸出信號有更低的總體諧波失真。在本論文有關功率放大器的開關及負載網路品質因素關係之探討亦可提供在設計此級功率放大器之負載網路及開關有效的設計準則。

關鍵詞：E 級，功率放大器，負載品質因素，互補式。

使用於影像壓縮系統之二維離散餘弦正/反 轉換器之架構設計與實現

呂學坤* 陳忠陽

輔仁大學電子工程系

摘 要

在本篇論文中，我們提出了具行列分解 (Row-Column Decomposition) 特性，與無須外置轉置記憶體的全平行化心脈式陣列 (Fully Parallel Systolic Array) 的二維離散餘弦正/反轉換器 (2D-FDCT/IDCT) 之架構，其具有如下之優點：(1) 擁有最規則的電路設計、(2) 最簡易的控制訊號、(3) 符合 MPEG-2 各項規格需求的高效能輸出量 (它能在每 N 個 Clock 即完成一筆 $N \times N$ 轉換的高效能輸出量，而且電路的輸出延遲量 (Latency) 只有 $2N + 1$ 個 Clock)、(4) 具高度彈性化架構等特點；我們使用 Altera 公司 Quartus II 軟體作 P&R 後的模擬與除錯平台，而且以 C 語言的模擬結果與我們電路模擬的輸出結果作相互比較與驗證後得以驗證我們所提出的架構之正確性；在考慮晶片測試方面，我們先利用 Synopsys 的 TextraMax 軟體產生測試向量，經故障模擬後得到 99.77% 的故障涵蓋率 (Fault Coverage)。未來我們將使用可測試性設計，使其符合 IP 的設計需求。

關鍵詞：行列分解 (Row-Column Decomposition) 演算法，平行化心脈式陣列 (Fully Parallel Systolic Array)，二維離散餘弦正/反轉換 (Two-Dimensional Forward /Inverse Discrete Cosine Transform)

*Corresponding author. Tel.: +886-2-29031111 ext. 3800; fax: +886-2-29042638
E-mail address: sklu@ee.fju.edu.tw

前 言

在視訊與影像壓縮處理的技術上 [1][15]，離散餘弦正轉換 (FDCT) 與離散餘弦反轉換 (IDCT) 已被廣泛的應用在許多數位系統的核心單元，如靜態畫面的壓縮 (JPEG)、動態影像的編解碼 (MPEG1、MPEG2)、影像電話與視訊會議 (H.261)、與數位廣播電視/HDTV...等；而其要達到空間上多餘訊息 (Spatial Redundancy) 的壓縮原理，皆是源自人類視覺系統本身對於高頻較不敏感的特性作考量而來。

我們可利用統計相依的性質，將任何介於像素間多餘及可被刪除的訊息找出來，除去這些多餘訊息的其中一個方法，便是利用影像轉換的方法將其去除；例如離散餘弦轉換 (DCT)，它為正交函數，所以轉換基底 (Basis) 彼此間正交；一般來說，每個待轉換的數位影像資料的區塊 (Block) 大小皆為 8×8 個像素 (Pixels) 值，所以，經由轉換後的整組區塊值均是唯一，而其中的每個係數皆表示不同基底影像 (Basis Image) 的頻率內容，所以當其係數的索引值越大時，表示其所代表的基頻影像有較高的頻率；因此我們可以知道，經由離散餘弦轉換完成後，係數間的特性為 (1) 係數彼此之間並無關聯 (2) 大部分的訊息內容會較集中在低頻區；轉換後較高索引值的係數部份，可經由量化的過程後被刪去；若要再將其還原則要利用離散餘弦反轉換，使其還原為空間域的資料。

現在已有許多的 2D-DCT/IDCT 演算法被提出，大略可分為下列數種，如直接運算演算法 (Direct Computation Algorithm)，它是未經過簡化的演算法，所以，轉換一個 $N \times N$ 的 DCT 區塊時，它分別需要 N^4 個乘法與加法運算；行列分解演算法 (Row/Column Decomposition) 則是利用二維離散餘弦轉換的可分離性，分解成兩個一維離散餘弦轉換，再利用一維離散餘弦轉換的對稱性質以奇偶分解 (Even/Odd Decomposition) 方法，將二維轉換之計算複雜度降低，因此它分別需要 $2N^3$ 個乘法與加法運算，但是會因為輸入轉換與轉換完成所輸出的行列方向不一致，所以需要大小為 N^2 的轉置記憶體 (Transposition Memory)；快速演算法 (Fast Algorithm) 也是要先將二維的離散餘弦轉換經由行列分解成一維轉換後，再以類似於快速傅利葉轉換 (FFT) 的方式，將 N 點的一維離散餘弦轉換以 $\log_2 N$ 方式持續分解，直到最後的 2 點為止，其運算的複雜度可大幅

降低，只需要 $N^2 \log_2 N$ 個乘法與加法運算，然而因為它是先經由行列分解的方法降低維度，所以仍會有轉換完的輸出值行列方向不一致的情形發生，所以也需要 N^2 個轉置記憶體；而分散式運算 (Distributed Arithmetic) 演算法，當 N 的大小固定且所要輸入的空間位置與所轉換後所相對應的位置為已知時，其離散餘弦轉換的餘弦函數項即為定值，此定值即為我們轉換時所需相乘的係數，所以此演算法是以 ROM 紀錄這些定值為基礎的技術，以利用查表方式代替乘法，可減去乘法計算所需的時間、功率損耗、與節省硬體實現的成本，它只需要 $32N$ 個加法運算、無需乘法運算，但是卻需大小為 $2 \times N \times 2^N$ 的額外 ROM 面積，且仍需 N^2 大小的轉置記憶體 (RAM)；結合分散式運算與流程圖 (Mixed Distributed Arithmetic/Flow Graph) 演算法將分散式運算所要用的 ROM 大小作簡化，但是要外加 2 個加法器以達到此功能，所以，它需要 $34N$ 個加法運算、無需乘法運算外，此演算法只需 $N \times 2^{N/2}$ 的 ROM 面積，可是仍需 N^2 大小的轉置記憶體；另外仍有多項式演算法 (Polynomial Algorithm)，文獻 [5] 是截至目前所知運算量最低的方法、其他尚有向量-基底演算法 (Vector-Radix Algorithm)、捲積結構 (Convolution structures)、質因子演算法 (Prime Factor Algorithm)...等。

然而，我們上述所提到的眾多演算法，並非全都適合運用超大型積體電路的技術來實現與應用，例如多項式演算法，雖然它所需的運算量是最低的，但是卻因連線繁複與複雜的乘法運算等問題，在上市時程與良率等種種因素的考量下，使得實現此演算法時並不易以合理的成本來實現；因為，除了行列分解演算法之外，直接演算法、和所有經過化簡所得到的快速演算法，都有程度不一的不規則連線問題出現，而若連線與電路架構的複雜度過高，則不容易以硬體方式實現，可能較適用於軟體實現方式；因為若電路的連線複雜度已超過蝶型架構 (Butterfly Structure) 甚多，則在電路佈局 (Layout) 時，無法產生很密實的電路排列與連線方式，相對於規則性的電路，其電路佈局後面積會較大，所製造後的成本也會較高；然而，在已邁入深次微米製程的考量後，最不規則的電路所受到 cross talk 和 coupling 的效應影響也會相對的越複雜，且用於製造測試的 DfT (Design for Testability) 方面的設計也越不易，若電路的錯誤涵蓋率不高，則會直接影響到良率的改進，無法有效降低製造成本；所以下一段的內

容，我們針對已經提出且被實作出來的架構部分作簡短的比較。

用來實作二維離散餘弦轉換器的方法，如先前所討論的，約略可分為兩種，一種是非行列分解方法 (NRCM)，另一種則為行列分解方法 (RCM)，大致上我們的結論如下；在非行列分解的實作方法中，以文獻 [3] 等人所提出的二維直接轉換架構較具代表性，因為它保存了直接二維轉換的低運算量特性，且相對於其它已被提出的二維直接轉換架構而言，它具有較好的規則度與高輸出量的特點；然而，與所有的行列分解方式所實作的電路比較起來，它的電路規則性與連線複雜度卻較高，而且輸出結果的順序並不規則；而由文獻 [5] 所提出的多項式轉換方式，經由他們的分析後發現，利用此方式在作直接二維離散餘弦轉換所需的運算量，在 $N = 8$ 時只需 96 個乘法過程，而且每個時脈可完成一個完整的二維離散餘弦轉換的高輸出量 (throughput) 能力，是目前所知的運算需求量最少與輸出量較高的電路，但是此方法並未被大量的運用於電路設計上，因為它的連線結構複雜度太高所致；另一種最常被用於硬體實現二維離散餘弦轉換/反轉換電路的方式，便是使用行列分解的方式 (RCM) 來實現，以目前所有的電路設計方式而言，雖然它擁有最具規則的電路結構與連線方式，然而以典型的電路架構來說，它的執行方式是先以列向量輸入一維離散餘弦轉換電路作運算，運算完之後再依序將所得到的結果，輸入到轉置記憶體中，待每個 8×8 區塊內所有的值皆被轉換完成存到記憶體後，再將其轉置記憶體內的值再次依序的輸入一維的離散餘弦轉換電路內，作第二次的一維離散餘弦轉換，此次所轉換完的值便是所要對應到的頻域係數值；然而此種類型的電路，例如文獻 [2]，因為有使用到轉置記憶體作轉置動作，所以無論是在面積上與速度的考量上都不盡理想；因此，有許多的研究成果發表均是利用分散式運算的架構來實作，例如文獻 [13]，他們所發表的架構，即是利用 ROM 的查表方式來代替乘法器，而且它的輸入與內部運算均是以單一位元的方式進行，所以它的面積很小，但缺點是它仍需要轉置記憶體，而且它只適用於低輸出位元率 (Low Bit Rate) 的系統；而在文獻 [11] 中，除了利用 PLA 查表方式取代乘法器提升速度與節省面積外，還巧妙地運用控制內部暫存器的技巧達成轉置的動作，取代了內部轉置記憶體的需求，另一個特點是資料的輸出和輸入格式都是 Rowwise 的方式，這樣更可省去外部轉置記憶體的設置 (如圖 1 所示)，而且更重要的是它

也符合 MPEG-2 的即時壓縮/解壓縮處理系統的硬體需求，缺點是它每一級 (Stage) 的架構與內部字組長度 (Wordlength) 都不相同，所以此型電路在作 DfT 時設計的考量上，便需要針對每一級作出不同的設計，在整體的 DfT 設計整合上的問題就會更為困難。

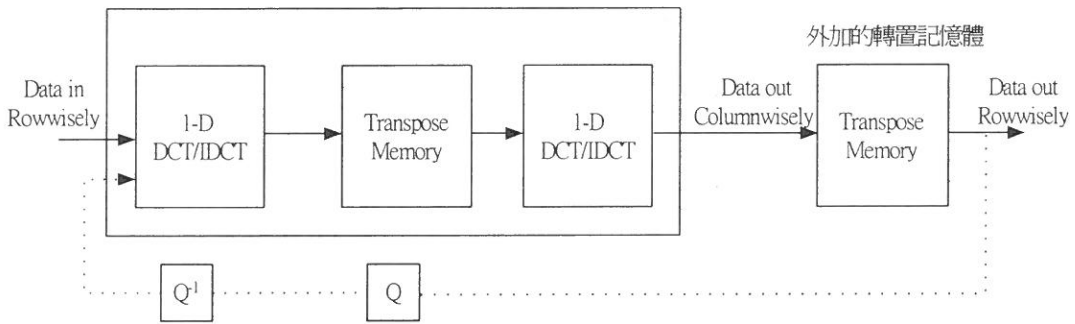


圖 1：包含外加轉置記憶體的行列分解演算法電路架構之方塊圖

文獻 [4] 則是運用行列分解演算法，實作出無須轉置記憶體的心脈式陣列架構，雖然它 N^2 的乘法器個數與每個時脈完成一個完整的二維離散餘弦轉換 (2-D transforms / cycles) 為 $1/N$ 的電路輸出量 (throughput) 均與我們所設計的架構在硬體的資源使用上是相同的，但是在與我們的架構比較下，它有電路規則性較差且連線複雜度較高等因素存在。

在本篇論文中我們提出了具行列分解特性與無須外置轉置記憶體的全平行化 (fully Parallel) 心脈式陣列架構；有別於文獻 [8] 所用的 serial-parallel 電路架構，雖然他們強調此種設計能達到效能與電路複雜度的折衷設計，然而 (1) 實際上的情形卻是它須求過多暫存器 (約為我們所設計架構的 2.5 ~ 3 倍) 的電路設計架構造成面積過大的問題，而在電路本身密集度很高的情形下，導致功率損耗提高的主因不再只是電路內部的轉態切換 (transition) 造成，反而是由電路本身面積過大所主導 [17]；(2) 我們所提出的架構於每個細胞單元只須非常簡潔的 load、clear、phase、FDCT/IDCT 四種控制訊號，相較之下，文獻 [8] 至少需要 $\Phi_0 \sim \Phi_6$ 六種不同相位的複雜控制訊號來協調單個細胞內部的電路動作，況且每級心脈式陣列的細胞內，為了解決細胞彼此間垂直方向的資料傳輸

問題上，仍須外加為數不等的移位暫存器來達成資料對齊的額外設計，更成了降低電路模組化設計的主因，然而，於我們所提出的架構設計上，便無須考慮此等問題；(3) 雖然在資料的輸入與輸出格式以序列方式傳遞時，很適用於低元率 (Low Bit Rate) 序列傳輸通訊上的使用，但是此架構所設計的電路，其所需傳遞的資料量卻太過於龐大而不適於此類的應用；且對於只能嵌入在影像與視訊編解碼系統的二維離散餘弦轉換/反轉換電路方面的應用而言，其 bit-serial 的傳輸方式則又更不適合。

與文獻中所提出的眾多電路架構比較時，我們所提出的架構具有如下之優點：(1) 擁有最規則的電路設計、(2) 最簡易的控制訊號、(3) 符合 HDTV 各項規格需求的高效能輸出量 (它能在每 N 個 Clock 即完成一筆 $N \times N$ 轉換的高效能輸出量，而且電路的輸出延遲量 (Latency) 只有 $2N + 1$ 個 Clock)、(4) 具高度彈性化架構等特點；在最後我們用 Synopsys 的 Design Compiler 軟體和由 CIC 所提供的 $0.35 \mu\text{m}$ 1P4M 的 Cell Library 作合成，在我們給了輸出負載電容、輸入訊號強度、輸入/輸出延遲、上升/下降延遲、操作在較差的工作環境等限制 (constraints) 後，發現此電路的核心可於 50MHz (worst-case) 的速度下操作，而它的資料輸出量可達 400M pixels/sec，可符合我們在表 1 所列 MPEG-2 的所有應用層級；我們使用 Altera 公司 QuartusII 作為模擬與除錯的平台，而且以 C 語言的模擬結果 (圖 13) 與我們電路模擬的輸出結果 (圖 14、圖 15) 作相互比較與驗證後可確知，我們所提出的架構是正確的；我們利用 Synopsys 的軟體產生測試向量，經故障模擬後得到 99.77% 的故障涵蓋率。未來我們將使用可測試性設計，使其符合 IP 的設計需求。

(I)、MPEG-2 的系統簡介與對其所需運算量的估算

MPEG-2 的標準是由 MPEG 組織約於 1980 年中期所制定的，其主要的特點 [12] [14] 約有下列幾點：(1) 與 MPEG-1 相容，(2) 更好的影像品質，(3) 彈性化的輸入格式，(4) 影像可依隨機存取顯示，(5) 具快速播放、倒帶播放與慢動作播放的功能，(6) 可增縮性的串流位元，(7) 於雙向傳輸時有較低的延遲時間，(8) 當有錯誤位元產生時，可用重新將巨集區塊 (macroblock) 任意切割的方式

重新選取所要傳遞的位元後，再將正確的訊息重新傳出；MPEG 的編解碼流程如圖 2 所示 [1]，於其輸入的部分有一前處理 (Preprocessing) 單元，它泛指將類比的色像轉換為數位的亮度 (Luminance) 與彩度 (Chrominance) 步驟的意思，亦即 R (Red)、G (Green)、B (Blue) 的類比訊號經由此步驟轉為 Y、Cb、Cr 的數位訊息，然而此時的輸出是以一個巨集區塊 (16×16 pixels) 的大小，作為輸入編碼系統的一個單位，然而在進行 DCT 轉換前，須將其巨集區塊分割為數個較小的區塊 (8×8 pixels)，作為待轉換影像的單位；如何將巨集區塊化為區塊取樣的比率問題，則依其所應用的領域與規格不同而有所差別，例如，在 CCIR601 其取樣格式為 [4:2:2]、SIF 是 [4:2:0]、JPEG 則有 [4:1:1] 與 [2:1:1] 兩種取樣方式、MPEG-2 則有 [4:2:0]、[4:2:2]、[4:4:4] 三種不同的取樣方式，可參考圖 3 [14] 所示，而 MPEG-2 的大略分類 (Profiles) 與分級 (Levels) 方式如表 1 [14] [12] 所示。

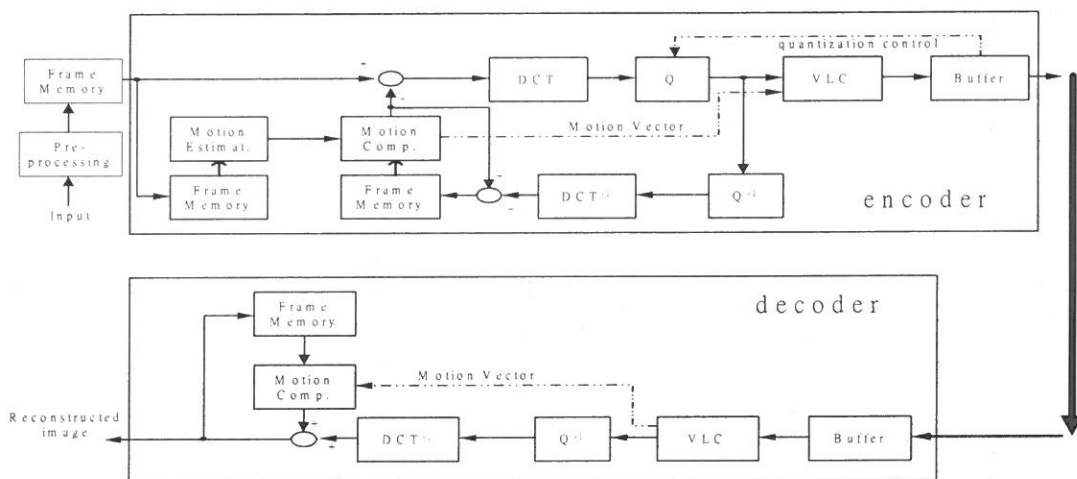


圖 2：MPEG 系統的編解碼方塊圖

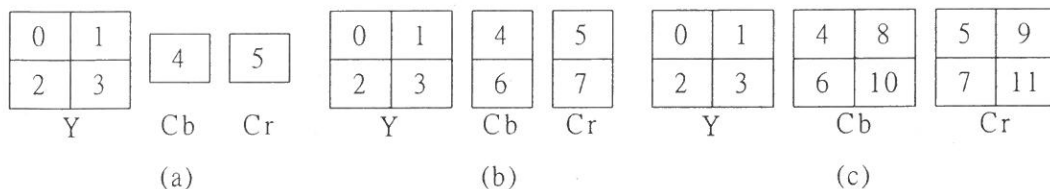


圖 3：(a) [4:2:0]、(b) [4:2:2] 與 (c) [4:4:4] 的取樣方式與其相對的取樣順序

Profiles Levels	Simple 4:2:0	Main 4:2:0	SNR 4:2:0	Spatial 4:2:0	High 4:2:2/4:2:0
High (Base) 1920 × 1152 × 60 HDTV		MP@HL HDTV			HP@HL
High-1440 1440 × 1152 × 30 HDTV		MP@1440		SSP@H1440	HP@H1440
Main (Base) 720 × 480 × 30 (720 × 576 × 25) CCIR601	SP@ML Digital CATV	MP@ML Direct TV	SNP@ML		HP@ML
Low (Base) 352 × 288 × 30 SIF		MP@LL	SNP@LL		

表 1：MPEG-2 的 Profiles 與 Levels

我們針對 Direct TV 的 MP@ML 規格 720 × 480, 30 f/sec 或 720 × 576, 25 f/sec 估算其所需的運算量如下：

$$720 \times 480 \times 30 = 720 \times 576 \times 25 = 10368000 \text{ (pixels)} = 162000 \text{ (blocks)}$$

其中 64 pixels/block，而取樣方式為 [4:2:0]，所以每秒的總取樣區塊為

$$162000 + 2 \times (162000 \div 4) = 243000 \text{ (blocks)}$$

因此若要達成即時影像與視訊編解碼系統的設計，則我們處理一個 block 最慢要在 4.115 μ s 內完成才行，換言之，要完成一個 pixel 的二維離散餘弦轉換/反轉換最慢要在 6.43 ns 內，又假設我們所設計的系統是每個 clock 完成 1 個 pixel 的轉換則電路速度只需執行逾 15.552 MHz 的速度便可滿足低於 MP@ML 規格的所有系統需求。若要滿足 HP@HL 取樣為 [4:2:2] 的系統需求，則電路的執行速度便要高於 265.4MHz，其分析過程如下：

$$1920 \times 1152 \times 60 = 132710400 \text{ (pixels)}$$

所以我們可以估算出要完成一個 pixel 的二維離散餘弦轉換/反轉換最慢要在 $1 \div [132710400 + 2 \times (132710400 \div 2)] \div 3.768 \text{ ns}$ 內完成，由此可知若我們設計的電路輸出量 (throughput) 能達到 266 M pixels / sec 則能適用於表 1 中所有的應用層級。

(II)、離散餘弦轉換/反轉換演算法

我們沿用文獻 [16] 對離散餘弦轉換與反轉換的定義，其中 $x(n)$ 代表空間領域的待轉換資料序列，而 $y(k)$ 則為其相對應完成轉換的係數值；

$$y(k) = \sqrt{\frac{2}{N}} u(k) \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N}, \text{ for } k=0, 1, \dots, N-1 \quad (1)$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} u(k) y(k) \cos \frac{(2n+1)k\pi}{2N}, \text{ for } n=0, 1, \dots, N-1 \quad (2)$$

$$\text{其中 } u(k) = \begin{cases} \sqrt{\frac{1}{2}}, & \text{if } k=0 \\ 1, & \text{otherwise} \end{cases}$$

我們可以發現 (1) 式與 (2) 式為一維的離散餘弦轉換與反轉換演算法，一般來說，若要用它們執行二維的離散餘弦轉換與反轉換的架構設計時，則必須使用文獻 [2] [10] 所提到的記憶體陣列轉置器 (RAM Array Transposer) 才行，如圖 4 所示；

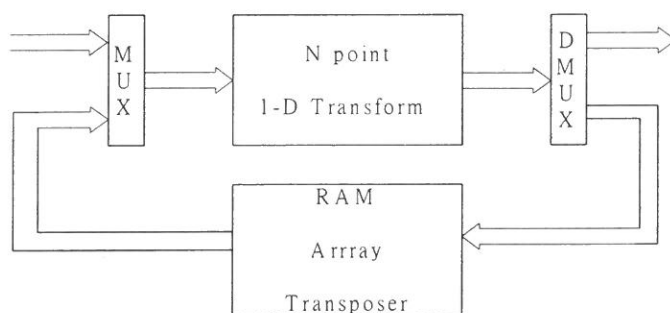


圖 4：使用單個一維離散餘弦轉換單元搭配記憶體陣列轉置器的電路方塊圖

在此我們先將 (1) 式與 (2) 式利用矩陣的表示法，將一維的離散餘弦轉換/反轉換分別重新改寫成 $[Y_N]_F = \sqrt{\frac{2}{N}} [X_N][C_N]^T$ 與 $[Y_N]_I = \sqrt{\frac{2}{N}} [Z_N][C_N]$ 的形式，而所執行的第二次一維離散餘弦正/反轉換也可用矩陣表示法重新改寫成 $[Z_N] = \sqrt{\frac{2}{N}} [C_N][Y_N]_F$ 與 $[X_N] = \sqrt{\frac{2}{N}} [C_N]^T[Y_N]_I$ ，其所要執行的二維離散餘弦正/反轉換的矩陣表示法則可表示成 $[Z_N] = \frac{2}{N} [C_N][X_N][C_N]^T$ 與 $[X_N] = \frac{2}{N} [C_N]^T[Y_N][C_N]$ ；其中，上述所使用的 $[C_N]$ 代表的數值矩陣則為：

$$[C_N] = \begin{cases} \frac{1}{\sqrt{2}}, & \text{for } k = 0 \\ \cos \frac{(2n+1)k\pi}{2N}, & \text{otherwise} \end{cases}$$

$$k, n \in \{0, 1, \dots, N-1\}$$

在仔細的觀察 $[C_N]$ 後，不難發現其實此為一常數矩陣，因為當 N 的大小決定後則其中的 k, n, N 均為常數，而整個 $[C_N]$ 即為一個常數矩陣，對 $N = 8$ 而言，我們可求得整個 $[C_N]$ 矩陣的常數為：

$$[C_N] = \begin{bmatrix} D & D & D & D & D & D & D & D \\ A & C & E & G & -G & -E & -C & -A \\ B & F & -F & -B & -B & -F & F & B \\ C & -G & -A & -E & E & A & G & -C \\ D & -D & -D & D & D & -D & -D & D \\ E & -A & G & C & -C & -G & A & -E \\ F & -B & B & -F & -F & B & -B & F \\ G & -E & C & -A & A & -C & E & -G \end{bmatrix}$$

$$\text{其中} \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{bmatrix} = \begin{bmatrix} \cos \frac{\pi}{16} \\ \cos \frac{2\pi}{16} \\ \cos \frac{3\pi}{16} \\ \cos \frac{4\pi}{16} \\ \cos \frac{5\pi}{16} \\ \cos \frac{6\pi}{16} \\ \cos \frac{7\pi}{16} \end{bmatrix} = \begin{bmatrix} 3EC\ 53 \\ 3B\ 20\ D \\ 3536\ D \\ 2\ D\ 414 \\ 238\ E\ 7 \\ 187\ DE \\ 0C\ 7C\ 6 \end{bmatrix}, \quad \begin{bmatrix} -A \\ -B \\ -C \\ -D \\ -E \\ -F \\ -G \end{bmatrix} = \begin{bmatrix} \cos \frac{15\pi}{16} \\ \cos \frac{14\pi}{16} \\ \cos \frac{13\pi}{16} \\ \cos \frac{12\pi}{16} \\ \cos \frac{11\pi}{16} \\ \cos \frac{10\pi}{16} \\ \cos \frac{9\pi}{16} \end{bmatrix} = \begin{bmatrix} C\ 13\ AD \\ C\ 4\ DF\ 3 \\ CAC\ 93 \\ D\ 2\ BEC \\ DC\ 719 \\ E\ 7822 \\ F\ 383\ A \end{bmatrix}$$

而且我們也求出了它們 2 的補數表示法的 16 進制值，可以看出我們已經決定了每個 [CN] 內元素的字組長度 (Wordlength) 為 20 個位元 (bits)，這是根據文獻 [9] 利用機率之平均值與變異數的基本原理推導與模擬後所得到的結論，於是當我們將此架構的輸入字組長度取 20 個位元時，可符合文獻 [7] 即 IEEE 於 1990 年末，針對以 8×8 為一處理區塊大小時，用不同的硬體架構所實作出的反離散餘弦轉換電路，可能因為不同電路架構的每一級產生的截斷誤差 (truncation errors) 累計，造成轉換完成輸出時的數值精確度不同，而在彼此間分別搭配作編碼系統與解碼系統時，在系統被更新 (refreshed) 的週期前，所累計產生的相對誤差值不超過所制定的範圍，則可確保此間影像與視訊品質不至於下降太多所訂定下來的標準規格。

在明瞭由 IEEE 所制定的規格後，我們也瞭解了在文獻 [11] [6] 中為何要對其提出的架構作多級的細部分析，而且每一級的有限字組長度 (finite wordlength) 都不相同，也因為如此，它每一級之間的連線必然較文獻 [12] [4] 更不規則；然而我們所使用的架構在同一個相位操作時，每個處理單元 (PE) 內卻仍能保持著完全相同的連線結構，這在對此架構作 DfT 的設計上會有很大的幫助，然而，我們得先明瞭電路在正常操作時的架構，才能對其作 DfT 的設計，因此接下來的部分，我們會對電路架構的特性和操作，作完整而詳細的分析。

(III)、內部架構說明

我們所使用的架構是將暫存器內嵌於每個處理單元內，如圖 5 [10] 所示，即為每個處理單元的行為描述：

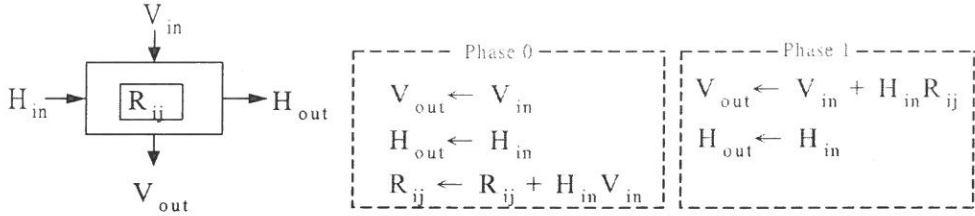


圖 5：在兩種不同相位時每個處理單元的行為描述

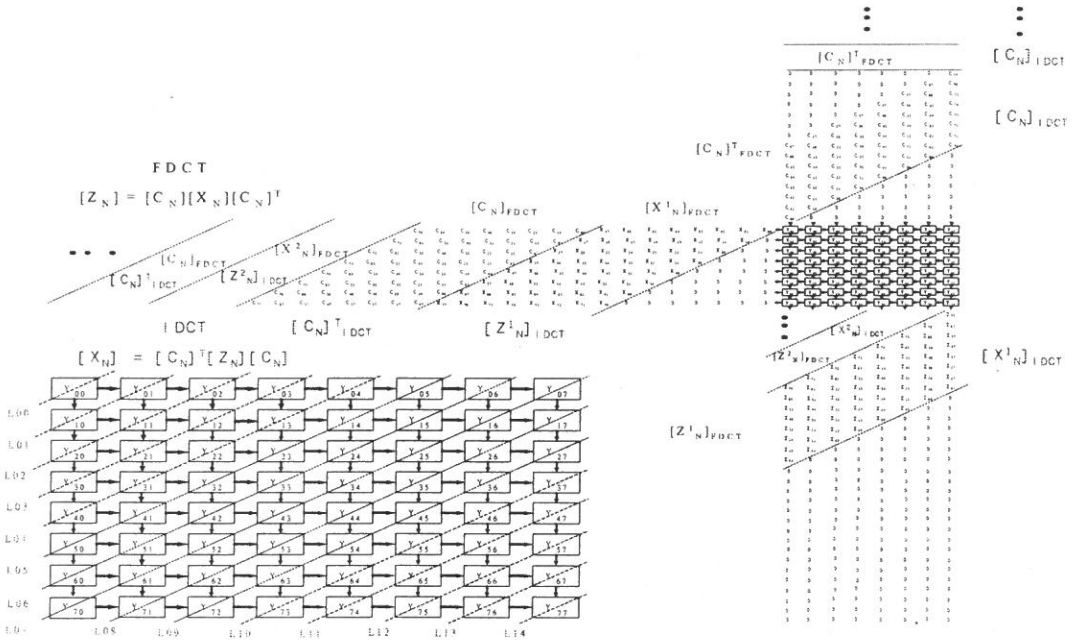


圖 6：二維離散餘弦轉換/反轉換的心脈式陣列架構

在詳細說明由資料輸入至處理單元內，到每個處理單元如何完成整個轉換的動作前，我們得先說明資料輸入到整個陣列的格式與整個陣列的操作動作；圖 6 表示我們所要進行轉換的輸入資料彼此間的關係，在前面的第 (II) 部分曾說明二維離散餘弦轉換可用線性代數的原理將它寫成矩陣的表示法，其中 $[Z_N]$

$$= \sqrt{\frac{2}{N}} [C_N][Y_N]_F \text{ 代表二維離散餘弦正轉換，而 } [X_N] = \sqrt{\frac{2}{N}} [C_N]^T[Y_N]_I \text{ 代表二維離}$$

散餘弦反轉換，其中 $[Y_N]_F = \sqrt{\frac{2}{N}} [X_N][C_N]^T$ 與 $[Y_N]_I = \sqrt{\frac{2}{N}} [Z_N][C_N]$ ，在圖 6 的說明中，我們暫時忽略 $\sqrt{\frac{2}{N}}$ 這個常數因子；首先，整個陣列處理單元彼此間的資料傳遞方式是以 45° 角 (L00 ~ L14) 的方式擴散給鄰接的單元，也因為輸入訊號資料流之間的關係是以 L00 ~ L14 的方式行進，所以其處理單元間控制訊號的連接關係也是如此，例如 L07 這條線穿過了 Y_{07} 、 Y_{16} 、 Y_{25} 、 Y_{34} 、 Y_{43} 、 Y_{52} 、 Y_{61} 、 Y_{70} 這 8 個編號的處理單元，所以它們的控制訊號皆相同 (Load、Clear、Phase、FDCT/IDCT 四條控制線)，並以並聯的方式連接，可依此類推其餘處理單元間控制訊號的連線方式，然而整個陣列所有處理單元間的 Clock 則都是以並聯的方式連接；接著因為我們所要進行轉換的 $N \times N = 8 \times 8$ ，所以矩陣皆為 8×8 的大小，也因此要算出 $[Y_N]_F$ 的值則要經過 8 筆值的相乘累加後存在各處理單元的暫存器內，例如要求得正轉換時 $Y_{00} = X_{00} \cdot C_{00} + X_{01} \cdot C_{01} + X_{02} \cdot C_{02} + X_{03} \cdot C_{03} + X_{04} \cdot C_{04} + X_{05} \cdot C_{05} + X_{06} \cdot C_{06} + X_{07} \cdot C_{07}$ 的值；我們以輸入訊號資料流的觀點依序解釋每個 Clock 的分解動作：在第 0 個 Clock 時，先將整個陣列的處理單元由相位 1 依序切換到相位 0 後 (請參考圖 10)，再依序啟動 Clear 訊號 (請參考圖 11)，將每個處理單元內先前暫存器的值清除掉；在第 1 個 Clock 時，作完 $X_{00} \cdot C_{00}$ 後，致能 (Enable) 編號第 Y_{00} 細胞的 Load 訊號 (請參考圖 12) 將值存到暫存器內；在第 2 個 Clock 時，輸入到處理單元 Y_{00} 的值作完 $X_{01} \cdot C_{01}$ 運算且累加了暫存器內之前的內含值 ($X_{00} \cdot C_{00}$) 後，則再次致能 Load 訊號將新的值存入處理單元 (Y_{00}) 的暫存器 (R_{00}) 中，在此同時 X_{00} 與 C_{10} 也已經傳到了處理單元 Y_{01} 中，且 X_{10} 與 C_{00} 也正好傳到處理單元 Y_{10} 中，而 Y_{01} 與 Y_{10} 此時則是重複 Y_{00} 在第 1 個 Clock 時的動作；依上述的步驟類推，到第 8 個 Clock 時，處理單元 Y_{00} 內的暫存器 R_{00} 的內含值即是一維離散餘弦正轉換的值 Y_{00} (忽略常數因子)；所以當第 9 個 Clock 時，處理單元 Y_{00} 則切換到相位 1 的模式下操作 (請參考圖 5，處理單元於相位 1 操作時的行為描述)，此時它會將水平方向的輸入值 C_{00} 與 R_{00} 的值作相乘後加上此時的垂直方向輸入值 0，得到垂直方向的輸出為 $C_{00} \cdot R_{00} + 0$ ，此時處於相位 0 的處理單元 Y_{10} 與 Y_{01} ，也正好完成了一維離散餘弦正轉換的值 Y_{10} 與 Y_{01} ，且分別儲存在 R_{10} 與 R_{01}

內；而在第 10 個 Clock 時，處理單元 Y_{00} 、 Y_{01} 與 Y_{10} 皆須保持在相位 1 的模式，因此編號 Y_{00} 的細胞所執行的功能如同前一個 Clock 一樣，它會將水平方向的輸入值 C_{10} 與 R_{00} 的值作相乘後加上此時的垂直方向輸入值 0，得到垂直方向的輸出為 $C_{10} \cdot R_{00} + 0$ ；而處理單元 Y_{10} 此時會將水平方向的輸入值 C_{01} 與 R_{10} 的值作相乘後累加垂直方向的輸入值是 $C_{00} \cdot R_{00}$ 而不再是 0；可依此類推，到第 11 個 Clock 時，處理單元 Y_{20} 的垂直方向輸出值為 $C_{00} \cdot R_{00} + C_{01} \cdot R_{10} + C_{02} \cdot R_{20}$ ；不難發現，當第 17 ($2N+1$) 個 Clock 時，第一筆二維離散餘弦轉換的係數值即可由 Y_{70} 的垂直輸出方向得到；依照上述的原則來推導，我們推導出整個陣列包含相對 Clock 訊息的輸入與輸出的關係圖 (即圖 6)。

於圖 6 的另一個須要說明的部分，是我們對離散餘弦正/反轉換輸入資料格式間所標示的相對應關係方式，可很清楚的看出 $[Z_N] = [C_N][X_N][C_N]^T$ 與 $[X_N] = [C_N]^T[X_N][C_N]$ 正好與所要輸入資料的位置關係相吻合，因此若我們想得到 $[Z_N]^T$ 輸出方式則只須將 $[Z_N]$ 取矩陣的轉秩變換即可得到： $[Z_N]^T = \{[C_N] \cdot ([X_N] \cdot [C_N]^T)\}^T = ([X_N] \cdot [C_N]^T)^T \cdot [C_N]^T = [C_N] \cdot [X_N]^T \cdot [C_N]^T$ ，同理， $[X_N]^T = [C_N]^T[X_N]^T[C_N]$ ；因此，我們只須將輸入的 $[X_N]$ 以轉秩的方式輸入，則可得到其相對應輸出的轉秩輸出格式，而無須更動任何輸入的 $[C_N]$ 部分，此即為電路的一個特點；另一個特點即是它的彈性化架構，也就是我們可用 8×1 (單一系列) 大小的陣列來執行 2 維離散餘弦轉換的功能，此時我們只須重新排列所要輸入的數值即可，簡單的說即是將垂直方向的輸入數值依序作串疊輸入，只要它們是以先輸入 N 個係數再輸入 N 個 0 的週期作循環即可達成。

而圖 5 是每個處理單元分別操作於兩個不同相位 (Phase) 時的行為描述如下；由於在此兩個相位時，每個處理單元分別作了兩個動作，所以我們將其分別說明如下；於相位 0 時：(1) 垂直方向與水平方向的輸入值分別傳遞給鄰接的處理單元，(2) 垂直方向與水平方向輸入值作相乘的動作後，加上暫存器 R_{ij} 的內含值後存回暫存器內；於相位 1 時：(1) 水平方向輸入值與暫存器 R_{ij} 內含值作相乘的動作後加上垂直方向的輸入值，然後再由垂直輸出的方向將結果輸出給鄰接的處理單元 (2) 將水平方向的輸入值輸出至鄰接的處理單元。

圖 7 則是我們所提出每個處理單元的電路架構詳細方塊圖，我們列出了它的控制訊號與相乘累加單元的詳細構造圖；而整個電路的操作則與前段所描述

的行為相同，其中最特別且仍未提及的部分是 Mux_D，它的功能是：(1) 將乘法器所產生的乘積作截斷 (truncation) 的動作，(2) 因為所要輸入作離散餘弦正/反轉換值的大小範圍並不相同，而且當暫存器 R 的內含值與水平輸入值作相乘的動作後，要再加上垂直方向的輸入值時，會有資料對齊的問題發生；所以在離散餘弦正/反轉換各需要作這兩個動作 2 次，因此在乘法器與加法器的連接中間必須加入此 4 對 1 的多工器。

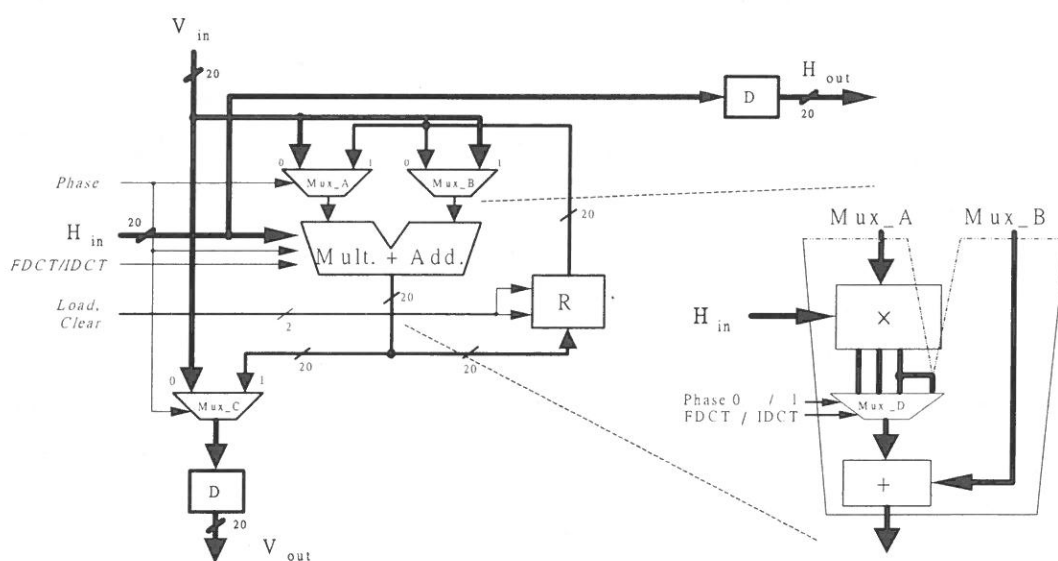


圖 7：每個處理單元內部架構的詳細方塊圖

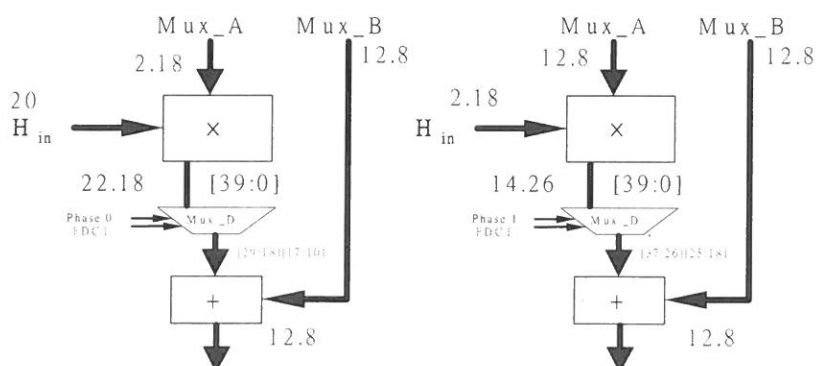


圖 8：在執行離散餘弦正轉換模式下的兩個不同相位時，資料格式的表示圖

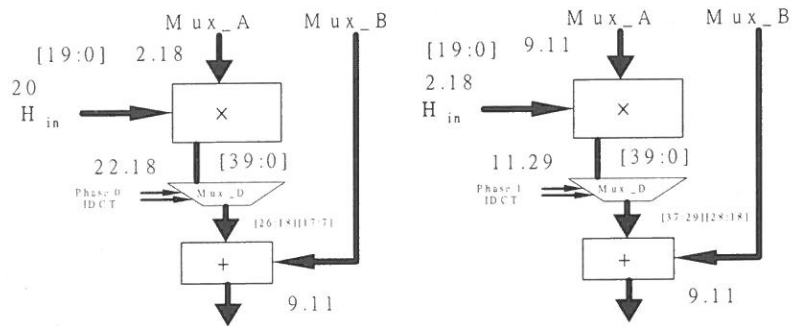


圖 9：在執行離散餘弦反轉換模式下的兩個不同相位時，資料格式的表示圖

在前一段內容中，我們推導了整個陣列的操作功能後，接下來的部分我們就可針對每個處理單元的內部資料格式作探討，其中我們依據 IEEE Std. [7] 的說明，在正轉換時的待輸入資料格式的範圍介於 $+255 \sim -256$ 之間，轉換完成輸出係數的資料格式範圍則介於 $+2047 \sim -2048$ 之間，所以它的輸入資料與輸出資料若用 2 的補數表示則分別為 9 bits 與 12 bits；在反轉換時的待輸入資料格式的範圍則介於 $+2047 \sim -2048$ 之間，轉換完成輸出係數的資料格式範圍則介於 $+255 \sim -256$ 之間，所以它的輸入資料與輸出資料若用 2 的補數表示則分別為 12 bits 與 9 bits；我們在第 (II) 部分後面有提到，我們使用了輸入為 20 bits 的架構，所以在正轉換與反轉換輸入資料格式在 MSB 的不足部分是作 Sign Extension 來補齊 20 bits。首先說明我們在圖 8 和圖 9 使用的資料表示格式：**ii.ff** 中的 **ii** 部分代表整數，**ff** 則代表小數，所以圖中的 2.18 表示有 2 位整數 18 位小數的二進位表示法；圖 8 左手邊的部分表示於相位 0 時，電路在執行離散餘弦正轉換模式下，資料的輸入與輸出格式；我們於 $[C_N]$ 的 2 的補數表示法是採用 **SX.XXXX** 的格式，其中 **S** 代表 sign bit，而 **X.XXXX** 則是 cosine 的 2 補數表示法的二進位數值，以長度為 6bit 表示時的例子來說，01.0000 表示 1，而 11.0000 則表示 -1，而在我們的電路設計中，則是使用了長度 20 bits 來表示 cosine 的 2 補數表示法的二進位數值，所以當輸入為 $[C_N]$ 我們則用 2.18 來表示。

而在資料的對齊問題方面，我們以圖 8 左手邊為例說明，輸入資料為 2.18 與 20 的格式輸入至乘法器會產生 22.18 (整數為 22 位元，小數則為 18 位元) 的 40 bits 的結果，若要給 20 bits 的加法器當輸入則要先將 40 bits 截斷成 20 bits 才行，而我們知道當相位 0 經過 8 個時脈 (對 $N = 8$ 而言)，所完成相乘累加的結果存於暫存器後，接著便要切換到相位 1 的操作，因為輸入數值的範圍介於

+255 ~ -256 之間，所以當與小於 1 的數值相乘累加 8 次後所產生的最大值必然介於 +2047 ~ -2048 間，此時的輸出格式的整數部分 12 bits 就可被決定出來，且總長度 20 bits 也已經固定了，所以得到的小數部分則為 8 bits，所以輸出以 12.8 來表示，而經由 Mux_B 所輸入到加法器的值，因為只是上個加法結果的輸出值暫存，所以必然與現在加法器的輸出格式相同；同理，圖 9 的左手邊的部分表示於相位 0 時，電路在執行離散餘弦反轉換模式的情形，內部數值的推論方法同上所述，只是此時我們所要得到輸出整數部分的範圍介於 +255 ~ -256 之間，所以它輸出的整數部分以 9 bits 來表示，其餘的 11 bits 即為小數部分；同樣的過程也發生在圖 8 與圖 9 所表示的部分，只是它們分別操作於正轉換與反轉換的相位 1 模式下，只是此時乘法器的水平方向輸入值為 $[C_N]$ ，垂直方向的輸入值為上一級相位 0 的輸出，所以它的垂直方向輸入分別承接上一級的 12.8 與 11.9 格式，而此格式也正好分別是它們的輸出格式。

(IV)、模擬結果

在完成整個電路的設計之後，我們使用 Altera 公司 QuartusII 作為模擬與除錯的平台，我們選用了具有內建 DSP 單元的 Stratix 系列作為合成與驗證的模擬元件；我們的設計總共使用了 53,204 個 Logic Elements 與 176 個 DSP block 9-bit elements，且我們所提出架構的所需控制訊號與模擬結果如下各圖（圖 10 ~ 圖 15）所示；電路的相位控制訊號與 Clock 之間的關係如圖 10 所示，Clock 0 的正緣觸發由 L00 所連接處理單元的相位訊號由 1 變到 0 的轉變，然後再由第 N 個 Clock 的正緣訊號，改變相位的控制訊號；Clock 1 則會觸發由 L01 的相位控制訊號，然後依此類推，直到由 Clock 14 的正緣觸發由 L14 所連接處理單元的相位訊號後，即形成相位控制訊號的完整週期；圖 11 的清除訊號則是由圖 10 相對應的相位控制訊號負緣所觸發；而處理單元內暫存器載入訊號的載入動作週期如圖 12 所示。最後我們以 C 語言的模擬結果（圖 13）與我們電路模擬的輸出結果（圖 14、圖 15）作相互比較驗證後可發現，我們所提出的架構是正確的，只是有限字組長的模擬結果尚未完成，然而對我們所設計的架構而言，這也只需要非常少部分的微調即可完成。

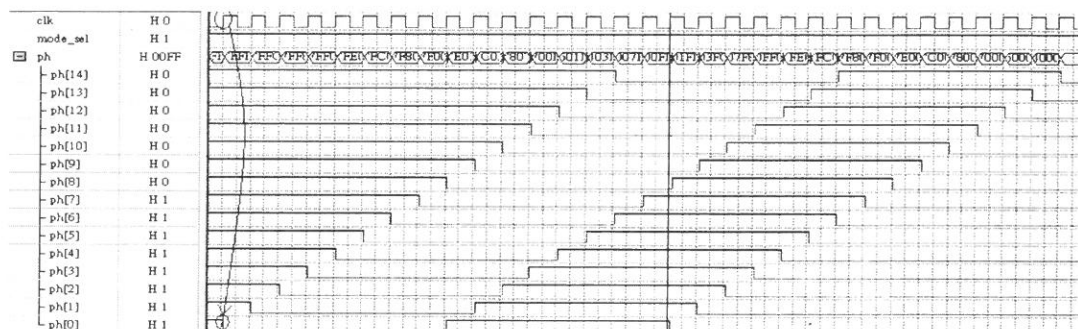


圖 10：於 Clock0 時，相位訊號與輸入時脈之間的關聯

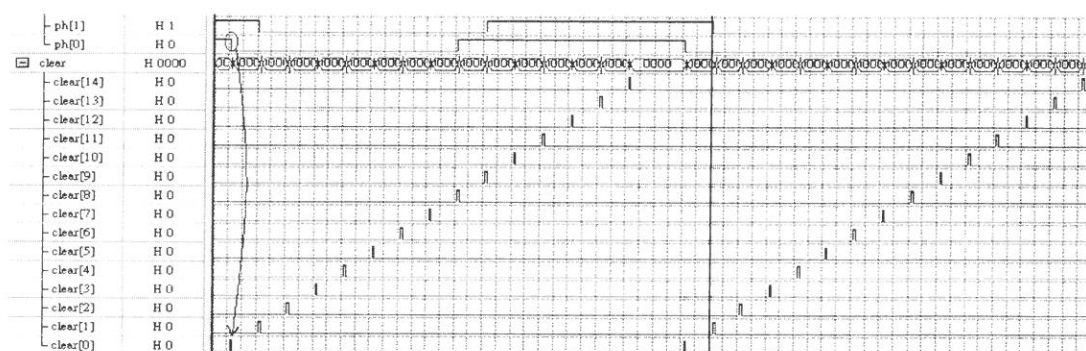


圖 11：於 Clock0 時，相位訊號與清除訊號之間的關係

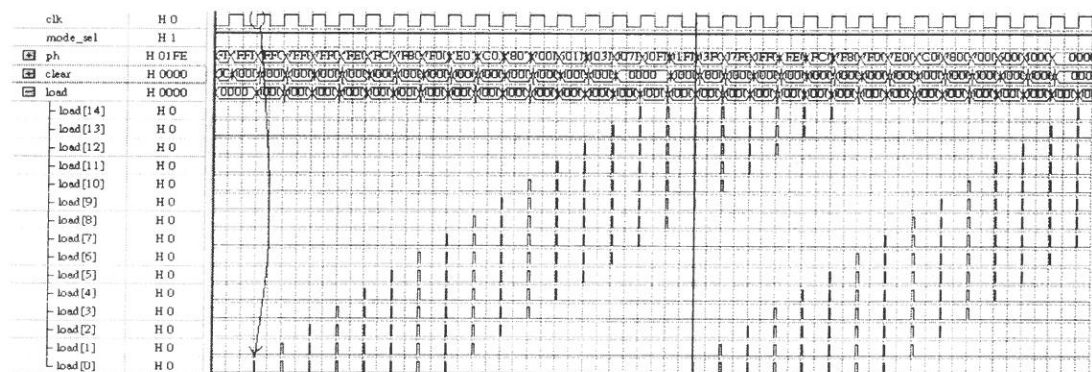


圖 12：於 Clock1 時，輸入時脈與載入訊號的關係圖

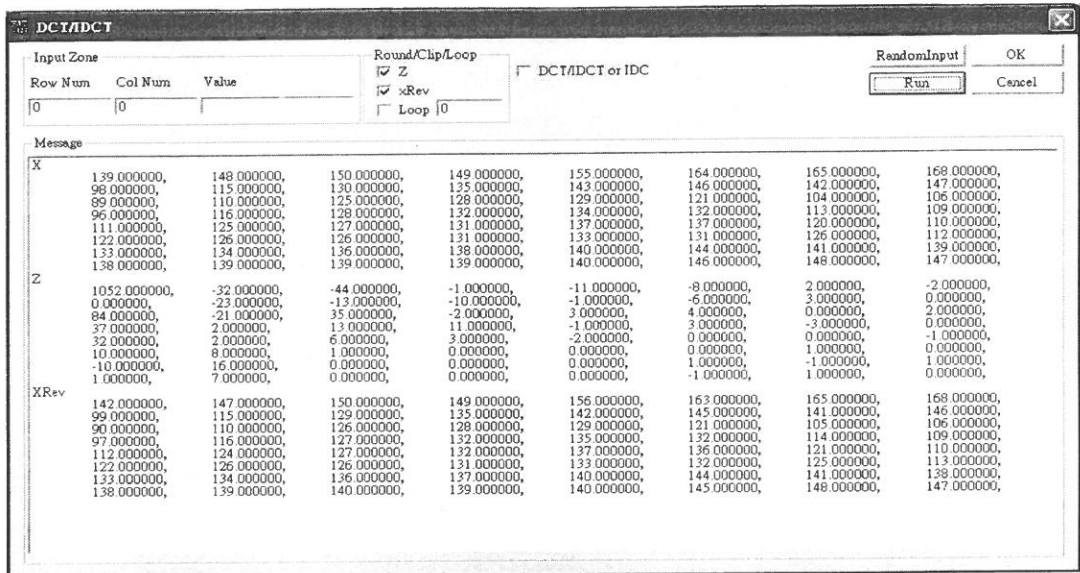


圖 13：C 語言的模擬結果

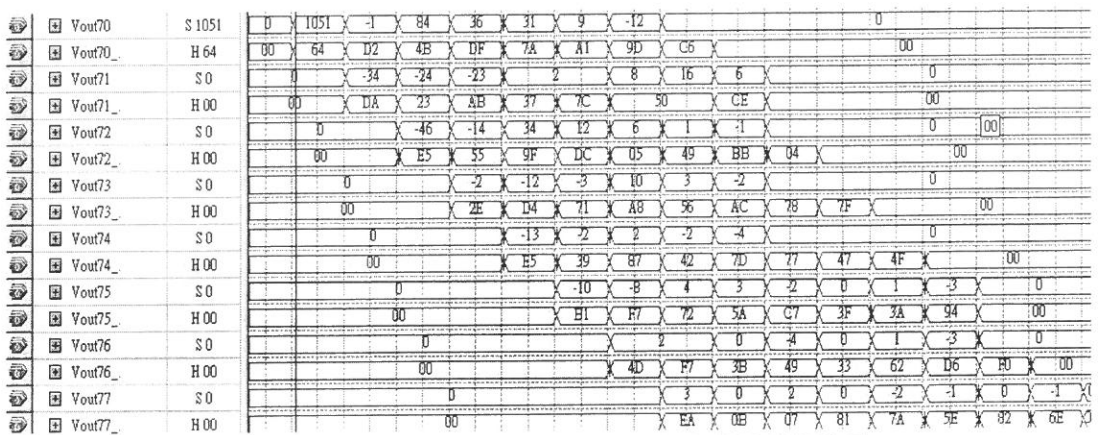


圖 14：我們所提出二維離散餘弦正轉換架構的模擬結果

1995

- [2] Han Bin Kim and Dong Sam Ha, "Design and Synthesis of Built-In Self-Testable Two-Dimensional Discrete Cosine Transform Circuits," *ASIC/SOC Conference, 2000. Proceedings. 13th Annual IEEE International*, pp. 3 -7, 2000
- [3] Yung-Ping Lee, Thou-Ho Chen, Liang-Gee Chen, Mei-Juan Chen, Chung-Wei Ku, "A Cost-Effective Architecture for 8×8 Two-Dimensional DCT/IDCT Using Direct Method," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 3, June 1997
- [4] Chin-Liang Wang and Yu-Tai Chang, "Highly Parallel VLSI Architectures for the 2-D DCT and IDCT Computations," *IEEE Region 10's Ninth Annual International Conference. Theme: Frontiers of Computer Technology. Proceedings of 1994*, pp. 295 -299 vol.1
- [5] P. Duhamel and C. Guillemot, "Polynomial transform computation of the 2-D DCT," *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, 1990, pp.1515 -1518 vol.3
- [6] Xia Wan, Yiliang Wang, and Wen H. Chen, "Dynamic Range Analysis for the Implementation of Fast Transform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 2, April 1995
- [7] CAS Standards Committee IEEE Circuits and Systems Society, "IEEE Standard Specifications for the Implementations of 8×8 Inverse Discrete Cosine Transform," *IEEE Std.* 1180-1990.
- [8] Hyesook Lim, Vincenzo Piuri, and Earl E. Swartzlander Jr. , "A Serial-Parallel Architecture for Two-Dimensional Discrete Cosine and Inverse Discrete Cosine Transforms," *IEEE Transactions on Computers*, vol. 49, no. 12, Dec. 2000
- [9] Hyesook Lim, Changhoon Yim and Earl E. Swartzlander, Jr. , "Finite Word-Length of an Unified Systolic Array for 2-D DCT/IDCT," *Application Specific Systems, Architectures and Processors, 1996. ASAP 96. Proceedings of International Conference on*, 1996, pp. 35 -44
- [10] Hyesook Lim and Earl E. Swartzlander, Jr. , "A Systolic Array for 2-D DFT and 2-D DCT," *Application Specific Array Processors, 1994. Proceedings.*

- International Conference on* , 1994, pp.123 -131
- [11] Yi-Fang Chiu, "A High-Throughput Two-Dimension DCT/IDCT Architecture for Real-Time Image and Video System and the VLSI Design," Master Thesis, Department of Electrical Engineering, Tamkang University, Taiwan 2001
 - [12] Yung-Ping Lee, "Architecture Design for Video Decoder and Discrete Cosine Transform," Ph. D. Dissertation, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan 1997
 - [13] Klaus Gaedke, Jens Franzen, Peter Pirsch, "A Fault-Tolerant DCT-Architecture based on Distributed Arithmetic," *Circuits and Systems*, 1993., ISCAS '93, 1993 *IEEE International Symposium on* , May 1993, pp.1583 -1586 vol.3
 - [14] 張嘉珮, "The MPEG Video Standards," Handout, NTUEE DIP Lab
 - [15] 陳同孝、張真誠、黃國峰著, "數位影像處理技術," 松崗電腦圖書資料股份有限公司
 - [16] A.V. Oppenheim and R.W. Shafer, "Digital Signal Processing," Englewood Cliffs, N.J.: Prentice Hall, 1975.
 - [17] Suzuki, H.; Yun-Nan Chang; Parhi, K.K., "Performance tradeoffs in digit-serial DSP systems," *Signals, Systems & Computers*, 1998. *Conference Record of the Thirty-Second Asilomar Conference on* , vol. 2 , 1998. pp. 1225 -1229 vol.2

received September 27, 2002
revised October 24, 2002
accepted November 8, 2002

Design and Implementation of a Two-Dimensional DCT/IDCT Processor for Video Compression Systems

Shyue-Kung Lu* and Chung-Yang Chen

Department of Electronic Engineering

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

In this paper, we proposed a fully parallel systolic array architecture by using row-column decomposition method for 2D-DCT/IDCT application. The architecture proposed possesses many advantages such as (1) the regularity of this architecture is higher than others. (2) control signals are greatly simplified. (3) it meets to all MPEG-2 specifications. It is a high throughput architecture which performs one complete $N \times N$ -point transform per N clock cycles, and the latency of the architecture is only $2N + 1$ clock cycles. and (4) the architecture is very flexible. We have implemented this processor with Synopsys Design Compile and COMPASS cell library. The core of the design can run at 50 MHz for the worst case. The throughput is 400Mpixels/sec which can be applied to all profiles and levels of MPEG-2 specifications. We also use QuartusII software of Altera Corp. to simulate the circuit function and compare the results with the simulation results of the software implementation with C language. The simulation results show that the architecture proposed can work correctly. For the test consideration, we use ATPG tool of Synopsys Corp. to generate the test patterns. Fault simulation is performed with Synopsys TetraMax. The fault coverage of the circuit is 99.77%. In order to meet the circuit requirements of an IP, design-of-testability techniques will be considered in our future work.

Key Words : Row-Column Decomposition, Fully Parallel Systolic Array, Two-Dimensional Forward / Inverse Discrete Cosine Transform

Supercapacitor Application in the Battery Energy System of the Electric Scooters

Jeremy C. C. Chang

Yuang- Shung Lee*

Department of Electronic Engineering,

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

This paper propose a hybrid energy system that contains supercapacitors (SCAPs) and acid- lead batteries for battery energy management system (BEMS). The BEMS strategy is proposed to improve the energy efficient of the hybrid energy system (HES) for the electric scooter (ES). The detailed power management model is composed of traction motor, wheels, road, transmission system, power converter/inverter, acid lead batteries, and supercapacitors. The system-level modeling, interactive simulations, and analysis package are developed in this study using Matlab/Simulink. The simulation result for the proposed BEMS strategy of HES show that it can reduce the over all power loss and extend the cruise range of the ES system.

Keyword : Supercapacitor, Battery energy management system, Hybrid energy system, Simulations, Electric scooter

*Corresponding author. Tel.: +886-2-29031111 ext. 3791; fax: +886-2-29042638
E-mail address: lee@ee.fju.edu.tw

1. Introduction

Recently, there have been energy shortage in Taiwan was lack, and most of the raw materials to produce energy depend on foreign exchange, so more efficient energy utilization is very important. In addition to the energy problem, there is also the even a more serious problem of air pollution in Taiwan. To solve these problem, the electric scooter is a promising technology for the long-range goal of energy efficiency and reduced atmospheric pollution, although their limited range and lack of supporting infrastructure may hinder their public acceptance [1]. The increasing the electric scooter's cruise range is an important research topic. First of all, a new energy storage device supercapacitor has been widely used in electric vehicles for supplementary power management [2] due to two important features. First, the energy density is higher than conventional capacitors and the power density is higher than rechargeable battery. In order to reach the goal of extending the ES's cruise range, the supercapacitor is a better choice for the second energy source. Several type of combined energy systems are currently in use, one of which is the supercapacitor system and cell stack connecting via dc-dc converter [3], and the another of which is only supercapacitor system connected to a dc-to-dc converter [4], [5]. Napoli and Crescimbeni showed three different combinations of battery, supercapacitor and fuel cell as derived, cascade and series [6], adopting a parallel combination in which each of the energy sources are connected on its own dc-to-dc converter. Using this type is easy to control the battery and supercapacitor tank input or output current. Jörg Lott and Helmut Späth also presented a similar hybrid energy storage system [7], But although in their system had only battery and supercapacitor. In this study use the combination in which only supercapacitors are connected to the dc-to-dc converter because this combination is more economical and effective.

The papers mentioned above discuss hybrid energy vehicles which have more than two energy sources. But in Taiwan, the conventional scooter is popular, and there are over 10 million conventional scooters in Taiwan. More than 330 thousand tons of carbon monoxide and 90 thousand tons of hydrocarbon are emitted by

conventional scooters every year. So some researches hope to use the ES's substituting for conventional ones. However, the ES has some drawbacks. One is that the ES is very heavy, about 100 kg; although this will be improved by reducing the battery weight. Also, the ES is more expensive than a conventional scooter, approximately 60 thousand new Taiwan dollars; although the Taiwan government will sponsor a part of this price. And the worst problem for ES is too short cruise range. Electric vehicles have longer cruise range, but they are too large for Taiwan's crowded conditions. Therefore, since people in Taiwan are more likely to buy ES, we work to popularize the ES in Taiwan.

To increase the ES's cruise range, we utilize a supercapacitor as a the second energy source for the ES. Hybrid electric scooters not only offer higher energy efficiency but also increased cruise range compared with conventional electric scooters.

This paper discusses the methodology for designing system-level scooters using the Matlab/Simulink. A hybrid electric scooter has been designed using the simulation package. The simulation results are discussed for our hybrid electric scooter.

2. System description

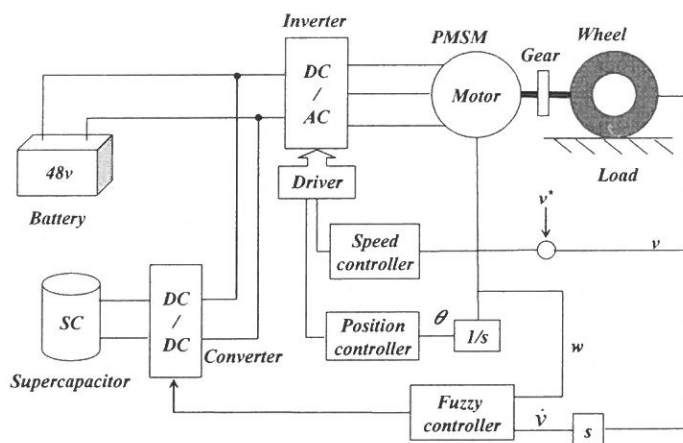


Fig. 1 Full electric scooter system with supercapacitor

Figure 1 shows the block diagram for the proposed HES system, which can be separated into five sub-systems: battery, supercapacitor, inverter, permanent magnetic synchronous motor (PMSM), transmission gear, and road load. The full electric scooter system has four lead acid cells in the battery system and total of 48v. This well developed technology provides more stability then that of lithium ion battery, and it is popular due to lower cost. however, its drawbacks are low specific energy, short life cycle, and environmental pollution. Thus, use the Li-Ion battery in place of lead acid one.

The supercapacitor is a relatively new electrical energy storage device with characteristics that lie between a battery and a dielectric capacitor. For the same volume, the capacitance of the supercapacitor is 100 times higher than that of standard dielectric capacitor. Their instantaneous specific power is at least 20 times higher than that of conventional battery, but their energy density is 20 to 50 times lower. In terms of technical maturity, ease of manufacture, expected performance and cost, the activated carbon based supercapacitor is the most promising device for most current applications. In addition, organic electrolytes are preferred since they provide acceptable levels of energy storage. Due to the highly reversible charge storage process in the supercapacitor, the cycle life is much longer than that of electrochemical batteries and is comparable to that of passive components. The main advantages of supercapacitors are high power density (7~10 kW/kg) and long cycle life (0.5~1 million cycle) [15].

Figure 2 shows a bidirectional dc-to-dc converter, which is employed the studied system to control the input/output current of the supercapacitor in the buck-boost operation mode. The control strategy of the buck-boost converter is a very important issue for the proposed BEMS of the HES. Under ideal conditions, the scooter can utilize an almost constant battery current, with the capacitor giving all the positive and negative variations around the constant battery current. However, we must also consider that the supercapacitor is currently quite expensive.

Control strategies depend on scooter acceleration:

If the ES is operating at high speed (higher than 10km/hr), the capacitor should be empty to be able to receive the energy coming from suddenly emergency stop. By contrast, when the ES is going to start, all the capacitor energy will be required. If the ES accelerates, the supercapacitor will provide energy for that; and when it decelerates, the supercapacitor will recover.

Control strategies depend on motor velocity:

When scooter speed increases, the motor velocity also increases, pulling more energy from the supercapacitor to the motor. Then, when the scooter speed is equal to drive command, the motor velocity will drop down to a stable value to keep the scooter moving. In this phase, there is also some energy pulled from the supercapacitor to the motor. Then, when scooter speed is decreased, the motor velocity is also drop down to zero and the energy which produced from decreasing the velocity will be recovered by the supercapacitor.

Operation mode of converter

If the battery voltage goes up rapidly, the controller activates the buck mode, and a given amount of energy is transferred to the supercapacitor. This situation happens when the ES is running and the brake is activated. Under this condition, the previous state of the overall system should have kept the supercapacitor voltage at low levels. By contrast, the boost mode will be activated when the battery voltage goes down (acceleration), and when the ES is going to move, or it is accelerating from low to high speeds (pulling current out from the batteries). Under these conditions, and if the battery is not fully charged, the supercapacitors should be at high levels of stored energy [3]. The entire active mode is show in Fig. 2. During the boost mode, IGBT2 is switched on and off in a controlled duty cycle, to transfer the required amount of energy from the supercapacitor to the motor and battery. When IGBT2 is turned on, energy is taken from the supercapacitor, and stored in the Ls. When IGBT2 is turned off, the energy stored in Ls is transferred into C, through D1,

and then into the battery and motor. During buck mode, the converter introduces energy from the battery to supercapacitor. When IGBT1 is switched on, the energy goes from the battery to the supercapacitor, and L_s stores part of this energy. When IGBT1 is turned off, the remaining energy stored in L_s is transferred inside the supercapacitor. A fuzzy logic controller is employed to control the operation mode of the buck-booster converter according to the control strategies that depend on motor velocity, and describing as above mentioned.

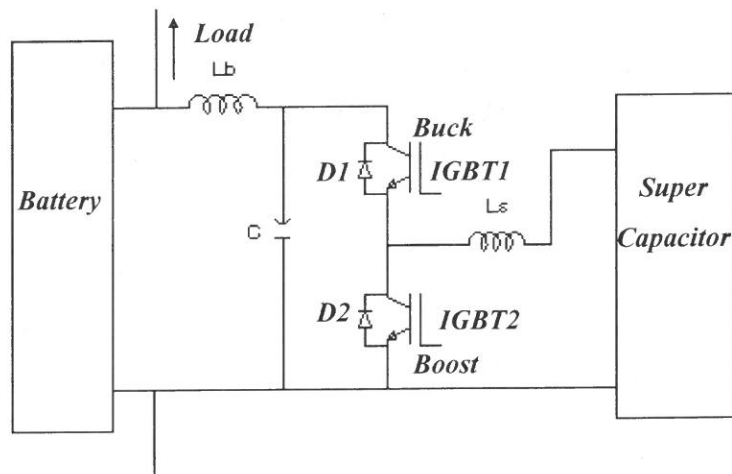


Fig. 2 Configuration of the bidirectional DC-DC converter

The PMSM offers several advantages for application of ES. For one, provides high acceleration and a good torque, namely, a high torque-to-inertia ratio, and an excellent power factor, which is close to unity since the copper losses are essentially located only in the stator. In addition, for the same delivered mechanical power, a PMSM needs a smaller line current value, which is favorable for the design of the electronic power converter feeding this drive [12]. The IGBT inverter is employed to control the PMSM. The IGBT has a high impedance gate, which requires only a small amount of energy to switch the device. In addition, the IGBT has a small on-state voltage even in devices with large blocking voltage ratings. Further more, the IGBT can be designed to block negative voltages [13].

The transmission gear has fixed ratios in the scooter mechanism, and the wheel

directly receives power from the motor through a transmission gear reduction with selected gear ratios. The road load utilizes an experimental parameter to imitate the road resistance, wind resistance, the grade resistance, etc. The grade resistance is a component of gravitation at force on the scooter in the direction of ramp and braked resistance.

3. Mathematical model

A. Battery

The battery is a conventional lead acid cell, and in a conventional ES, it provides input power to set the ES into motion. The physical equation of the Shepherd battery model is[10]:

$$V_b = E_s - iR - Ki \frac{Q}{Q - \int i dt} = E_s - iR - Ki \frac{1}{SOC} \quad (1)$$

where

R : battery internal resistance (Ω)

K : polarization resistance (Ω)

SOC : state of charge

Q : capacity (Ah)

Fig. 3 is a simple battery model for easier simulation.

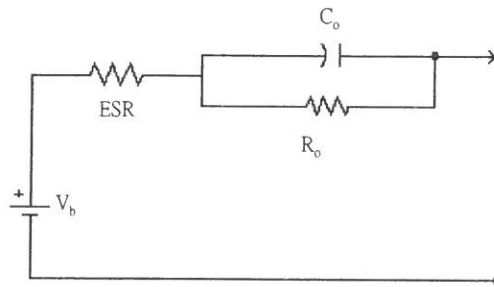


Fig. 3 Thevenin battery model [14]

where

ESR: a constant equivalent internal series resistance

V_b : battery voltage with no load voltage

C_o : overvoltage capacitance

R_o : overvoltage resistance

B. Supercapacitor

Figure 4 shows a physical model of the supercapacitor. The model consists of three elements. The capacity C represents the capacity of the ultra capacitor. R_s represents the charging and discharging resistance of the supercapacitor and accounts for the associated losses [2].

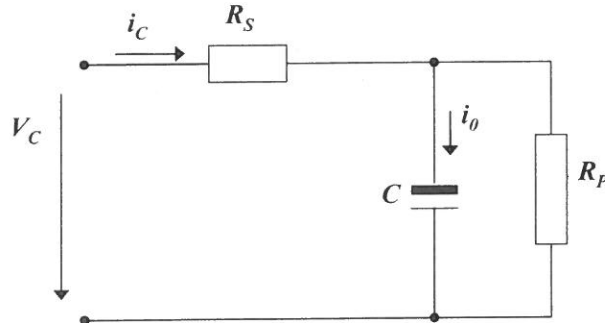


Fig. 4 Physical model of the supercapacitor system

$$i_c = i_0 + \frac{1}{R_p C} \cdot \int_0^t i_0 \cdot dt \quad (2)$$

$$v_c = R_s \cdot i_c + \frac{1}{C} \cdot \int_0^t i_0 \cdot dt \quad (3)$$

where

i_c : the current flow supercapacitor (A)

i_0 : net current stored in supercapacitor (A)

R_p : parallel resistance R_p responsible for the energy loss due to supercapacitor

self-discharge (Ω)

C. Motor

It is important that choose an ac motor with good performance to replace the convention engine as the driving machine of the ES. The PMSM's mathematical model can be shown as:

• Electrical system:

$$L_d \frac{d}{dt} i_d = v_d - R i_d + L_q \dot{\omega}_r i_q \quad (4)$$

$$L_q \frac{d}{dt} i_q = v_q - R i_q + L_d \dot{\omega}_r i_d - \lambda p \dot{\omega}_r \quad (5)$$

$$T_e = 1.5 p [\lambda i_q + (L_d - L_q) i_d i_q] \quad (6)$$

where

L_q, L_d : q and d axis inductances (H)

R : resistance of the stator windings (Ω)

i_q, i_d : q and d axis currents (A)

v_q, v_d : q and d axis voltages (V)

ω_d : angular velocity of the rotor (rad/s)

λ : amplitude of the flux induced by the permanent magnets of the rotor in the stator phase (Wb)

p : number of pole pairs

T_e : electromagnetic torque (Nm)

• Mechanical system:

$$J \frac{d}{dt} \omega_r = (T_e - F \omega_r - T_m) \quad (7)$$

$$\frac{d\theta}{dt} = \omega_r \quad (8)$$

where

J : combined inertia of rotor and load ($\text{kg}\cdot\text{m}^2$)

F : combined viscous friction of rotor and load ($\text{N}\cdot\text{m}\cdot\text{s}$)

θ : rotor angular position (rad)

T_m : shaft mechanical torque (Nm)

D. Load

Here, we discuss the transmission gear of the electric scooter, which has the equation:

$$\frac{T_L}{T_e} = \frac{\omega_r}{\omega_L} = \frac{r_L}{r_{motor}} \frac{N_L}{N_{motor}} = n \quad (9)$$

where

T_L, ω_L : torque and rotational velocity of the load

T_e, ω_r : torque and rotational velocity of the motor

r_L, N_L : radius and tooth number of the load side gear wheel

r_{motor}, N_{motor} : radius and tooth number of the motor side gear wheel

In addition, the transfer functions for torque and force can be expressed as:

$$F_{out} = \frac{T_e \times \omega_r}{v} \quad (10)$$

The load equation is:

- The Road resistance:

$$F_{road} = M \times (A_d + B_d \times v_c) \quad (11)$$

where

F_{road} : road resistance (N)

A_d, B_d : road resistance coefficient

M : weight of the scooter (Kg)

v_c : speed of the scooter (m/s)

- The grade resistance is a component of gravitation of ES in the direction of ramp, expressed as:

$$F_g = Mg \sin \theta \quad (12)$$

where

g : gravity (m/s^2)

θ : angle of the slope (rad)

- The aerodynamic resistance acts on the ES, expressed as:

$$F_{wind} = \frac{1}{2} \times C_d \times \rho \times A \times (v + v_w)^2 \quad (13)$$

where

F_{wind} : wind resistance (N)

C_d : wind resistance coefficient ($\text{N} \cdot \text{s/kg}$)

ρ : atmosphere density (kg/m^3)

A : measure of area with the wind (m^2)

v, v_w : speed of the scooter and wind (m/s)

- To calculate the ES speed:

$$F_{load} = F_{road} + F_r + F_{wind} \quad (14)$$

$$a = \frac{F_{out} - F_{load}}{M} \quad (15)$$

$$v = \int a dt \quad (16)$$

E. Loss

To observe the efficiency of whole system we calculate the power and energy loss, using the following equations.

- Input power:

$$P_{in} = V \times i \quad (17)$$

where

P_{in} : input power (w)

V : input voltage (v)

i : input current (A)

• Output power:

$$P_{out} = v \times F \quad (18)$$

where

P_{out} : output power (w)

v : scooter speed (m/s)

F : output force (N)

• Power loss:

$$P_{loss} = P_{in} - P_{out} \quad (19)$$

• Energy loss:

$$Energy\ loss = \int P_{loss} dt \quad (20)$$

Figure 5 shows the proposed system in Matlab/Simulink, using its the default IGBT inverter and DC-DC converter. Thus, we assume that the loss of inverter and converter can be neglected.

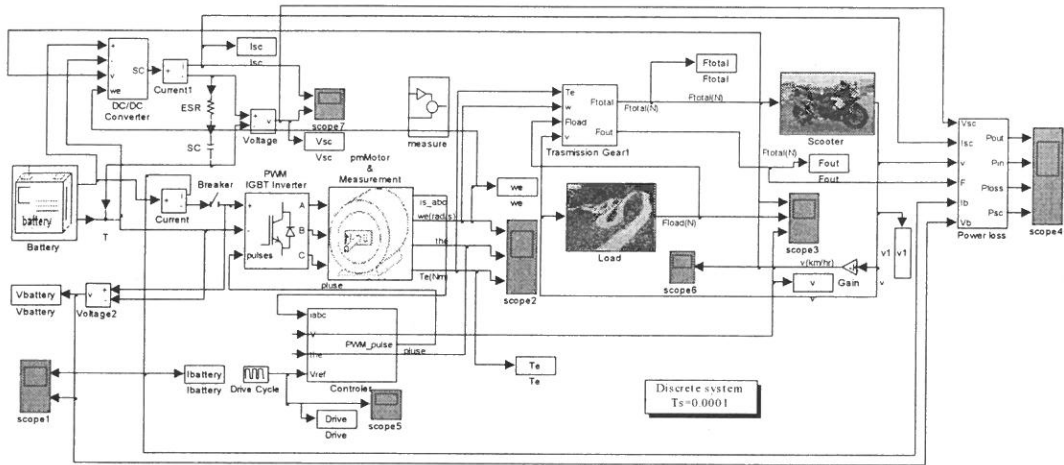
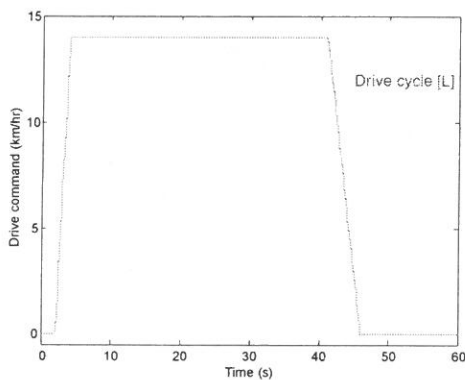


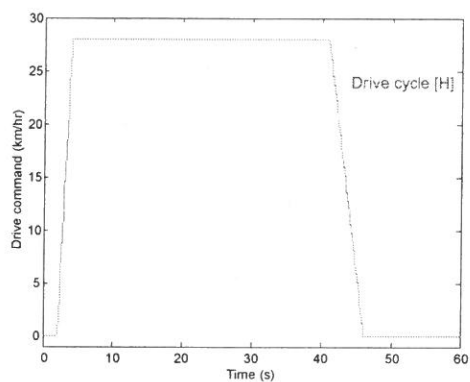
Fig. 5 Overall system model in Matlab/Simulink

4.Simulation

The model in Matlab/Simulink is shown in Fig. 5. For simulation of the scooter, the drive cycle will be the input of the whole system, including two kinds of drive cycles, as shown in Fig. 6. The simulation results will demonstrate the speed of scooter, motor velocity, power loss, supercapacitor energy and total energy of the studied system. Figure 7 shows the speed of scooter, and Figure 8 shows motor velocity. Comparing Fig. 7 and Fig. 8, when the speed of scooter is stable, the motor velocity will slow down and maintain a particular speed. There are some scooter speed transient during the change state of the driving cycle. The impulse and jitter in Fig. 8 would be reduced by a integrator which is included in the transfer function of torque and force and it acts as a low pass filter. It means that the motor no longer exports more power to accelerate. In Figures 9 and 10, we discover that the curve has a negative part when the scooter speed is stable. This part is called negative power in the physical meaning, and it reduces the energy loss. Comparing Fig. 9 and Fig. 10, the curve in Fig. 10 has more negative power than Fig. 9. So the hybrid electric scooter has less energy loss than convention ES.

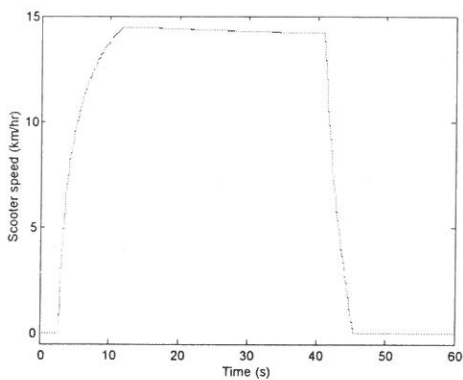


(a)

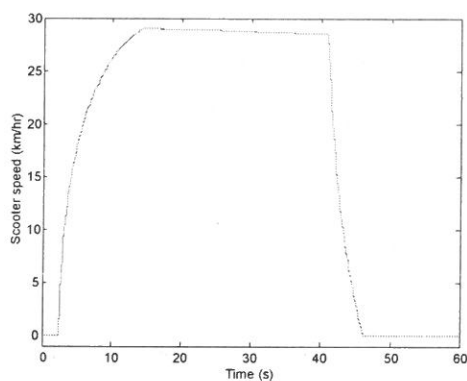


(b)

Fig. 6 Drive cycle : (a) top speed 14km/hr (drive cycle [L]), (b) top speed 28km/hr (drive cycle [H])

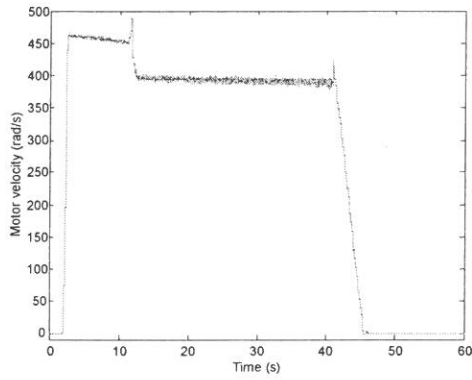


(a)

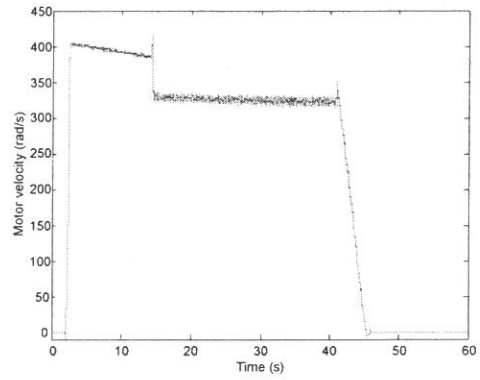


(b)

Fig. 7 Scooter speed vs. time : (a) drive cycle [L], (b) drive cycle [H]

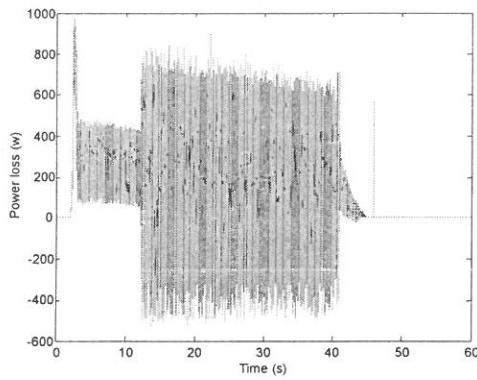


(a)

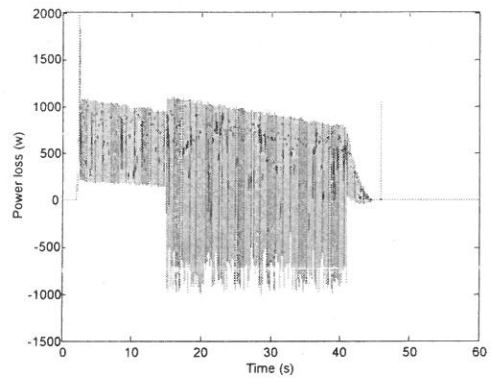


(b)

Fig. 8 Motor rotational velocity vs. time : (a) drive cycle [L], (b) drive cycle [H]



(a)



(b)

Fig. 9 Power loss vs. time for the conventional ES : (a) drive cycle [L], (b) drive cycle [H]

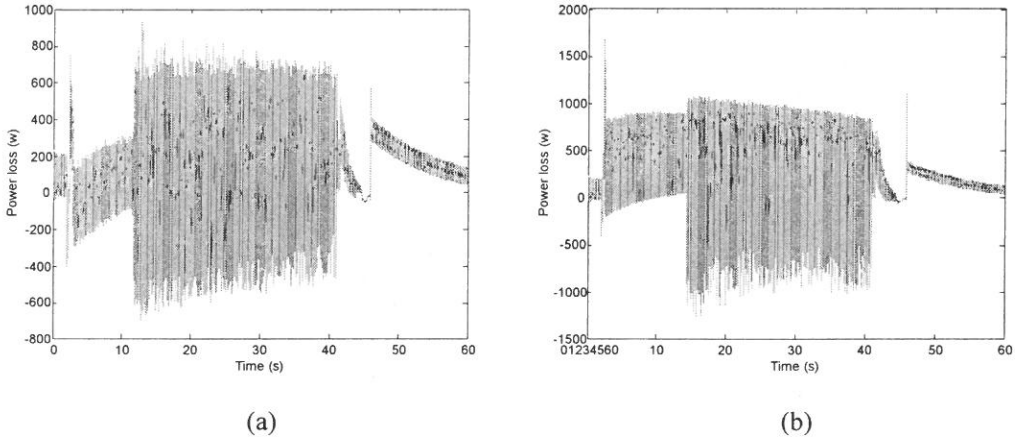


Fig. 10 Power loss vs. time for the proposed hybrid electric scooter : (a) drive cycle [L], (b) drive cycle [H]

We show the supercapacitor energy in Fig. 11, which also shows the operation mode of dc-to-dc converter versus time, implementing the strategy presented in Section 2. When the curve's slope is positive, then the supercapacitor is discharging; and when it is negative, it is recovering. The four curves in Fig. 12(a) are the integrals of Fig. 9 and 10, showing the energy loss in HES or convention ES, and drive cycle [L] or [H]. Line 1 and Line 2 in Fig. 12(a) are the energy loss of drive cycle [H]. Figure 12(b) shows that the HES is more energy-efficient (about 10%~30% increasing) than the conventional energy management system of ES. Consequently, we can confirm that if the battery has a parallel supercapacitor, its energy loss will be reduced.

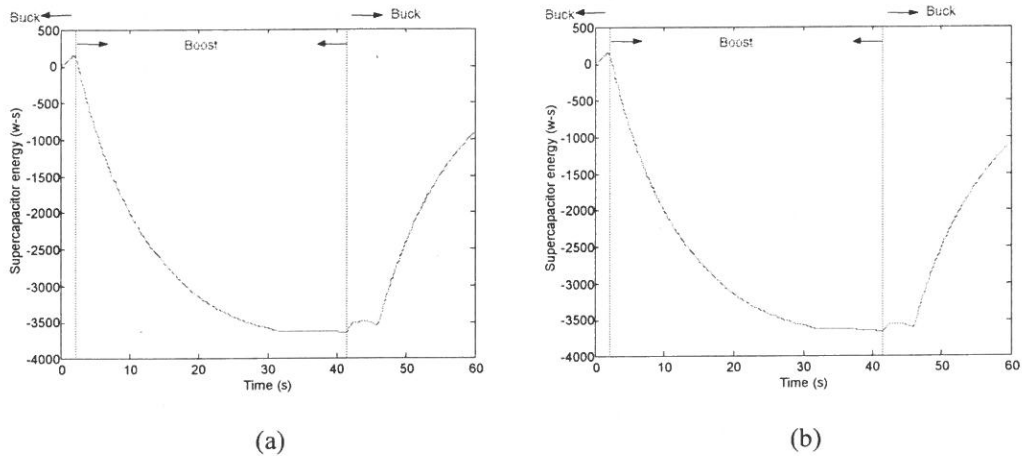


Fig. 11 Supercapacitor energy vs. time : (a) drive cycle [L], (b) drive cycle [H]

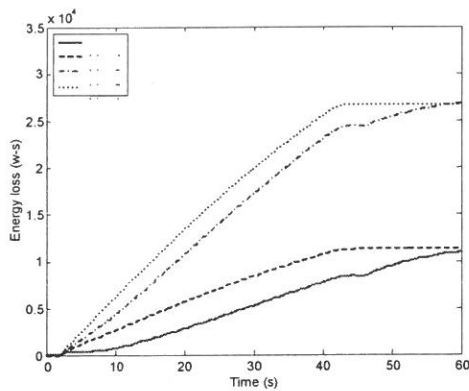


Fig. 12(a) Energy loss vs. time

Line 1: drive cycle [H], convention ES

Line 2: drive cycle [H], hybrid electric scooter

Line 3: drive cycle [L], convention ES

Line 4: drive cycle [L], hybrid electric scooter

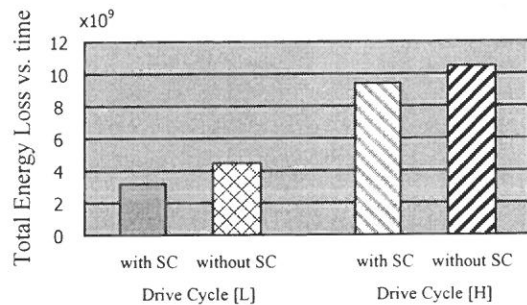


Fig. 12(b) Total energy loss vs. time

Fig. 12 Energy simulation result

5. Conclusions

This paper presents a new drive modeling for a hybrid electric scooter. The simulation is developed using Matlab/Simulink to compare the conventional ES and the hybrid ES. The simulation package use mathematical equations to establish the hybrid electric scooter model, computing the power loss and energy of the HES by using the proposed model. The simulation results show that HES is more energy-efficient (about 10%~30% increasing) than the conventional energy management system of ES. Furthermore, the proposed control strategy can be extend the ES cruise range.

Acknowledgement

The author wishes to express his thanks for the financial support of the Societas Verrbi Divini.

Reference

- [1] Karen L. Butler, Mehrdad Ehsani, and Preyas Kamath, "A Matlab-Based Modeling and Simulation Package for Electric and Hybrid Electric Vehicle Design," IEEE Transactions on Vehicular Technology, vol.48, no.6, pp.1770-1778, 1999.
- [2] Thomas Dietrich, "UltraCaps – A new Energy storage Device for Peak Power Applications," 18th Electric Vehicles Symposium, October, 2001.
- [3] Karl-Heinz Hauer, and Badri Narayanan, "Numerical Simulation of Two Different Ultra Capacitor Hybrid Fuel Cell Vehicles," 18th Electric Vehicles Symposium, October, 2001.
- [4] Juan W. Dixon, and Micah Ortúzar, "Test Results with Regenerative Barking based on Ultracapacitors and a Buck-Boost Converter," 18th Electric Vehicles Symposium, October, 2001.
- [5] Detlef Heinemann, and Dietrich Naunin, "Ultracaps in power-assist applications in Battery Powered Electric Vehicle- Implications on Energy Management

- Systems,” 18th Electric Vehicles Symposium, October, 2001.
- [6] A. Di Napoli, F. Crescimbeni, L. Solero, G. Pede, G. Lo Bianco, and M. Pasquali, “Ultracapacitor and Battery Storage System Supporting Fuel-Cell Powered Vehicle,” 18th Electric Vehicles Symposium, October, 2001.
- [7] Jörg Lott, and Helmut Späth, “Double Layer Capacitors as additional Power Source in Electric Vehicles,” 18th Electric Vehicles Symposium, October, 2001.
- [8] S. Buller, E. Karden, D. Kok, and R. W. De Doncker, “Simulation of Supercapacitors in highly dynamic Application,” 18th Electric Vehicles Symposium, October, 2001.
- [9] Michio Okamura, “A Progress Report of the Capacitor Hybrid System – ECS,” 18th Electric Vehicles Symposium, October, 2001.
- [10] 林昆民, “電動機車整體性能之動態模擬分析,” 國立清華大學動力機械工程研究所, 碩士論文, June, 2000.
- [11] 廖峻慶, 葉勝年, 黃仲欽, “電動機車動力測試平台之研製,” 行政院國家科學委員會輔導專題研究計畫, 成果報告, August, 2001.
- [12] Damien Grenier, L.-A. Dessaint, and Ouassima Akhrif, “Experimental Nonlinear Torque Control of a Permanent-Magnet Synchronous Motor Using Saliency,” IEEE Transactions on Industrial Electronics, Vol. 44, No. 5, pp.680-687, October 1997.
- [13] Ned Mohan, Tore M. Undeland, and William P. Robbins, “Power Electronics Converters, Applications, and Design,” John Wiley & Sons, Inc., 1995.
- [14] H.L. Chan, and D. Sutanto, “A New Battery Model for Use with Battery Energy Storage Systems and Electric Vehicles Power Systems,” Power Engineering Society Winter Meeting, IEEE, Vol. 1, pp.470-475, 2000.
- [15] Andrew Burke, and Marshall Miller, “Comparisons of Ultracapacitors and Advanced Batteries for Pulse Power in Vehicle Applications: Performance, Life, and Cost,” 19th Electric Vehicles Symposium, October, 2002.

received September 27, 2002

revised October 24, 2002

accepted November 8, 2002

超高電容器在電動機車電池能源系統中的應用

張佳祺 李永勳*

輔仁大學電子工程系

摘 要

本論文提出具有超高電容器組和鉛酸電池模組的電動機車混成能源系統的能源管理系統研究，研究中將建立混成能源管理系統的功率管理模型，完整的能源管理模型包括牽引馬達模型、車輪和道路模型、變速系統模型和電力電子轉換器導通模型、鉛酸電池模型和超高電容器模型，並利用 Matlab/Simulink 建立視窗交談式能源管理系統之模擬，驗證其能減少能源損失及增加行車距離的目的。

關鍵詞： 超高電容器、電池能源管理系統、混成能源系統、系統模擬、電動機車

An Evolving Bidding Strategy for Intelligent Agents in Multiple Auctions

Jong Yih Kuo*

Department of Computer Science and Information Engineering

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

Due to the desire of almost all departments of business organizations to be interconnected and to make data accessible at any time and any place, more and more multi-agent systems are applied to business management. As numerous agents are roaming through the Internet, they compete for the limited resource to achieve their goal. In the end, some of them will succeed, while the others will fail. However, when agents are initially created, they have little knowledge and experience with relatively lower capability. They should also strive to adapt themselves to the changing environment. It is advantageous if they have the ability to learn and evolve. This paper addresses evolution of intelligent agents in virtual enterprises. Agent fitness and fuzzy multi-criteria decision-making approach are proposed as evolution mechanisms, and fuzzy soft goal is introduced to facilitate the evolution process. Genetic programming operators are employed to restructure agents in the proposed multi-agent evolution cycle. We conduct a series of experiments to determine the most successful strategies and to see how and when these strategies evolve depending on the context and negotiation stance of the agent's opponent.

Key Words: Intelligent agent, Evolving strategy, Fuzzy logic, Genetic algorithm

*Corresponding author. Tel.: +886-2-29031111 ext. 2444; fax: +886-2-29023550

E-mail address: jykuo@csie.fju.edu.tw

1. Introduction

Doing business on the Internet is becoming more and more popular. The use of the Internet to facilitate commerce among companies and customers brings forth many benefits, such as automated transactions, greater access to buyers and sellers, and dramatically reduced costs. The agent-based e-commerce has emerged and become a focus of the next generation of e-commerce. The intelligent agent acts on behalf of customers to carry out delegated tasks automatically. They have demonstrated tremendous potential in conducting various e-commerce activities, such as comparison-shopping, auction, sales promotion, etc [1,2].

In order to solve a problem, an agent has to have certain skills and the ability to reason about these skills. We call the reasoning abilities as “mental” skills [3]. However, when agents are initially created, they have little knowledge and experience with relatively lower capability. They should also strive to adapt their negotiation strategies and tactics to the changing environment. It is advantageous if the agents have the ability to learn and evolve. Many issues are essential in agent evolution. Firstly, evolution of an agent is closely related with agent structure. Thus, a suitable agent structure is one of basic concerns in agent evolution. Secondly, agents should have their own mechanisms to advance evolution. Thirdly, in a multi-agent system, evolution of individual agent is also related with many social concerns, such as coordination, negotiation, communication, etc. Finally, some tools can be used to evaluate the fitness of agents in the evolution procedures..

In this paper, we address multi-agent evolution for agents in e-commerce. Section 2 summarizes our service-oriented negotiation model based on fuzzy theory and the BDI model. In Section 3, we adopt an evolutionary approach in which strategies and tactics correspond to the genetic material in a genetic algorithm. In Section 4, we present an empirical study showing the relative success of different strategies against different types of opponent in different environments. Section 5 contains our conclusion. The following diagram illustrates our flow of approach (see Figure 1).

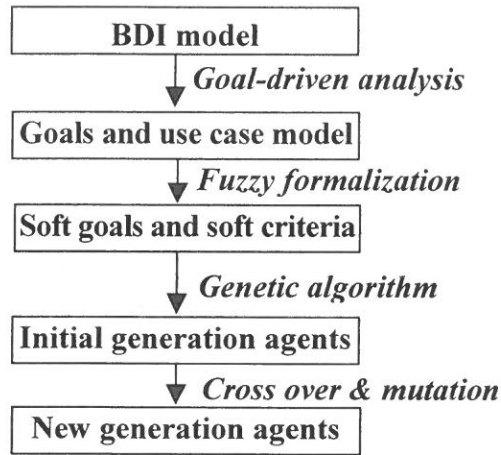


Figure 1: The architecture of our approach

2. The Service-oriented Negotiation Model

This paper addresses evolution of intelligent agents about the mental skills. Obviously, there is no limit to what one would like to include under what we call mental skills. We agree that BDI model [4,5,6] provides a simple but powerful formalism for the representation, the specification and the analysis of the mental attributes of intelligent agent: belief, desire and intention.

2.1 The BDI Model

In the BDI architecture an agent can be completely specified by the events that it can perceive, the actions it may perform, the beliefs it may hold, the goals it may adopt, and the plans that give rise to its intentions [7]. The figure 2 represents the relationships of BDI model.

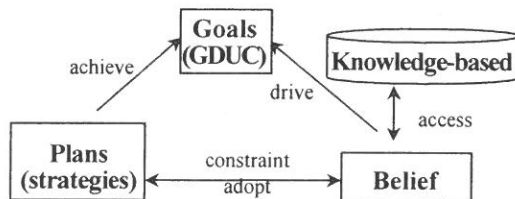


Figure 2: The BDI model of Agent

A belief model describes the information about the environment and internal state that an agent of that class may hold, and the strategies and tactics it may perform. A goal model (desires) describes the goals that an agent may possibly adopt, and the events to which it can respond. It consists of a goal set which specifies the goal and event domain and one or more goal states – sets of ground goals – used to specify an agent's initial mental state. There are soft and rigid goals specified by the users. We use fuzzy logic to represent the goals [8].

A plan model (intensions) describes the plans that an agent may possibly employ to achieve its goals. A plan is a sequence of strategies through reasoning mechanism (mental skills of the agent). The strategy is the combination of tactics with various weights.

2.2 The Goal-driven Analysis

To model user the BDI model, we use GDUC (goal-driven use case) approach [9] to structure the goals hierarchy and to analyze the plans or strategies achieving these goals. The steps describe below.

- (1) Identify actors and user's goals to construct belief model: First, we must analyze the organization of enterprise or the environment of e-commerce to extract the basic knowledge for the agent. The knowledge can be build into a general common ontology. We also identify the users and their preferences to build the specific user-defined ontology. The ontology hierarchy can be stored into the knowledge-based of belief model.
- (2) Analyze goal hierarchy to build goal model: A faceted classification is proposed for identifying goals from domain descriptions and system requirements. Each goal can be classified under three facets we have identified: competence, view and content. The facet of competence is related to whether a goal is completely satisfied or only to a degree. A rigid goal describes a minimum requirement for a target system, which is required to be satisfied utterly. A soft goal describes a desirable property for a target system, and can be satisfied to a degree. The facet

of view concerns whether a goal is actor-specific or system-specific. Actor-specific goals are objectives of an actor in using a system; meanwhile, system-specific goals are requirements on services that the system provides. We use the “use case” to structure goals hierarchy.

- (3) Analyze goal model to build plan model: According to the user’s goal and use cases, we can construct the scenarios of use cases and the possible planes to achieve the goals. Then we also evaluate the degrees of satisfaction about the planes. The ability of context sensitivity and evolution help agent adopt the negotiation strategies to achieve user’s goals.

2.3 Applying Fuzzy Theory to BDI Model

To model user goals, we apply GDUC to get a set of soft and rigid goals, a set of use cases, and a set of planes. For achieving these goals, agent must use particular strategies to change their mental states. We can continuously change the problem state to achieve the goal state. Thus we can apply the soft requirement [10] to formally represent the user goals.

A user goal, g , is specified by the properties of agent’s mental state-transition $\langle b, g, a \rangle$, where b is the state before a plan, and a is the state after invoking the plan. A plan or strategy can thus be specified using a pair $\langle precondition, post-condition \rangle$. The precondition and the post-condition describe properties that should be held by the state b and a . A rigid goal describes state properties that must be satisfied. The soft goal describes state properties that can be satisfied to a degree. We use Zadeh’s test-score semantic [11] to represent the user goals. A basic idea underlies test-score semantics is that a proposition p in a natural language may be viewed as a collection of elastic constraints, C_1, \dots, C_k , which restricts the values of a collection of variables $X = (X_1, \dots, X_n)$. In fuzzy logic, this is accomplished by representing p in the canonical form:

$$G \Rightarrow R(P) \text{ IS } A \quad (1)$$

in which A is a fuzzy predicate. The canonical form G implies that the possibility distribution of $R(P)$ is equivalent of the membership function of A , namely $\Pi_{R(P)} = \mu_A$. For example, the agent helps a user to buy high quality item and can be represented using the canonical form below:

$$G \Rightarrow \text{Quality(goods) IS HIGH} \quad (2)$$

Where **HIGH** is a fuzzy predicate. The rigid goal is the specialization of the soft goal, which membership function of fuzzy predicate is 1.0.

2.4 The Negotiation Strategies

A negotiation strategy is the combination of tactics with various weights. A tactic generates a value for a single negotiation issue based upon a single criterion (e.g. time remaining, resource remaining).

As negotiation proceeds, the goals of agent may become relevant and the relative importance of existing criteria may vary. To reflect this fact, an agent has a strategy that varies the weights of the different tactics over time in response to various environmental. We extended the research [12,13] to proposed four types of tactics described below.

(1) Time-dependent tactics

These tactics model the fact that the agent is likely to concede more rapidly as the negotiation deadline approaches. The negotiation must have completed at the pre-established deadline (t_{\max}). The maximum price is P_r . When the deadline is nearly up, the price approaches the P_r . The function of tactics:

$$f_t = \alpha_t(t) P_r \quad (3)$$

$$\alpha_t(t) = k_t + (1 - k_t) \left(\frac{t}{t_{\max}} \right)^{1/\beta_t}$$

$$0 \leq \alpha_t(t) \leq 1, \quad \alpha_t(0) = k_t, \quad \alpha_t(t_{\max}) = 1,$$

$$0 \leq k_t \leq 1, 1/200 \leq \beta_t \leq 1000.$$

(2) Resources-dependent tactics

These tactics generate offers depending on how a particular resource is being consumed; they become progressively more conciliatory as the quantity of resource diminishes. Here, we use the bidder tactics. The equation is:

$$f_r = \alpha_r(t) P_r. \quad (4)$$

$$\alpha_r(t) = k_r + (1 - k_r) \left(\frac{c(t)}{|A|} \right)^{1/\beta_r}$$

$$0 \leq k_r \leq 1, 1/200 \leq \beta_r \leq 1000.$$

$c(t)$ is the number of web at $0 \sim t$, $|A|$ is the number of active bidding web at $0 \sim t_{max}$.

(3) Prices-dependent tactics

Agent uses these tactics to maintain the goal of minimum price. Agent must get the bidding prices of all active bidding webs. The equation is:

$$f_p = \omega(t) + \alpha_p(t)(P_r - \omega(t)) \quad (5)$$

$$\alpha_p(t) = k_p + (1 - k_p) \left(\frac{t}{t_{max}} \right)^{1/\beta_p}$$

$$0.1 \leq k_p \leq 0.3, 1/200 \leq \beta_p \leq 0.5$$

$$\omega(t) = \frac{1}{|L(t)|} \left(\sum_{1 \leq i \leq |L(t)|} \frac{t - \sigma_i}{\eta_i - \sigma_i} v_i(t) \right)$$

$|L(t)|$ is the number of active bidding webs at time t . The η_i represents the start time of the i th bidding web. The σ_i is the end time of the i th bidding web. The $v_i(t)$ represents the highest price of the i th bidding web at time t .

(4) Desire-dependent tactics

Agent does the best to buy the high quality goods to achieve the user desired. The curve of the price will quickly approach the P_r . The equation is:

$$f_d = \omega(t) + \alpha_d(t)(P_r - \omega(t)) \quad (6)$$

$$\alpha_d(t) = k_d + (1 - k_d) \left(\frac{t}{t_{\max}} \right)^{1/\beta_d}$$

$$0.7 \leq k_p \leq 0.9, 1.67 \leq \beta_d \leq 1000$$

3. The Evolution of Intelligent Agent

Genetic algorithm (GA) operators are employed to restructure agents in the proposed multi-agent evolution cycle. How to encode a solution of the problem into a chromosome is a key issue for genetic algorithms. In Holland's work [14] encoding is carried out using binary strings. For many GA applications, the simple GA was difficult to apply directly because the binary string is not a natural coding. During the past ten years, various non-string encoding techniques have been created for particular problems, for example, real number coding for constrained optimization problems and integer coding for combinatorial optimization problems. In our research, a real-coded GA uses floating-point numbers to represent genes [15].

3.1 Coding Schema

In order to find proper intelligent agent, the agent's negotiation strategies are coded and represent an individual. A strategy that is the combination of tactics with various weights will determine the bidding price at time t . We have three categories of tactics: time- dependent, resource-dependent, behavior- dependent. The equation of a strategy describes below:

$$S(t) = w_t f_t + w_r f_r + w_p f_p + w_d f_d \quad (7)$$

$$0 \leq w_t, w_r, w_p, w_d \leq 1,$$

$$w_t + w_r + w_p + w_d = 1$$

Each agent is represented as a string of fixed length. The bits of the string (the

gene) represent the parameters of the agent's strategy.

$$G = (t_{max}, d, r, k_1, \beta_1, w_1, k_2, \beta_2, w_2, k_3, \beta_3, w_3, k_4, \beta_4, w_4) \quad (8)$$

3.2 Measuring a Strategy's Fitness

A fitness function is the survival arbiter for individuals. For finding the near-optimal intelligent agent, we propose the fuzzy multi-criteria decision-making (FMCDM) approach as the evolution mechanisms, and the fuzzy soft goals to facilitating the evolution process.

We may analyze the user's goals into a goal model. Each goal can have some criteria. The evaluation of soft goal is a satisfaction degree. The relationships between goals will exist conflicting and cooperative. Most of the existing approaches in multiple criteria making lack the aspect of an explicit modeling of relationships between goals. Carlsson and Fuller [16] advocated that much closer to MCDM in the real world than the traditional MCDM are the case with interdependent criteria. Our previous work on Criteria Trade-off Analysis (CTA) has been on the formulation of soft criteria based on Zadeh's canonical form in test-score semantics and an extension of the notion of soft condition [17]. The trade-off among soft goals is analyzed by identifying the relationships between goals. A compromise overall satisfaction degree can be obtained through the aggregation of individual goal based on the goals hierarchy. The steps describe below.

- (1) To compute the relationship between goals: The c, c' are two soft goal, and a is the strategy, CF and CP denote the set of conflicting and cooperative pairs. AP denotes the set of all pairs. The conflicting and cooperative degree between two goals is defined as:

$$cf(c, c') = \frac{\sum_{(a, a') \in CF} (|\mu(a) - \mu(a')| + |\mu'(a) - \mu'(a')|)}{\sum_{(a, a') \in AP} (|\mu(a) - \mu(a')| + |\mu'(a) - \mu'(a')|)} \quad (9)$$

$$cp(c, c') = \frac{\sum_{(a_i, a_j) \in CP (|\mu_c(a_i) - \mu_c(a_j)| + |\mu_{c'}(a_i) - \mu_{c'}(a_j)|)}{\sum_{(a_i, a_j) \in Ap (|\mu_c(a_i) - \mu_c(a_j)| + |\mu_{c'}(a_i) - \mu_{c'}(a_j)|)} \quad (10)$$

- (2) To covert connections of the goals into DNF (Disjunctive Normal Form), and to establish a goals hierarchy: We assume that goals specified by users are connected by linguistic connectives in natural language. To take these connectives into account, we proposed the use of DNF to obtain a uniform representation of the goals. According to the conflicting and cooperative degrees, a goals hierarchy of n levels is defined as a tree. This tree is important in the sense that the ordering established through the hierarchy helps alleviate the associative problem inherited in fuzzy aggregation operator.
- (3) By using fuzzy aggregation operator to compute the strategy's fitness: An extended goals hierarchical aggregation structure is proposed to facilitate goals aggregation through the fuzzy and / or operator. The fitness can be obtained through the aggregation of satisfaction degrees based on the aggregation structure.

3.3 The Evolution Steps

All GAs use some form of mechanism to chose which individuals from the current population should go into the mating pool that forms the basis of the next population generation. A selection mechanism known to work well in such circumstances is Tournament Selection [18]. The crossover process exchanges genetic material between individuals. We randomly select two individuals from the population. Crossover points are then randomly chosen and sorted in ascending order. Then the genes between successive crossover points are alternately exchanged between the individuals, with a probability. Mutation process works by randomly selecting some of the genes present in the population in order to mutate.

The evolution of agent describes below.

- (1) Initial population

A GA requires a population of potential solutions to be initialized at the beginning of the GA process. Here, we randomly generate some genes to create the initial population. We also generate some genes based on the agent's belief model.

(2) Selection Procedure

Selection procedure may create a new good population for the next generation based on either all parents and offspring or part of them [19,20]. A sampling space is characterized by two factors: size and ingredient (parent or offspring). In regular sampling, there are several replacement strategies to replace old parents with offspring when new offspring are produced. As mentioned before, the population of next generation was formed by roulette wheel selection [19]. When selection performs on enlarged sampling space, both parents and offspring have the same chance of competing for survival. An evident advantage of this approach is that we can improve GA performance by increasing the crossover and mutation rates. We don't worry that the high rate will introduce too much random perturbation if selection is performed on enlarged sampling space.

A reproduction operation allows strings that with higher fitness values would have larger number of copies while the strings with lower fitness values have a relatively smaller number of copies or even none at all. This is an artificial version of natural selection (strings with higher fitness values will have more chances to survive). For example, suppose that N strings are generated, and the fitness value of the i th individual string is f_i ($i=1, \dots, N$). Then, the probability of the i th individual string to be selected into the mating pool is

$$p_i = \frac{f_i}{\sum_{i=1}^N f_i}, \quad (11)$$

and the number of copies for the individual string is calculated by

$$n_i = N \cdot p_i \quad (12)$$

This strategy emphasizes the survival-of-the-fitness aspects of the GA. The better

strings receive more copies and go into the mating pool so that their desirable characters may be passed onto their offspring.

(3) Crossover

Crossover is a process to provide a mechanism for two high-fitness strings (parents) to produce two offspring by matching their desirable qualities through a random process [21]. The procedure of crossover is to select a pair of strings from the mating pool at random, then, an integer position k (called the crossover point) along the string is selected uniformly at random between 1 and $(l-1)$, where l is the string length greater than 1. Finally, according to the probability of crossover, two new strings are generated by swapping all characters between position $(k+1)$ and l inclusively. For example, consider two strings A and B of the population are mated for crossover

$A=011|11001$

$B=100|10011$

Suppose we obtain $k=3$ as indicated by the separator symbol "|". The resulting crossover yields two new strings as shown below

$A'=01110011$

$B'=10011001$

where A' and B' are the strings of the new generation. Although the crossover is done by random selection, it is not the same as a random search through the search space. Since it is based on the reproduction process, it is an effective means of exchanging information and combining portions of high-fitness solutions.

An agent is composed of some kinds of modules. Each module has a special feature for a particular plan. In multi-agent system, there are two phase of crossover: the inter-agent and intra-agent. With the inter-agent crossover, an agent exchanges a module with another agent. The intra-agent crossover exchanges some parameters or strategies between modules in the same agent.

(4) Mutation

Mutation is a process to provide an occasional random alteration of the value at a particular string position [22]. In the case of binary string, this simply means

changing the state of a bit from 1 to 0 and vice versa. A uniform mutation is first to produce a mask randomly, then change the selected string value in the position of mask where the bit value is "1". For example, consider the following selected string and generated mask:

$A' = 0\ 1\ 1\ 1\ 0\ 0$

mask $M = 1\ 0\ 0\ 1\ 0\ 1$

then, the 0-1 pattern of the string becomes as the following string,

$A'' = 111100$.

Mutation occurs with a small probability in the GA to reflect the small rate of mutation existing in the real world. Mutation is needed because some digits at a particular position in all strings may be eliminated during the reproduction and crossover operations. Such a situation is impossible to be recovered by using only reproduction and crossover operations. To ensure that reproduction and crossover do not lose some potentially useful genetic materials (1's or 0's at particular locations), in mutation phase, some bits will be changed in all the strings according to the mutation rate (P_m). In general, the mutation rate is less than 0.05. So the mutation plays a role as a safeguard in GA. It can help GA to avoid the possibility of mistaking a local optimum for a global optimum.

4. Experiments and Results

Over the last few years, the number of online auction houses has increased tremendously. To date there are more than 760 auction houses that conduct business online. Some examples of popular online auction house include eBay, Amazon, Yahoo!Auction, Priceline, Ubid, and FirstAuction. The types of auction that are conducted vary from site to site, but the most popular one are English, first-price sealed bid. In the English auction, the auctioneer begins with the lowest acceptable price and bidders are free to raise their bids successively until there are no more offers to raise the bid. The winning bidder is the one with the highest bid [23].

4.1 The Simulation Environment

In this paper, a digit video is the target item. The P_r is its reservation price for the target item. The bidder is given a deadline by when it needs to obtain the item. There are five predefined auctions running in the environment. These auctions have a finite start time and duration generated randomly from a standard probability distribution. The start time and the end time vary from one auction to another. The auction starts with a predefined small starting value. The process is repeated until the reservation price is reached or until the end time for the auction is reached. At the start of each auction, a group of random bidders are generated to simulate other auction participants. These participants operate in a single auction and have the intention of buying the target item and possessing certain behavior. They maintain information about the item they wish to purchase, their private valuation of the item (reservation price), the starting bid value and their bid increment. These values are generated randomly from a standard probability distribution. They start bidding at starting bid value; when making a counter offer, they add their bid increment to the current offer, and they stop bidding when they acquire the item or when their reservation price is reached.

4.2 The Strategy of the Bidder Agent

The bidder agent is allowed to bid in any of the auction at any time when the marketplace is active. The objective of the bidder agent is to participate across the multiple auctions, bid in the auctions and deliver the item to its consumer in a manner that is consistent with their preferences. The bidder agent utilizes the available information to make its bidding decision; this includes the user's reservation price, the time it has left to acquire the item, the current offer of each individual auction, and its set of tactics and strategies. The output of the bidding decision is the auction the agent should bid in and the recommended bid value that it should bid in that auction. The agent's overall behavior is the amalgamation of those strategies proposed in this paper, weighted by their relative importance to the user. Mapping this to an auction environment, the bidder agent needs to decide the new bid value

based on the current offer price. Let t be the current universal time across all auctions, where $t \in \tau$, and τ is a set of finite time intervals. Let t_{max} 爲 be the time by when the agent must obtain the good (i.e. $t_{max} \geq t \geq 0$), and let A be the list of all the auctions that will be active before time t_{max} . At any time t , there is a set of active auctions $L(t)$ where $L(t) \subset A$. Since each auction has a different start and end time, the bidder agent needs to build an active auction list to keep track of all the auction that are currently active in the marketplace. The agent identifies all the active auctions and gathers relevant information about them. It then calculates the maximum bid it is willing to make at the current time using the agent's strategy. Based on the value of the current maximum bid, the agent selects the potential auctions in which it can bid and calculates what it should bid at this time in each such auction. The auction and corresponding bid with the highest expected utility is selected from the potential auctions as the target auction. Finally, the agent bids in the target auction.

4.3 Experimental Evaluation

Our experiments were run in an environment with 10 agents of the first generation, $t_{max} = 100$, 5 English auctions running concurrently, and for each auction, there are 10 participants. If the reservation price of the agent is reached or until the end time for the all auction is reached, we can get the bidding price, the bidding time, and the quality of the target item for each agent. We apply the fuzzy decision making approach to compute the fitness value of agent. Then we apply the evolution approach to generate the next generation agents. In this particular experiment, the mutation rate is 0.02. The simulations stop when the population is stable (95% of the individuals have the same fitness) or the number of iterations is bigger than a predetermined maximum (200 in our case). The results of the simulation is shown as the following figure (Figure 3).

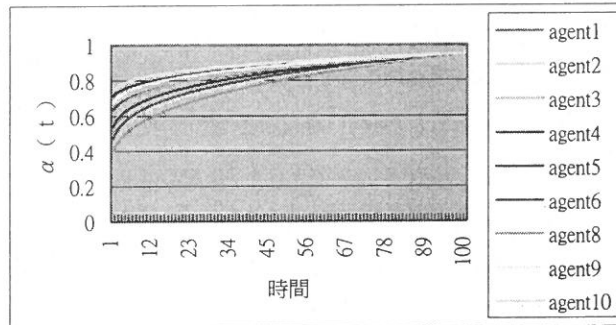


Figure 3: The results of simulation

5. Conclusion

In this paper, we present a new approach for evolving intelligent agents in e-commerce. A goal-driven approach can construct the user's soft and rigid goals based on fuzzy set theory. The proposed BDI model represents the mental skills of the intelligent agent, including belief, desire, intension, and strategy. A FMCDM approach is applied to evaluate the agent's strategy. Agent fitness and life cycle are proposed to facilitate and control the process of agent evolution. We construct multi-agent evolution cycle, which includes states of restructuring, selection, and growing. We conduct a series of experiments to determine the most successful strategies and to see how and when these strategies evolve depending on the context and negotiation stance of the agent's opponent. Finally, we have demonstrated the usefulness of agents employing a cocktail of tactics – both for different negotiation issues and, in combination, for a single issue.

6. Acknowledgements

The author wish to express his thanks for the financial support of the Societas Verrbi Divini.

References

- [1] R.H. Guttman, and P. Maes. Agent-mediated negotiation for retail electronic

- commerce. In *Agent-mediated Electronic Commerce: First International Workshop on Agent Mediated Electronic Trading*, Berlin: Springer, pp. 70-90, 1999.
- [2] T.H. Wang, S.U. Guan, and S.H. Ong. An agent-based auction service for electronic commerce. In *Proceedings of International ICSC Symposium on Interactive and Collaborative Computing*, Australia, 2000.
- [3] D. C-H Han and N. Parameswaran. Multi Agent Problem Solving with Mental States. IEEE. pp. 288-292. 1994.
- [4] P. Cohen and H.J. Levesque. Intention is Choice with Commitment. *Artificial Intelligence*, 42(3): pp. 213-216, 1990.
- [5] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.
- [6] A. Rao and M. Georgeff. Modeling rational agents within a BDI architecture. J. Allen, R. Fikes, and E. Sandewall (editors). KR-91. Morgan Kaufmann, 1991.
- [7] M. Barbuceanu and M.S. Fox. COOL: A language for describing coordination in multi-agent systems. In *Proceedings of the International Conference on Multi-Agent Systems, ICMAS-95*, San Francisco, CA, 1995.
- [8] J. Lee and J.Y. Kuo. New approach to requirements trade-off analysis for complex systems. *IEEE Transactions on Knowledge and Data Engineering*. 10 (4): 551-562. 1998.
- [9] J. Lee, N.L. Xue and J.Y. Kuo. February 2001. Structuring Requirement Specifications with Goals. *Information and Software Technology*, Vol. 43, pp. 121-135.
- [10] J. Lee, J. Y. Kuo and A. Liu. Modeling User Requirements with a Front-end Analysis. *International Journal on Artificial Intelligence Tools*. 6(3): 313-324, Sep. 1997.
- [11] L.A. Zadeh. Test-score Semantics as a Basis for a Computational Approach to the Representation of Meaning. *Literacy Linguistic Computing*, 1:24, 1986.
- [12] P. Anthony, W. Hall, V. Dung, N.R. Jennings. Autonomous for Participating in Multiple Online Auctions. *Proceedings IJCAI Workshop on E-Business and the*

- Intelligent Web, Seattle, WA, pp. 54-64, 2001.
- [13] N. Matos, C. Sierra and N. R. Jennings. Determining successful negotiation strategies: an evolutionary approach. Proceedings 3rd Int. Conf. on Multi-Agent Systems (ICMAS-98), Paris, France, pp. 182-189, 1998.
 - [14] J. Backer, "Adaptive Selection Methods for Genetic Algorithms", Proceedings ISCGA, pp. 100-111. 1994.
 - [15] M. Michalewicz, Genetic Algorithms + Data Structure = Evolution Program, Springer-Verlag, Berlin, 1992.
 - [16] C. Carlsson and R. Fuller. Interdependence in fuzzy multiple objective programming. Fuzzy Sets and System, 65:19-29, 1994.
 - [17] J. Lee and J.Y. Kuo. Fuzzy decision making through trade-off analysis between criteria. Information Sciences, 107(1-4): 107-126. 1998.
 - [18] D. Goldberg. Genetic Algorithms in Search Optimization and Matching Learning. Addison-Wesley. 1989.
 - [19] Baker, J.B., "Adaptive Selection Methods for Genetic Algorithms", Proceedings of the First International Conference on Genetic Algorithms, pp.101~111, 1985.
 - [20] D. Goldberg. Genetic Algorithms in Search. Optimization and Matching Learning. Addison- Wesley. 1989.
 - [21] Chen, T.Y. and Chen, C.J., "Improvements of Simple Genetic Algorithm in Structural Design", International Journal for Numerical Methods in Engineering, Vol.40, pp.1323-1334, 1997.
 - [22] Z. Michalewicz, T. Logan, and S. Swaminath. Evolutionary Operators for Continuous Convex Parameter Spaces", In Proceedings of the 3 rd annual Conference on Evolutionary Programming. World Scientific, pp. 84-97, 1994.
 - [23] R. P. McAfee and J. McMillan. Bidding rings. American Economic Review, Vol. 82 No. 3. pp. 579-599. 1992.

received September 22, 2002
revised October 27, 2002
accepted November 14, 2002

應用智慧型代理人漸進式拍賣策略於多重拍賣網站

郭忠義*

輔仁大學資訊工程系

摘 要

隨著電子商務和網際網路的蓬勃發展，虛擬企業系統的應用越來越受到學術研究與工商業界的重視。當前虛擬企業系統中，軟體代理人漫遊在網際網路上，與其他代理人競爭有限資源以達成本身目標，有些代理人較易達成目標，有些無法完成任務，新手代理人須經過不斷努力增加策略行為以適應環境。在代理人創造之初，只有少許能力、知識與經驗，若代理人有能力可以學習與演進其知識與能力將有很大的助益。本論文提出智慧型代理人演進法，以 BDI 模型描述代理人智慧模組，包括目標與計畫。運用目標導向使用案例分析法擷取並分析使用者賦予的任務，利用模糊邏輯模型化不精確之代理人目標，使用基因演算法與模糊多準則決策發展代理人演進方法。

關鍵詞：代理人、模糊邏輯、基因演算法。

Low Power Programming Design for Recharge Cycle Extension of a Mobile Computer

Ying-Wen Bai* and Chou-Su Chow

Department of Electronic Engineering

Fu Jen Catholic University

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

Abstract

Reducing power consumption is important for mobile computer design in order to reduce the power exhaustion probability by extending the recharge cycle. Thus, in this paper we first propose a queueing model to represent the power exhaustion probability of a mobile computer that is dependent on power consumption and power support. Second, to reduce power exhaustion probability, we propose a low power programming design in order to reduce power consumption. By utilizing assembly programming, we can reduce the execution time and thus provide the potential of reducing power consumption and extending the recharge cycle of a mobile computer. In our experiment, there are five testing programs with different functions for each version in the module of a mixed mode consisting both of assembly and C language, and of a pure C language design. In addition, by using maximum speed options of two compilers for comparison testing, the execution times for each version of the five testing programs are evaluated. Overall, the experimental results show that we have obtained a reduction in execution time, so the power consumption is reduced from one to two and half times.

Key Words: Mobile Computer, Low Power Programming, Mixed Mode, Improvement Proportion.

*Corresponding author. Tel.: +886-2-29031111 ext. 3792; fax: +886-2-29042638
E-mail address: bai@ee.fju.edu.tw

1.Introduction

Although the power consumption of computer operation depends on both the hardware used and the software design, most researchers have investigated computer power consumption from only a hardware point of view, resulting in substantial improvements [1-3]. On the other hand, considering software, power management methods and low power programming can also provide another important factor to reduce power consumption by reducing the execution time [5-10]. Because a low power programming technique can be used with different hardware platforms, thus the software methods used are also important for portable information applications, especially considering the recharge cycle of their very limited battery power sources. Thus, by utilizing a more efficient programming language, such as assembly language, in a mobile computer will consume less power and extend the battery recharge cycle [4].

The user operation characteristics representing the specific service request process and the service departure process in a mobile computer can be very random. Therefore, in order to design a mobile computer with a long recharge cycle, we must consider an average performance analysis of the power exhaustion probability [4]. Hence, we can use a single equivalent M/M/1 queueing model because the Poisson arrival process and the exponential service time distribution result in very fair assumptions for the operating characteristics of a typical mobile computer. The defining assumptions for a Poisson arrival process are as follows. First, the probability of one particular event arrival in a time slot is independent of arrivals in adjacent (past or future) time slots. Second, the interval of a time slot is small enough that the number of the event arrivals will be one or zero in those time slots [5-6].

In order to reduce the power consumption and power exhaustion probability, we investigate possible assembly language techniques and count the execution machine cycles of a mobile computer powered by a rechargeable battery as the experimental platform. There are five different testing programs used, each of which emphasizes different software features. These programs include: data I/O, memory access, and

computation. For each testing program, we design two different versions, which consist of either a mixed mode of assembly and C language, or a pure C language. In addition, by using the maximum speed options of two compilers, “A” and “B”, we record the CPU time of the five testing programs to compare each version [7-10].

The rest of this paper is organized as follows. In Section 2, we discuss the queueing model with respect to the power charging and discharging behavior of a rechargeable battery. In Section 3, we present Amdahl’s Law as the comparison index for the improvement proportion by using the testing programs of a mixed mode of assembly and C language, and of a pure C language. In Section 4, the detailed characteristics of five testing programs, their functions and their flowcharts are presented. Section 5 shows the experimental results of the execution time in comparison with testing programs, including a mixed mode of assembly, a pure C language and the maximum speed options of two compilers. Finally, in Section 6, we provide some conclusions drawn from our research.

2.Queueing Model for Charging and Discharging of a Rechargeable Battery in a Mobile Computer

Because of the individual differences in the operational behavior of various users, it can be very difficult to give a precise analysis of the charging and discharging process of a rechargeable battery. However, an approximate analysis can be made, which can provide certain estimated characteristics of the event, the charge arrival, and the departure processes. For the approximate analysis of the charge arrival process and the departure process, we use a single equivalent M/M/1 queueing model, as shown in Fig. 1. The event occupancy is fulfilled by the program execution that is in proportion to the usage of energy. And the event occupancy can be represented in proportion to the discharging rate in the rechargeable battery. Therefore, the equivalent model is similar to a single queue with a single-server, a charge buffer, an average charge arrival rate of λ , and an average charge departure rate of μ . Utilization of a low power programming technique is a method to both

reduce the execution time and also reduce the average charge departure rate. Overall, by utilizing this technique, the power consumption and the power exhaustion probability can be reduced and hence, the recharge cycle can be extended.

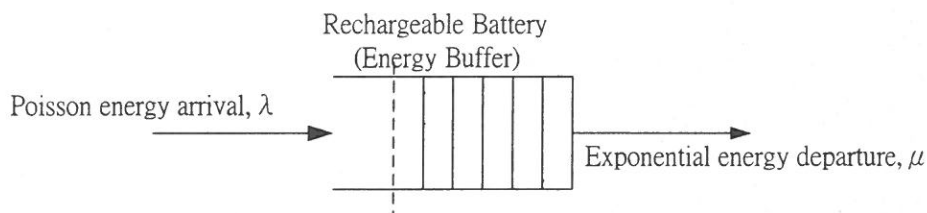


Fig. 1 Equivalent M/M/1 queueing model for a dual rechargeable battery

As noted in reference [4], regarding the static properties of the M/M/1 queue, both the average rechargeable battery energy occupancy and the probability of using-up the energy of the rechargeable battery are readily determined once we find the probabilities of state P_n of the rechargeable battery. Here, P_n represents the probability that there are “n” units of charges in the battery and P_0 represents the probability that there are zero units of charge in the battery. In other words, P_0 is the probability of using up the energy in a rechargeable battery. If P_0 is small enough and is based on a certain operating environment, then the power supply of the rechargeable battery powered mobile computer system can be acceptable. Specifically, we assume that the charge arrival process to the rechargeable battery

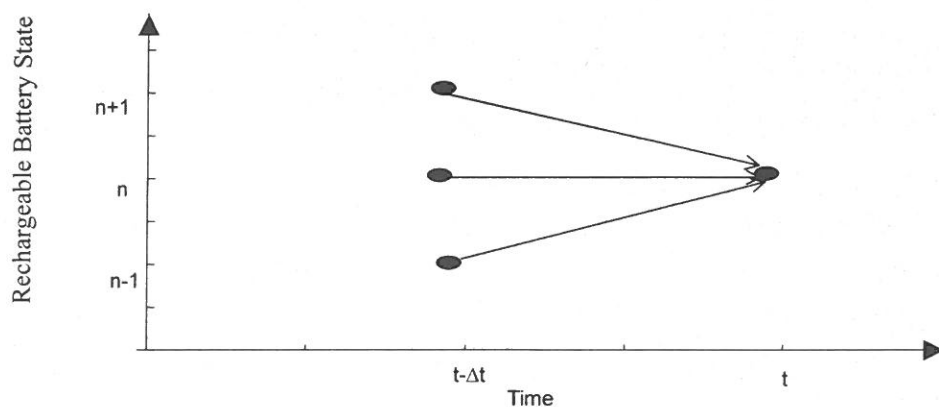


Fig. 2 M/M/1 state-time diagram for the model of a rechargeable battery

has a Poisson distribution, with a charge rate λ . If we assume the charge departure rate is exponential, with a discharge rate μ , then the probability $P_n(t + \Delta t)$ means that there are n units of charges in the rechargeable battery at time “ $t + \Delta t$ ”. Referring to the state-time diagram shown in Fig. 2, if we assume that the rechargeable battery is in state “ n ” at time “ $t + \Delta t$ ”, then it can only have been in states “ $n-1$ ”, “ n ”, or “ $n+1$ ” at time “ t ”.

The charging and discharging process of a rechargeable battery can be represented by a typical queue. Therefore, the probability $P_n(t + \Delta t)$ of the queue in state “ n ” at time “ $t + \Delta t$ ” can be the sum of the state probability in state “ $n-1$ ”, “ n ”, or “ $n+1$ ” at time “ t ”, each multiplied by the probability of charges arriving at state “ n ” in the intervening “ Δt ” units of time. Thus we can generate the following equation for $P_n(t + \Delta t)$,

$$\begin{aligned} P_n(t + \Delta t) = & P_n(t) [(1 - \lambda\Delta t)(1 - \mu\Delta t) + \mu\Delta t\lambda\Delta t + O(\Delta t)] \\ & + P_{n-1}(t) [\lambda\Delta t(1 - \mu\Delta t) + O(\Delta t)] \\ & + P_{n+1}(t) [\lambda\Delta t(1 - \mu\Delta t) + O(\Delta t)] \end{aligned} \quad (1)$$

Since $O(\Delta t)$ includes terms of orders $(\Delta t)^2$ and higher, the terms involving $(\Delta t)^2$ in Eq. (1) should be incorporated in $O(\Delta t)$. Simplifying Eq. (1) by dropping $O(\Delta t)$ terms and higher order terms, one can obtain $P_n(t)$, as shown in Eq. (2), by solving differential-difference equations.

$$P_n(t) = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n \left[1 - \sum_{k=0}^{\infty} \left(\frac{(\lambda t)^k e^{-\lambda t}}{k!} e^{-\mu t} \sum_{m=0}^{n+i+1+k} \left(\frac{\mu t}{m!}\right)^m\right)\right] \quad (2)$$

In taking the limit as $t \rightarrow \infty$, we see that the terms $e^{-(\lambda+\mu)tm}$ go to 0. Hence, when $t \rightarrow \infty$ the result is,

$$\lim_{t \rightarrow \infty} P_n(t) = P_n = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n \quad (3)$$

When $\lambda/\mu < 1$ we get a valid steady-state probability distribution, and when $n=0$,

we obtain the power exhaustion probability P_0 .

$$P_0 = 1 - \frac{\lambda}{\mu}, \quad 0 \leq \frac{\lambda}{\mu} < 1 \quad (4)$$

where λ is the charging rate from the power generation of the sources and μ is discharging rate for the power consumption of a mobile computer system. Because typical users do not use a mobile computer continuously, its utilization should be less than 100%. Hence, we can consider the utilization (u.t.) and then obtain an approximate equivalent power consumption rate of,

$$\hat{\mu} = (u.t.)\mu \quad (5)$$

If we rewrite Eq. (4) with the equivalent power consumption rate $\hat{\mu}$ and power generation rate λ , we obtain

$$P_0 = 1 - \frac{\lambda}{(u.t.)\mu} \quad (6)$$

Eq. (6) shows the relationship between the power exhaustion probability and the utilization (u.t.) of a mobile computer system, demonstrating the smaller utilization and discharging rate (μ) with respect to the smaller power exhaustion probability for a given set of parameters.

3. Amdahl's Law for Performance Comparison

Amdahl's Law provides a quick way to find the overall speedup from an enhancement that depends on the following two factors:

1. The fraction of the computation time in the mobile computer that can be converted to take advantage of the enhancement.
2. The improvement gained by the enhanced execution mode; that is, how much faster the task would run if the enhanced mode were used for the entire program.

The execution time using the original machine with the enhanced mode will be the time spent using the unenhanced portion of the machine plus the time spent using the enhancement [3].

We can transfer the overall “Speedup” of Amdahl’s Law from the primary equation as an overall “Improvement Proportion” which can be one of the following: execution time and power consumption:

$$Im\ provement_{overall} = \frac{1}{(1 - Pr\ oportion_{enhanced}) + \frac{Pr\ oportion_{enhanced}}{Im\ provement_{enhanced}}} \quad (7)$$

Then

Improvement_{overall}: The overall improvement of the program performance

Pr oportion_{enhanced}: The proportion of the program including assembly

Im provement_{enhanced}: The improvement proportion of the efficiency between assembly and C language.

In Section 4, we obtain the overall improvement proportion by using Eq. (7) of Amdahl’s Law.

4. Five Testing Programs and Their Flowcharts

In general, the program utilizing assembly language has both a faster execution time and a lower program code size providing a potential for reducing power consumption by means of the higher controllability of computer hardware using the natural characteristics of assembly language. However, the main disadvantages of programs utilizing assembly language is that they are difficult to design, implement and maintain. On the other hand, the C version of the program has the characteristics of a typical high-level programming language. In other words, C programming language is easily use to design, implement and maintain, but has the disadvantage of a longer execution time with higher power consumption. Therefore, if we can integrate the characteristics of the high-level and the assembly program languages, the mixed mode of the testing programs will provide a scalable ability in execution

time and power consumption.

To demonstrate the different characteristics of the testing programs, we use mobile computer systems powered by rechargeable battery as the experimental platform of the five testing programs. For each of the five testing programs we use two versions: a mixed module of C and assembly and a pure C module respectively. In addition, both versions are compiled by compiler A and B, which selects whether to use a maximum speed option or not. Therefore, for each program, we can obtain eight classes of execution times. By inspecting the experimental data, we can therefore investigate the execution time of low power programming.

In these five testing programs, program 1 emphasizes input and output data; program 2 emphasizes computing data before input/output; program 3 emphasizes retrieving data from a table, and the resulting input or output data; programs 4 and 5, instead of emphasizing input and output, emphasize “computation mathematics”. We use different testing program structures for the same measurement of the software characteristics, and then we explore the proportion of improvement due to programming style.

These five testing programs and their software flowcharts are shown as follows.

3.1 Testing Program 1 (The Display Digit Shifts Left and Right)

Testing program 1 contains a function that can control the digit display by shifting digits from left to right. When this program is executed, as it outputs data to an I/O port of the computer system, it outputs the digit from the I/O port and then goes back to form loops. Fig. 3 shows the software flowchart of testing program 1.

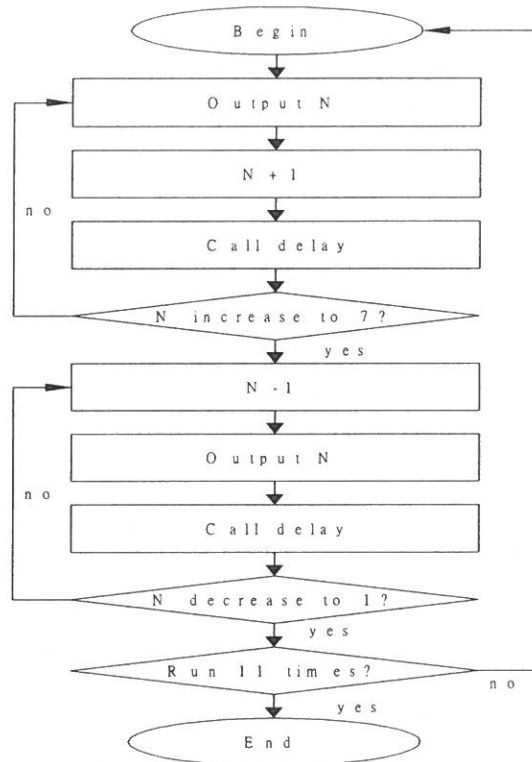


Fig.3 Flowchart of testing program 1

In Fig. 3, the subprogram “DELAY” generates a delay time between the “on” and “off” states. Fig. 4 shows the assembly code of part of testing program 1 with a mixed module of C language and assembly language.

```

asm{
    MOV EAX, OH
    MOV ECX, aa
STAR: ADD EAX, 01H
    CMP EAX, ECX
    JNE STAR
}
  
```

Fig. 4 Assembly codes of part of testing program 1 with a mixed design of C and assembly (delay function)

Figure 5 is the source code of testing program 1 when using C language. We use the statement “for” to control the number of times that the digit “c” is shifted either to the left or to the right.

```

void delay( double  n){
    double  a;
    for(a=0.0 ; a < n; a++){;}
}
main()
{
    int c =  1;
    double  i, g=0.0,b,e,f;
    b=clock ();
    printf("%d\n", b);
    while(g<= 10) {
        for(i = 0;i<7;i++) {
            printf("%d\t", c);
            delay(0x1989680);
            c = c+1;
        }
        printf("\n");
        for(i = 0; i<7 ; i++){
            printf("%d\t", c-1);
            delay(0x1989680);
            c=c-1;
        }
        printf("\n");
        g++;
    }
    e=clock ();
    f=(e-b)/CLK_TCK;

```

```

printf("\n%f\n%f",e, f);
return 0;
}

```

Fig. 5 Source codes of testing program 1 using C language

3.2 Testing Program 2 (Timer)

Testing program 2 uses an I/O port computer system to control a 7-segment digit display for 0~9 of decimals. The program will show numbers between 00 and 99 like a timer, and then repeat the operation by looping. Fig. 6 shows the software flowchart.

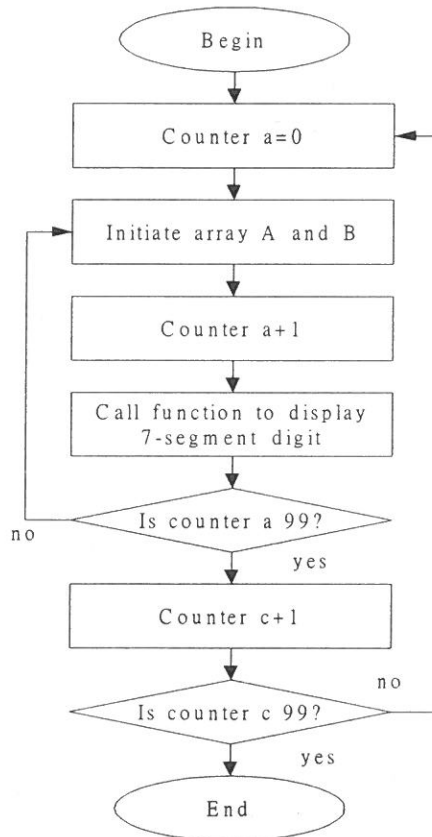


Fig. 6 Flowchart of testing program 2

Figure 7 shows the portion of source codes rewritten by assembly code. In

Figure 7, the machine cycles of C source codes require 98 cycles; on the other hand, the machine cycles of assembly codes require only 45 cycles. This is the major reason why we can reduce the execution time by using a mixed module of C language and assembly language.

a1 = a%10;	→	_asm{
c1 = a/10;	Become	MOV EAX, a
		CDQ
		DIV X
		MOV c1, EAX
		MOV a1, EDX
		}

Fig. 7 Assembly portion of the source codes of testing program 2, utilizing a mixed of C language and assembly language

3.3 Testing Program 3 (5×7 Dot Matrix's Digit)

The rows of the dot matrix digits are searched in a table by the computer program. The mobile computer outputs the data by means of an I/O port. The program outputs and displays symbols from 0 to 9 and A to H. Figure 8 shows the software flowchart of testing program 3.

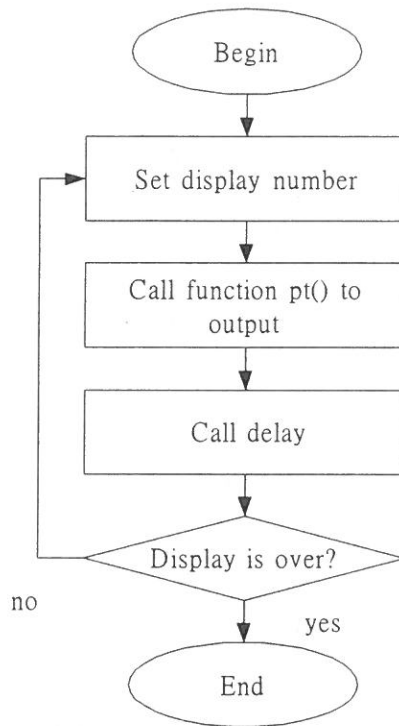


Fig. 8 Software flowchart of testing program 3

3.4 Testing Program 4 (Prime Numbers between 1 and 500)

In order to understand the different characteristics of programming language, one must remember that testing programs 4 and 5 are only for computing data, and they lack input/output. However, we want to know the effects on computation performance in the testing program, and Fig. 9 shows the flowchart of testing program 4.

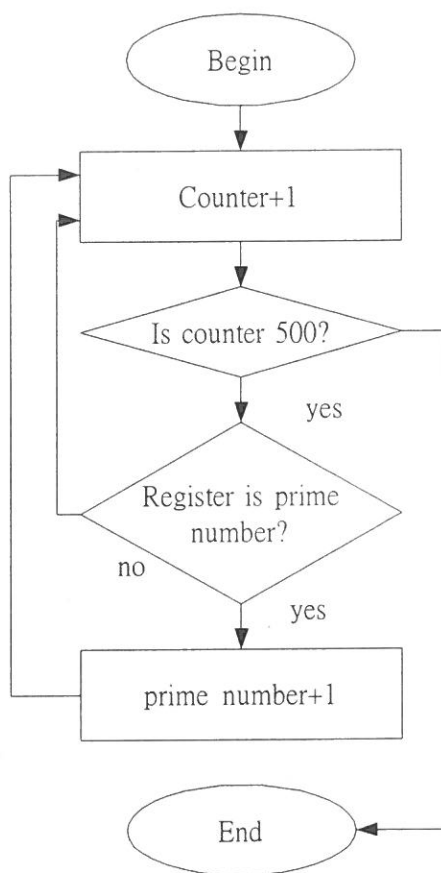


Fig. 9 Flowchart of testing program 4

	→	_asm{
	Become	MOV EDX, i
for (j = 2; j < I; j++) {		MOV AX, DX
x = 1%j;		MOV CL, 08H
if(x == 0)		SHR EDX, CL
n++;		SHR EDX, CL
}		DIV j
		CMP DX, 0H
		JNE TT

```
INC n  
TT  
}
```

Fig. 10 Assembly portion of testing program 4 in a mixed design

In C language, we use a “division” statement to find prime numbers. The register variable “x” saves remainders from dividing into two numbers, and the statement “if” checks whether the remainder is equal to zero or not. In Fig. 10, the number of machine cycles needed by the “for” loop is 61; on the other hand, the number of machine cycles needed for the equivalent assembly codes by the “for” loop is 54. This reduction again can reduce both the execution time and the power consumption of mobile computers.

3.5 Testing Program 5 (Multiply Two 100*100 Matrices)

Testing program 5 only emphasizes data computation, without utilizing input/output. “SEG1” and “SEG2” save two 100*100 matrices separately. In assembly language, we use the instruction “MOV” to obtain the data from the two matrices separately, and we save them to register variables “A” and “B”, respectively. Then to obtain the answer, we multiply two variables in the registers. In the C language design, we obtain the answer that uses the statements “for”.

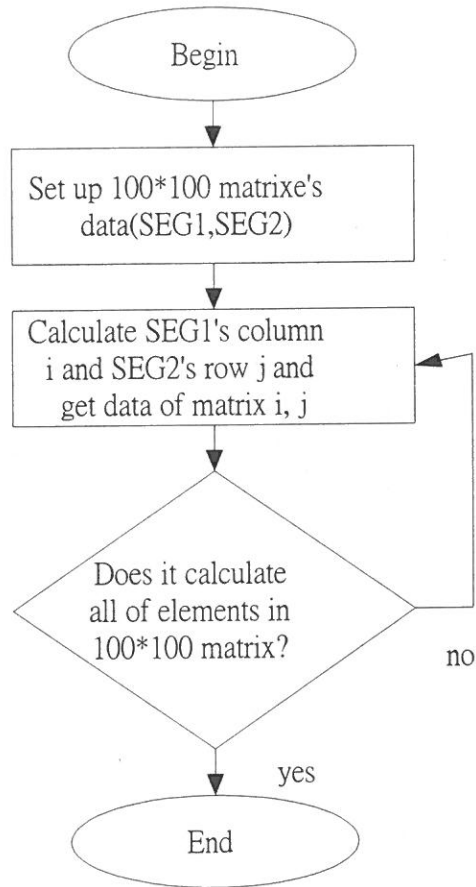


Fig. 11 Flowchart of testing program 5

$$c = (c + A[i][a] * B[a][j]);$$

→
Become

```

b = A[i][a];
d = B[a][j];
_asm{
    MOV EAX, b
    IMUL EAX, d
    ADD c, EAX
}
  
```

Fig. 12 Assembly part of testing program 5 in a mixed design

In the mixed mode version, we again use assembly code to replace the part of the testing program in the “for” structure statement that, by looping, computes each element of the matrices. In Fig. 12, the number of machine cycles is 55 for the C codes loop; on the other hand, the number of machine cycles is 53 for the equivalent assembly codes used in the matrix operation. This reduction is very limited, so the reduction of the execution time is also limited, and we can save only a little power consumption.

5. Experimental Data of Five Testing Programs

The five testing programs have been designed for the different functions, as described and shown in the above section. Furthermore, since the five testing programs are implemented by utilizing different versions of various programming languages, the execution times can be different. Therefore, in this section, we will record the differences in the execution time and power consumption for the two versions of the five testing programs.

5.1 Execution Time

Table 1 and Fig. 13 show the relationship for the different execution times of the three versions of five testing programs.

Table 1 Execution time of five testing programs

“A” compiler	Program1	Program2	Program3	Program4	Program5
C language	193.511sec.	15.801sec	8.947sec.	2.181sec.	9.723sec.
C language (max speed)	31.338sec.	16.049sec.	3.922sec.	0.439sec.	9.887sec.
Mixed mode	18.158sec.	15.566sec.	2.196sec.	0.428sec.	9.281sec.
Mixed mode (max speed)	18.197sec.	15.703sec.	2.209sec.	0.44sec.	9.349sec.
“B” compiler					

C language	187.031sec.	17.682sec.	22.195sec.	2.416sec.	14.236sec.
C language (max speed)	184.702sec.	18.105sec.	11.473sec.	1.306sec.	13.780sec.
Mixed mode	19.311sec.	16.957sec.	2.245sec.	1.169sec.	13.247sec.
Mixed mode (max speed)	180.520sec.	17.308sec.	2.239sec.	1.237sec.	12.907sec.

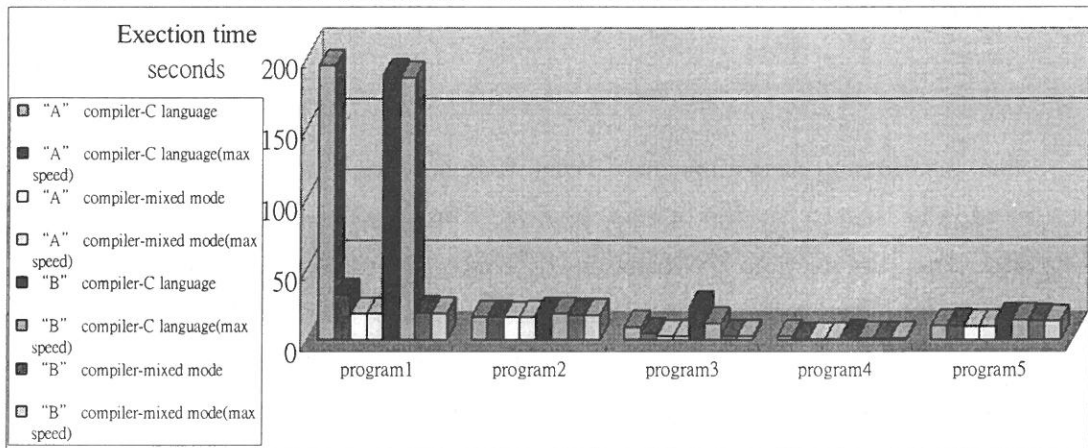


Figure 13 Execution time of five testing programs

Clearly, by examining the relationships between the various execution times, we can obtain the improvement proportion for the execution time by using Eq. (7) in Section 3, which can be shown as follows.

For the testing program 1, we have

$$Pr oportion_{enhanced} = (193.511 - 18.158) / 193.511 = 0.906165541$$

$$Im provement_{enhanced} = 193.511 / 18.158 = 10.65706576$$

The overall improvement proportion of testing program 1 can be shown by Eq. (8)

$$Im provement_{overall} = \frac{1}{(1 - Proportion_{enhanced}) + \frac{Proportion_{enhanced}}{Im provement_{enhanced}}} = 5.587 \quad (8)$$

The overall improvement proportion of the testing program 2~5 can be obtained using the same procedure as shown Eq. (8).

The average improvement proportion of the five testing programs is:

$$(5.587+1.0003+2.326+2.825+1.002)/5 = 2.5 \quad (9)$$

From the above comparison, we can see that the execution time using assembly language is shorter than by using C language. In other words, the average power consumption will be less by using assembly language.

In our experiment, the difference between assembly and C language execution time can reach 1 ~ 2.7 times. This is a very important factor for the reduction of power consumption of portable information systems. If we can reduce the power consumption of a portable information system, then the recharge cycle can be extended and the convenience of usage will also be increased. Table 2 shows, from an execution time point of view, the average proportion of improvement.

Table 2 Relationship of average proportion of improvement of five testing programs

	Improvement ratio of mixed mode design by using compiler "A" without optimum option	Improvement ratio of pure C design by using compiler "A" with optimum option	Improvement ratio of mixed mode design by using compiler "B" without optimum option	Improvement ratio of pure C design by using compiler "B" with optimum option
Program 1	5.587	1.214	5.104	5.249
Program 2	2.326	1.235	5.2069	2.838
Program 3	2.825	1.004	1.3627	1.003
Program 4	1.0003	1.0016	1.0017	1.0461
Program 5	1.002	1.003	1.004	1.004
Average Improvement Ratio	2.54806	1.09152	2.73586	2.22802

From Table 2, we determined that an average proportion of improvement of five testing programs was an energy improvement of 1.7 times. Especially, in the mixed mode design, we can reduce the discharging rate (μ) by about 2.5 times that obtained from the average of the improvement ratios in Table 2 by using compilers “A” and “B” without a maximum speed option. Therefore, we rewrite equation (6) as

$$P_0 = 1 - \frac{\lambda}{\hat{\mu}/2.5} = 1 - \frac{2.5 \cdot \lambda}{(u.t.)\mu} \quad (10)$$

Eq. (10) shows the initial relationship and the improved relationship with low power programming for ratio $\lambda / \hat{\mu}$, utilization u.t., and operation time. The recharge cycle, or operation time depends on u.t., λ and μ . The discharge rate, μ , is reduced 2.5 times, therefore, we can extend the recharge cycle of a mobile computer by 2.5 times. In other words, the operation time can be extended to 7.5 hours, as compared to 3.0 hours without using any low power programming technique.

6. Conclusions

For the study of the low power programming design of the mobile computer system powered by a rechargeable battery, we use C language, and a mixed mode of C language and assembly language. In our experiment, we use the maximum speed options of two compilers to investigate five testing programs. The experiment results show that there is a potential energy improvement of 1.7 times as a result of the total effect of all versions from Table 2. Finally, if the average proportion of improvement makes use of low power programming without the maximum speed option of compiler “A” and “B”, the operation time can be extended to 7.5 hours from 3.0 hours by using a mixed mode design. From the results of the above experiment, it can be seen that the low power programming design can be a viable scalable method based on software technology to extend the recharge cycle.

In the future, our research results can be incorporated into a low-power compiler that can automatically generate low-power machine code to handle mobile information applications such as “AutoCad”, “MatLab” and “Words” with less

power consumption. Before this compiler has been designed, we can trace high-level source program and locate some often used modules, and then rewrite them by assembly language to reduce the execution time. Therefore, we can obtain scalability for the reduction the power consumption of mobile computers based on the use of the portion of assembly code in the application software.

References

- [1] E. P. Harris, S. W. Depp, W. E. Pence, S. Kirkpatrick, M. Sri-Jayantha and R. R. Trountman, "Technology Directions for Portable Computers," Proceedings of the *IEEE* Vol. 83, No. 4, pp. 636-658, April 1995.
- [2] Z. J. Limnios and K. L. Gabriel, "Low-Power Electronics," *IEEE Design & Test of Computer*, Vol. 11, No. 4, pp. 8-13, Winter 1994.
- [3] John L. Hennessy and David A. Patterson, *Computer Architecture a Quantitative Approach*, Second Edition, Morgan Kaufmann Publishers, Inc, CA, USA.
- [4] Ying-Wen Bai and Cheng-Lun Chang, "Feasibility Analysis of a Queueing Model for a Dual Rechargeable Batteries in Solar Cells Powered Mobile Computers," Proceedings of IASTED International Conference Applied Modelling and Simulation, P. P. 136-142, May. 2000
- [5] Rulnick, John, M. and Bambos, Nicholas, "Mobile Power Management for Maximun Battery Life in Wireless Communication Network," Proceedings-*IEEE INFOVOM V2*, pp. 443-450, 1996.
- [6] Luca Benini and Giovanni De Micheli, "System-Level Power Optimization: Techniques and Tools," Proceedings of the International Symposium on Low Power Electronics and Design, pp. 288 -293, 1999.
- [7] V. Tiwari, S. Malik, A. Wolfe, "Power Analysis of Embedded Software: A First Step Towards Software Power Minimization," *IEEE Transactions on VLSI*, vol. 2, no. 4, pp. 437-445, Dec. 1994.
- [8] V. Tiwari, S. Malik, A. Wolfe, "Instruction Level Power Analysis and Optimization of Software," *Journal of VLSI Signal Processing*, vol. 13, no. 1-2,

pp. 223-233, 1996.

- [9] H. Mehta, R. Owens, M. Irwin, R. Chen, D. Ghosh, "Techniques for Low Energy Software," International Symposium on Low Power Electronics and Design, pp. 72-75, Aug 1997.
- [10] C. Gebotys and R. Gebotys, "An Empirical Comparison of Algorithmic, Instruction, and Architectural Power Prediction Model for High-Performance Embedded DSP Processors," International Symposium on Low Power Electronics and Design, pp. 121-123, Aug. 1998.

received September 29, 2002
revised October 12, 2002
accepted November 22, 2002

利用低功率程式設計延長行動電腦再充電周期

白英文* 邱國書

輔仁大學電子工程系

摘 要

降低行動電腦功率消耗，延長再充電周期在可攜式資訊應用日益重要，因此本文依據先前排隊理論模型，界定缺電機率與電腦功率消耗和供給所形成的相依關係，再者，爲了降低缺電機率和延長再充電周期，我們提出低功率程式設計用以降低行動電腦功率消耗，經由使用組合語言，降低程式執行機械週期個數，加速完成程式執行，降低平均功率消耗，達到降低功率消耗的效果。本論文採用五個不同功能之測試程式，分別以組語、組語與 C 語言混合、純 C 語言設計對應版本，再以兩種編譯器編譯五個測試程式之三種版本，比較其執行時間，平均而言，有潛力達成縮短 2.5 倍。透過狀態控制將行動電腦轉移至低功率狀態，達到節電效果，也就是有潛力達成節電 2.5 倍的效果。

關鍵詞：行動電腦，低功率程式設計，混合模式，改善比率

A Practical Receipt-Free e-Voting Scheme Based on Smart Cards

Wei-Chi Ku* and Chun-Ming Ho

Department of Computer Science and Information Engineering

Fu Jen Catholic University

Taipei, Taiwan 242, R.O.C.

Abstract

Bribe and coercion are common in traditional voting systems and usually lead to a biased result that imparts the desired democracy. However, these problems become harder to solve in e-voting schemes. Unfortunately, existing e-voting schemes can not effectively provide receipt-freeness and uncoercibility. This paper presents a secure e-voting scheme that can be realized with current techniques. By using simple physical voting booths, bribers and coercers can not monitor the voter while he votes. By using smart cards to randomize part of content of the vote, the voter can not construct a receipt. Since the added randomness can not be controlled and learned by the voter, receipt-freeness and uncoercibility are effectively achieved or at least lessened. Particularly, the proposed e-voting scheme is practical in that its assumptions are quite appropriate for realistic environments. Furthermore, its performance is optimal in that time and communication complexity for the voter is independent of the number of voting authorities.

Key Words: e-voting scheme, receipt-freeness, smart cards, voting booth

*Corresponding author. Tel.: +886-2-29031111 ext. 3894; fax: +886-2-29023550

E-mail address: csie2023@mails.fju.edu.tw

1. Introduction

Voting is regarded as one of the effective methods for individual to express their opinions on a given topic. Generally, not only the intentions of the voters but also the voting methods will affect the voting results. Traditional paper-based voting methods are inconvenient for voters and then will more or less affect the accuracy of the voting results. For example, voters living far from their domiciled homes may waive their voting rights. Hence, we are motivated to develop an e-voting scheme capable of providing more efficient voting services than the traditional paper-based voting methods. Inevitably, the development of a secure e-voting scheme is difficult since it also allows for the possibility of adversaries to affect or even disrupt voting in an easier way once it contains a security flaw.

The e-voting scheme is an example of secure multi-party computations. In 1982, Chaum [Cha81] pioneered the notion of e-voting, and several concrete schemes, e.g., [DLM82] and [Yao82], were subsequently proposed. However, these early e-voting schemes are unsuitable for being deployed in large-scale environments because a failure of a single voter would disrupt the entire voting. Later, many e-voting schemes, e.g., [CF85], [Cha88], and [FOO92], for large-scale environments have been proposed. Chaum [Cha88] described an e-voting scheme based on the sender untraceable email system, which assumes that at least one mix is trust. Based on multiple key ciphers, Boyd [Boy87, Boy90] proposed an e-voting scheme, which is limited because the voting authority can falsify the ballots. Since knowledge of the intermediate results could distort further voting, Fujioka *et al.* [FOO92] proposed an e-voting scheme capable of solving the fairness problem by using the bit-commitment function. No one, including the voting authority, can know the intermediate result of the voting. They also proposed another e-voting scheme [OFO93] based on a public bulletin board, which is realized by a committee of several members that can perform the same function as the mix in [Cha81]. Unfortunately, the security of their e-voting schemes relies on the cooperation of the voters. Cohen and Fisher [CF85] initially proposed an e-voting scheme based on the

homomorphism encryption technique, which can conceal the content of votes. Next, similar e-voting schemes have been proposed by Benaloh and Yung [BY86] and Sako and Kilian [SK95], respectively, with each one having its merits and limitations.

In traditional voting systems, a voting booth not only allows voters to keep their vote secret, but also prevents vote-selling and coercion. The notions of receipt-freeness and uncoercibility for e-voting were introduced by Benaloh and Tuinstra [BT94]. Receipt-freeness ensures that the voter is convinced that his vote is counted without getting a receipt. Uncoercibility ensures that the voter is not able to convince the coercer of the value of his vote. Although the concepts of uncoercibility and receipt-freeness are not the same, the term ‘receipt-freeness’ has been used in the literature as the widespread expression to represent both the receipt-freeness and uncoercibility for their similarity. In this paper, we also adopt this convention. Most e-voting schemes without using voting booths sacrifice receipt-freeness in order to establish correctness for the voting results, i.e., the voter can get a receipt to check his vote. In these schemes, the voter may be bribed or persecuted to vote for a certain candidate and is possibly obliged to vote under the supervision of the bribers or coercers. Preventing such threats in e-voting schemes has been the subject of recent research. So far, most receipt-free e-voting schemes [BT94, NR94, SK95] rely on some physical assumptions such as:

- An untappable channel from the voter to the voting authority: a channel that can guarantee the message sent from the voter to the authority is perfectly secret to all other parties.
- An untappable channel from the voting authority to the voter: a channel that can guarantee the message sent from the authority to the voter is perfectly secret to all other parties.
- A voting booth: a physical booth that the voter can cast his vote through a confidential channel between him and the voting authority.

However, the briber or the coercer can prescribe private random bits that the voter must use. Hence, these schemes do not effectively provide receipt-freeness. To

achieve actual receipt-freeness with current techniques, physical voting booths should be employed. Although it is inconvenient for the voter to cast his vote from a physically isolated voting booth in a dedicated network, however, there is no way to prevent the briber or the coercer from monitoring the voter while he votes by using a personal computer.

In summary, a secure e-voting scheme can be defined as in the following:

Definition 1. An e-voting scheme is secure if it satisfies:

- Completeness: all votes are counted correctly.
- Privacy: all votes must be secret.
- Unreusability: no voter can vote twice.
- Eligibility: only the eligible voters can vote.
- Fairness: no one can know the intermediate results of the voting.
- Verifiability: the result of the voting can be verified.
- Receipt-Freeness: the voter cannot reveal his vote to others.

Undoubtedly, a theoretically secure e-voting scheme is impractical if at least one of its assumptions is unreasonable. For example, schemes with multiple stages based on the assumption that no voter abstains from the voting in the intermediate stages are impractical. In this paper, we will describe a practical receipt-free e-voting scheme based on smart cards. The proposed e-voting scheme is an improved version of Cramer-Gennaro-Schoenmakers e-voting scheme [CGS97]. Physical voting booths are used to achieve receipt-freeness. To provide convenience to voters, sufficient voting facilities are supplied in sufficient public voting booths. In contrast to the traditional voting system, the voter can choose any voting booth convenient and safe to him to vote.

The rest of this paper is organized as follows. Section 2 briefly reviews related works on receipt-free e-voting schemes. Section 3 describes the proposed e-voting scheme. Next, Section 4 analyzes the security of the proposed scheme. Conclusions are finally made in Section 5.

2. Related Works

Benaloh and Tuinstra [BT94] initially introduced two receipt-free e-voting schemes using physical voting booths. To achieve universal verifiability, they employed a special bulletin board, which is like a broadcast channel with memory to the extent that any party can see the contents of it and each voter can post ballot by appending the ballot to his own designated entry. In particular, no party can erase anything from the bulletin board. The ballot does not reveal any information on the vote but it is ensured by an accompanying proof that the ballot contains a valid vote and nothing else. After that, many receipt-free e-voting schemes have been proposed. Hirt and Sako [HS00] showed that Benaloh-Tuinstra e-voting scheme is not receipt-free, and then proposed an efficient receipt-free e-voting scheme that assumes the existence of one-way untappable channels from the authority to the voter.

Cramer, Franklin, and Schoenmakers [CFSY96] described an e-voting scheme that can provide information-theoretic privacy by employing multiple voting authorities. The time and communication complexity for the individual voter is linear in the number of voting authorities. Later, Cramer, Gennaro, and Schoenmakers [CGS97] proposed another e-voting scheme, whose performance is claimed to be optimal in that time and communication complexity for the voter is independent of the number of voting authorities. The above two e-voting schemes also assume the untappable channel, and satisfy all requirements except for receipt-freeness. Recently, Magkos *et al.* [MBC01] proposed a receipt-free e-voting scheme based on the virtual voting booth, which is implemented with a smart card. In particular, untappable channels are not required in their scheme. Receipt-freeness is achieved by distributing the voting procedure between the voter and the smart card. The voter and the smart card jointly contribute randomness for the encryption of the vote. Within the virtual voting booth, the voter interactively communicates with his smart card. It is assumed that, during the very moment of voting, the briber or the coercer does not monitor the voter. However, it is an unreasonable assumption, i.e., the use of the virtual voting booth cannot effectively provide receipt-freeness in practical

environments.

3. The Proposed e-Voting Scheme

In this section, we will describe a practical receipt-free e-voting scheme based on smart cards. Most of the techniques used in our scheme are adopted from [CGS97] and [MBC01]. The model of the proposed scheme involves many voters and n voting authorities. A bulletin board on which each active participant can publish information is installed to store all ballots. All communications through the bulletin board is public and can be read by any party. No party can erase any information from the board, but each active participant can append messages to its own designated entry. It is assumed that no more than $t-1$ voting authorities conspire. Moreover, we employ physical voting booths that can perform voter authentication and generate random numbers, and are only protected by guards. However, we do not assume that the voting booth should guarantee the security of the communication between the voter and the voting authorities. The system architecture of the proposed e-voting scheme is shown in Fig. 1.

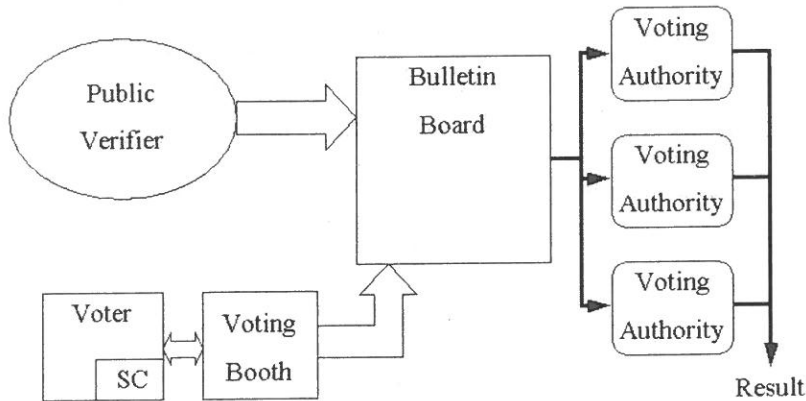


Fig. 1. System Architecture of the Proposed e-Voting Scheme

The proposed e-voting scheme relies on a homomorphic version of the ElGamal cryptosystem [ElG84]. Our construction works in subgroups G_q of order q of Z_p^* , where p and q are large primes such that $q|p-1$, and g and G are generators of G_q .

Given a message m , the encryption of m is the ElGamal encryption of G^m with base g , i.e., $(x, y) = (g^a, h^a G^m)$, where $h = g^s$ is the public key, s is the secret key, and a is a random number. All operations are modulo p . For clearness, we drop the operator $\text{mod } p$ throughout this paper. Due to the homomorphic properties of the used encryption method, a particular ElGamal encryption scheme, the final tally is verifiable to any observer. Privacy of individual ballot is guaranteed partly by the security of the ElGamal cryptosystem used to encrypt the vote. To join the system, the voter has to register first to the voting authority. After authentication, the voter will get a personal smart card, which contains the public encryption key of its owner, the public encryption key of the distributed voting authorities, and a secret signature key, with the corresponding certificate. It is assumed that the smart card is tamper-proof and can be activated only by using the correct password.

During voting, the voter posts a ballot accompanied with a proof that the ballot contains a valid vote, without revealing the actual content of the vote. Both the voting booth and the smart card contribute randomness to the ballot so that the voter cannot construct a receipt. The voter has to be convinced that the voting booth has done things correctly, but without finding out the randomness added by it. Before the ballot is submitted to the bulletin board, the voter must be given a proof of correctness of the encryption performed by the voting booth. The proof must be non-transferable; otherwise, it may be used as a receipt for this ballot. For this purpose, the verifier can be implemented by using either the beacon [Rab83, Ben87] as a trusted source of random bits or the Fiat-Shamir heuristic [FS86], which is being used in the proposed e-voting scheme. There should be an authenticated channel, which only the voter uses to submit the encrypted vote to the bulletin board. A digital signature scheme, e.g., RSA or DSA, is used to control access to the various entries on the bulletin board. The vote should be encrypted with the public key shared by the voting authorities to achieve vote secrecy, and the voter must not know the corresponding private key for decryption. A key generation protocol is used to generate the private key, denoted by s , jointly shared by all the n voting authorities. The secrecy of the votes is protected against coalitions of up to $t-1$ authorities.

Encrypted votes are accumulated when ballots are aggregated. A decryption protocol [Ped91, Ped92] is invoked by the voting authorities to jointly decrypt the tally of the accumulated votes without explicitly reconstructing s .

Let SC_i denote the smart card of voter i , VB a physical voting booth, and BB the bulleting board. The expression ' $Alice \rightarrow Bob: m$ ' represents that $Alice$ sends m to Bob . The protocol of the proposed e-voting scheme can be described as in the following.

Ballot Generation Phase (please refer to Fig. 2)

Ballot Generation Phase can be invoked at any time prior to the deadline of voting.

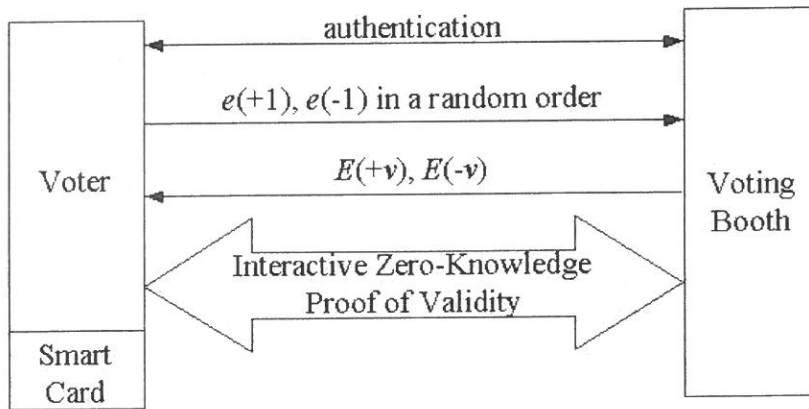


Fig. 2. Ballot Generation Phase

Step G1. Voter i :

- goes to a VB that is convenient and/or safe for him.
- authenticates to VB together with his SC_i that has been activated by the correct password.

Step G2. Voter i :

- uses SC_i to generate two random numbers r_1 and r_2 .
- uses SC_i to compute $e(+1) = (g^{r_1}, h^{r_1} G)$ and $e(-1) = (g^{r_2}, h^{r_2} / G)$.

- \rightarrow VB : $e(+v), e(-v)$. // v is randomly selected from $\{1, -1\}$ //

Step G3. VB:

- generates two random numbers R_1, R_2 .
- computes

$$E(+v) = (e_1(+v)g^{R_1}, e_2(+v)h^{R_1})$$

$$E(-v) = (e_1(-v)g^{R_2}, e_2(-v)h^{R_2}),$$

where $e(x) = (e_1(x), e_2(x))$.

- generates a random number W and computes $(P_a, P_b) = (g^W, h^W)$.
- generates a random number W_{bt} and computes $(x_{bt}, y_{bt}) = (g^{W_{bt}}, h^{W_{bt}})$.
- \rightarrow Voter i : $E(+v), E(-v), (P_a, P_b), (x_{bt}, y_{bt})$.

Then, Steps G4 ~ G6 are executed K times. The value of K depends on how Voter i can be convinced that VB has done things for him correctly in Step G3.

Step G4. Voter $i \rightarrow$ VB: c . // c is a random challenge //

Step G5. VB:

- computes $res_1 = W + R_1c$ and $res_2 = W + R_2c$.
- \rightarrow Voter i : res_1, res_2 .

Step G6. Voter i :

- computes (g^{R_1}, h^{R_1}) and (g^{R_2}, h^{R_2}) .
- if $(g^{res_1} \neq P_a(g^{R_1})^c \vee (g^{res_2} \neq P_a(g^{R_2})^c \vee (h^{res_1} \neq P_b(h^{R_1})^c \vee (h^{res_2} \neq P_b(h^{R_2})^c$,
goes to Step G2.

Ballot Casting Phase (please refer to Fig. 3)

Ballot Casting Phase can be invoked at any time prior to the deadline of voting. For Voter i , he will enter this phase after finishing the steps specified in Ballot Generation Phase.

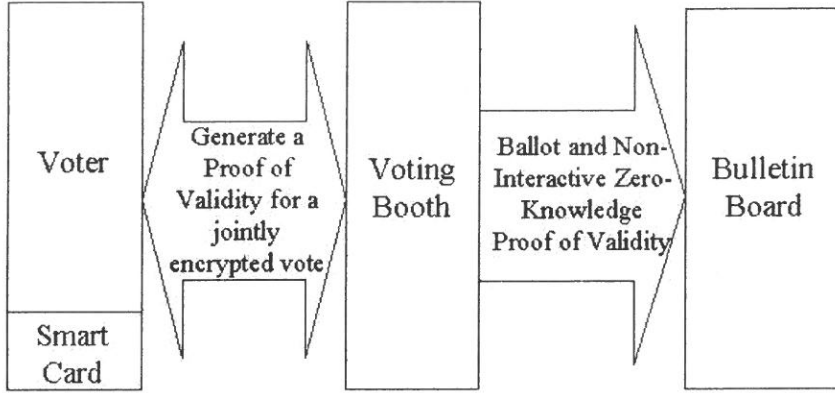


Fig. 3. Ballot Casting Phase

(Case 1) if Voter i chooses to vote $E(+v) = (x, y) = (g^{r+R_1}, h^{r+R_1}G^v)$,

$$\text{where } r = \begin{cases} r_1 & \text{if } v = +1 \\ r_2 & \text{if } v = -1 \end{cases} :$$

Step C1. Voter i :

- uses SC_i to generate random numbers W_{vi} , ran_1 , and d_1 , and compute $(a_1, b_1) = (g^{ran_1}(x)^{d_1}, h^{ran_1}(yG^v)^{d_1})$ and $(a_2, b_2) = (g^{W_{vi}+W_{bt}}, h^{W_{vi}+W_{bt}})$.
- uses SC_i to compute $B = H(ID_i, x, y, a_1, b_1, a_2, b_2)$ and $d = d_2 = B - d_1$.

Step C2. Voter $i \rightarrow$ VB: B, d .

Step C3. VB \rightarrow Voter i : $r_{bt1} (= W_{bt} + R_1d)$ and $r_{bt2} (= W_{bt} + R_2d)$.

Step C4. Voter i :

- uses SC_i to compute $ran_2 = W_{vi} - rd + r_{bt1}$.
- \rightarrow BB : $E(+v), a_1, b_1, a_2, b_2, d_1, d_2, ran_1, ran_2, B$, with signature.

(Case 2) if Voter i chooses to vote $E(-v) = (x, y) = (g^{r+R_2}, h^{r+R_2}G^v)$,

$$\text{where } r = \begin{cases} r_1 & \text{if } v = +1 \\ r_2 & \text{if } v = -1 \end{cases} :$$

Step C1. Voter i :

- uses SC_i to generate random numbers W_{vt} , ran_2 , and d_2 , and compute $(a_1, b_1) = (g^{W_{vt}+W_{bt}}, h^{W_{vt}+W_{bt}})$ and $(a_2, b_2) = (g^{ran_2}(x)^{d_2}, h^{ran_2}(yG^v)^{d_2})$.
- uses SC_i to compute $B = H(ID_i, x, y, a_1, b_1, a_2, b_2)$ and $d = d_1 = B - d_2$.

Step C2. Voter $i \rightarrow$ VB: B, d .

Step C3. VB \rightarrow Voter i : $r_{bt1} (= W_{bt} + R_1d)$ and $r_{bt2} (= W_{bt} + R_2d)$.

Step C4. Voter i :

- uses SC_i to compute $ran_1 = W_{vt} - rd + r_{bt2}$.
- \rightarrow BB : $E(-v), a_1, b_1, a_2, b_2, d_1, d_2, ran_1, ran_2, B$, with signature.

Public Verification Phase (please refer to Fig. 4)

Public Verification Phase can be invoked at any time prior to the tallying phase.

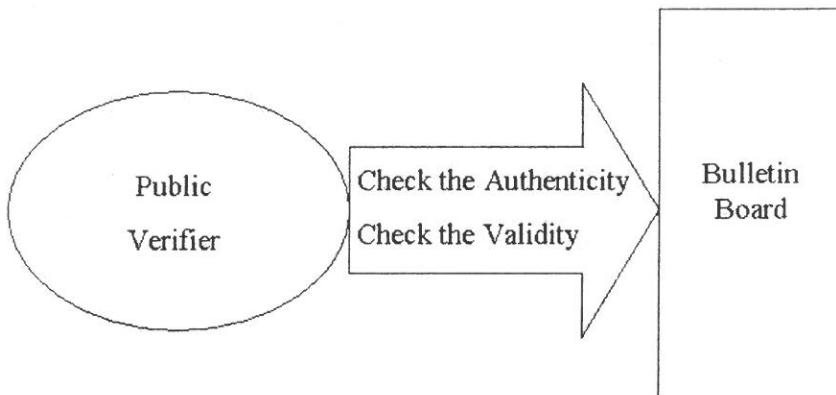


Fig. 4. Public Verification Phase

Step V1. Any party can:

- check the authenticity of $(E(+v)/E(-v), a_1, b_1, a_2, b_2, d_1, d_2, ran_1, ran_2, B)$ by verifying its corresponding signature in each entry on BB. If fails, the representative Voting Authority should clean the entry; otherwise, any party can ask the representative Voting Authority to do so.

- verify the validity of $E(+v)$ (or $E(-v)$) $= (x, y)$ by using its corresponding $(a_1, b_1, a_2, b_2, d_1, d_2, ran_1, ran_2, B)$ in each entry on BB. If fails, the representative Voting Authority should mark the entry as INVALID BALLOT; otherwise, any party can ask the representative Voting Authority to do so.

Tallying Phase (please refer to Fig. 5)

Tallying Phase is invoked only once after deadline of the whole voting.

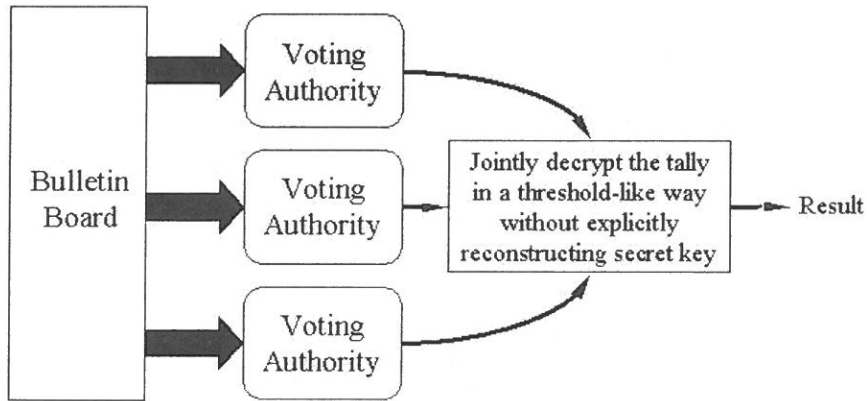


Fig. 5. Tallying Phase

Step T1. Voting Authorities:

- compute $(X, Y) = (\prod x_i, \prod y_i)$, where x_i and y_i denote the valid x and y of Voter i , respectively.
- t voting authorities jointly compute $W = \frac{Y}{X^s} = G^T$ by using the decryption protocol specified in [Ped91, Ped92].
- determine final tally T from W .

- announce the result.

4. Security Analysis of the Proposed e-Voting Scheme

In this section, we analyze the security of the proposed e-voting scheme.

Lemma 1 (completeness). All votes are counted correctly in the proposed e-voting scheme.

Sketch of Proof: Since the bulletin board is open to the public and no ballot can be erased from the bulletin board, any party can verify the validity of each ballot. Therefore, no valid ballot posted on it can be dropped or wrongly handled. Due to the homomorphic properties of the used encryption method, the final tally is the sum of the votes contained in all valid ballots, i.e., all votes are counted correctly. Hence, the proposed e-voting scheme satisfies completeness.

Lemma 2 (privacy). All votes must be secret.

Sketch of Proof: Since the vote is encrypted with the public key shared by the voting authorities, it can be individually decrypted from its corresponding ballot only by using the private key jointly shared by voting authorities. This decryption can be correctly performed only when t or more voting authorities participate. However, it contradicts the assumption that no more than $t-1$ voting authorities conspire. On the other hand, the adversary can directly decrypt the cast ballot only when he has solved the discrete logarithm problem, which is computationally infeasible by using current techniques. Therefore, the proposed e-voting scheme ensures privacy.

Lemma 3 (unreusability). In the proposed e-voting scheme, no voter can vote twice.

Sketch of Proof: Each ballot is posted in the entry designated to the voter on the bulletin board in an authenticated manner. Since a digital signature technique is used to control access to the entries of the bulletin board, any party can verify the authenticity of each ballot by verifying its signature. If a voter wants to vote twice,

he has to cast his extra ballot in the entry designated to another voter. However, his extra ballot will be rejected because he can not generate the correct signature for it. In addition, since each ballot should be accompanied with a proof that it contains a valid vote, i.e., $+1$ or -1 , he can not cast a ballot containing the vote with an invalid value, e.g., $+2$, -2 , and so on. Therefore, the proposed e-voting scheme satisfies unreusability.

Lemma 4 (eligibility). In the proposed e-voting scheme, only the eligible voters can vote.

Sketch of Proof: As stated in proving Lemma 3, each ballot, containing the signature of the voter, is posted in the entry designated to the voter on the bulletin board. No one except the certification authority can impersonate as an eligible voter to vote. In addition, the certification authority will be caught if he undertakes such an impersonation. Note that even if the voting authorities conspire, they can not impersonate as an eligible voter to vote. Therefore, the proposed e-voting scheme satisfies eligibility.

Lemma 5 (fairness). No one can know the intermediate results of the voting in the proposed e-voting scheme.

Sketch of Proof: The private key jointly shared by voting authorities can only be used in the tallying phase once to decrypt the encrypted final tally. It is only when t or more voting authorities conspire prior to the tallying phase that the decryption can be performed to obtain intermediate result. However, it clearly contradicts the assumption that no more than $t-1$ voting authorities conspire. Hence, the proposed e-voting scheme provides fairness to the voters.

Lemma 6 (verifiability). In the proposed e-voting scheme, the result of the voting can be verified.

Sketch of Proof: Since the public can verify the authenticity of each ballot by

verifying its signature and no ballot can be erased from the bulletin board, we can also verify the integrity of the encrypted final tally. Due to the homomorphic properties of the used encryption method, the final tally is verifiable to any observer of the election. Thus, the proposed e-voting scheme satisfies verifiability.

Lemma 7 (receipt-freeness). In the proposed e-voting scheme, the voter cannot reveal his vote to others.

Sketch of Proof: By employing physical voting booths, no one can monitor the voting process of other voters. Since the voter can not control the randomness added to the vote, he can not construct a receipt. During voting, the voter posts a ballot accompanied with a proof that it contains a valid vote, without revealing its actual content. Both the voting booth and the smart card contribute randomness to the ballot so that the voter cannot construct a receipt. The voter has to be convinced that the voting booth has done things correctly, but without finding out the randomness added by it. Before the ballot is submitted to the bulletin board, the voter must be given a proof of correctness of the encryption performed by the voting booth. By employing the Fiat-Shamir heuristic [FS86], the proof for each ballot is non-transferable and can not be used as a receipt. In summary, the proposed e-voting scheme satisfies receipt-freeness.

Theorem 1. The proposed e-voting scheme is secure.

Sketch of Proof: According to Lemma 1 ~ Lemma 7, we conclude that the proposed e-voting scheme is, by Definition 1, secure.

5. Conclusion

In the past, the development of e-voting schemes focuses on providing the utmost convenience to the voter in that the voter can vote at any place by using a personal computer with Internet accessing capability. Bribe and coercion are common in traditional voting systems and usually lead to a biased result that imparts

the desired democracy. These problems become more terrible and harder to solve in an e-voting scheme. Although many e-voting schemes are claimed to provide receipt-freeness, bribe and coercion can not be effectively prevented since the voter may be bribed or coerced to vote for a certain candidate and is obliged to vote under the supervision of the briber or the coercer. To effectively solve or at least lessen these problems, physical voting booths should be used again. So far, several modern e-voting schemes using physical voting booths have been proposed. These schemes are good, but they are not ideal enough. In these e-voting schemes, the briber or the coercer can prescribe private random bits that the voter should use, i.e., these schemes do not effectively provide uncoercibility.

This paper has described a secure e-voting scheme that can be realized in practical environments. As all votes are counted correctly, it satisfies completeness. Moreover, no voter can vote twice, therefore, it satisfies unreusability. Since only the eligible voters can vote, it meets the requirement of eligibility. Because no one can know the intermediate results of the voting, it satisfies fairness. Additionally, it satisfies verifiability since the result of the voting can be verified. In particular, since the voter can not reveal his vote to others, it satisfies receipt-freeness, which actually represents both uncoercibility and receipt-freeness. By using physical voting booths, bribers and coercers can not monitor the voter while he votes. By using smart cards to randomize part of content of the vote, the voter can not construct a receipt. Since the added randomness can not be controlled and learned by the voter, receipt-freeness and uncoercibility are effectively achieved or at least lessened. In summary, the proposed e-voting scheme is practical in that its assumptions are quite appropriate for realistic environments. Furthermore, the performance of the proposed e-voting is optimal in that time and communication complexity for the voter is independent of the number of voting authorities. Although the proposed e-voting scheme only offer the voter a choice between two options, it can be extended, e.g., by using the techniques described by Cramer, Franklin, and Schoenmakers [CFSY96], to provide multi-way voting, i.e., the voter can choose to vote between several options.

Acknowledgment

This research was supported by The Society of the Divine Word (Societas Verbi Divini) China Province at Fu Jen Catholic University, Taiwan, Republic of China, under Grant SVD-9017. The authors are also grateful to the reviewers for their valuable comments.

References

- DH76. W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6), pp. 644-654, 1976.
- Cha81. D. Chaum. Untraceable electronic mail, return addresses and digital pseudonyms. *Communications of ACM*, 24(2):84-88, 1981.
- DLM82. R. Demillo, N. Lynch and M. Merritt. Cryptographic protocols. In *Proc. 14th Annual ACM Symposium, Theory of Computing*, pp. 177-182, 1982.
- Yao82. A. Yao. Protocols for secure communications. In *Proc. 23rd Annual IEEE Symposium Foundations of Computer Science*, pp.160-164, 1982.
- Rab83. M. Rabin. Transaction protection by beacons. *Journal of Computer Systems Science*, 27(2), pp.256-267, 1983.
- ElG84. T. ElGamal. A public-key cryptosystem and a signature scheme based on discrete logarithms. In *Advances in Cryptology —CRYPTO'84*, vol. 196 of *LNCS*, pp.10-18. Springer-Verlag, 1984.
- CF85. J. Cohen and M. Fischer. A robust and verifiable cryptographically secure election scheme. In *Proc. 26th IEEE Symposium on Principles of Distributed Computing (PODC)*, pp.52-62, 1985.
- BY86. J. Benaloh and M. Yung. Distributing the power of a government to enhance the privacy of voters. In *27th IEEE Symposium on Principles of Distributed Computing (PODC)*, pp.52-62, 1986.
- FS86. A. Fiat and A. Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *Advances in Cryptology —*

- CRYPTO*'86, vol. 263 of *LNCS*, pp.186-194. Springer-Verlag, 1986.
- Ben87. J. Benaloh. Verifiable Secret-Ballot Elections. PhD thesis, Yale University, Department of Computer Science Department, New Haven, CT, September 1987.
- Boy87. C. Boyd. Some applications of multiple key ciphers. In *Advances in Cryptology — EUROCRYPT*'88, pp. 234-238. Springer-Verlag, 1987.
- Cha88. D. Chaum. Elections with unconditionally secret ballots and disruption equivalent to breaking RSA. In *Advances in Cryptology — EUROCRYPT*'88, pp. 177-182. Springer-Verlag, 1988.
- Boy90. C. Boyd. A new multiple key ciphers and an improved voting scheme. In *Advances in Cryptology — EUROCRYPT*'89, pp.617-625. Springer-Verlag, 1990.
- Cha91. D. Chaum. Zero-knowledge undeniable signatures. In *Advances in Cryptology — EUROCRYPT*'90, vol. 473 of *LNCS*, pp. 458-464. Springer-Verlag, 1991.
- Ped91. T. Pedersen. A threshold cryptosystem without a trusted party. In *Advances in Cryptology — EUROCRYPT*'91, vol. 547 of *LNCS*, pp. 522-526. Springer-Verlag, 1991.
- FOO92. A. Fujioka, T. Okamoto and K. Ohta. A practical secret voting scheme for large scale elections. In *Advances in Cryptology — AUSCRYPT*'92, pp. 244-251, 1992.
- Ped92. T. Pedersen. Distributed Provers and Verifiable Secret Sharing Based on the Discrete Logarithm Problem. PhD thesis, Aarhus University, Computer Science Department, Aarhus, Denmark, March 1992.
- CP93. D. Chaum and T. P. Pedersen. Wallet databases with observers. In *Advances in Cryptology — CRYPTO*'92, vol. 740 of *LNCS*, pp. 89-105. Springer-Verlag, 1993.
- OFO93. T. Okamoto, A. Fujioka and K. Ohta. A practical large scale secret voting scheme based on non-anonymous channels. In *Proc. of SCIS93, IC*. January 1993.

- BT94. J. Benaloh and D. Tuinstra. Receipt-free secret-ballot elections. In *Proc. 26th Symposium on Theory of Computing (STOC'94)*, pp.544-553, 1994.
- NR94. V. Niemi and A. Renvall. How to prevent buying of votes in computer elections. In *Advances in Cryptology – ASIACRYPT'94*, pp.405-419. Springer-Verlag, 1994.
- SK94. K. Sako and J. Kilian. Secure voting using partially compatible homomorphisms. In *Advances in Cryptology – CRYPTO'94*, vol. 839 of *LNCS*, pp.411-424. Springer-Verlag, 1994.
- Gen95. R. Gennaro. Achieving independence efficiently and securely. In *Proc. 14th ACM Symposium on Principles of Distributed Computing (PODC '95)*, 1995.
- SK95. K. Sako and J. Kilian. Receipt-free mix-type voting scheme—a practical solution to the implementation of a voting booth. In *Advances in Cryptology – EUROCRYPT'95*, vol. 921 of *LNCS*, pp. 393-403. Springer-Verlag, 1995.
- CFSY96. R. Cramer, M. Franklin, B. Schoenmakers, and M. Yung. Multi-authority secret-ballot elections with linear work. In *Advances in Cryptology – EUROCRYPT'96*, vol. 1070 of *LNCS*, pp.72-83. Springer-Verlag, May 1996.
- CGS97. R. Cramer, R. Gennaro and B. Schoenmakers. A secure and optimally efficient multi-authority election scheme. *European Transactions on Telecommunications*, 8:481-489, 1997. Preliminary version in *Advances in Cryptology – EUROCRYPT'97*.
- Oka97. T. Okamoto. Receipt-free electronic voting schemes for large scale elections. In *Proc. of Workshop on Security Protocols '97*, vol. 1361 of *LNCS*, pp. 25-35. Springer-Verlag, 1997.
- HS00. M. Hirt and K. Sako. Efficient receipt-Free voting based on homomorphic encryption. In *Advances in Cryptology – EUROCRYPT 2000*, vol. 1807 of *LNCS*, pp.539-556. Springer-Verlag, 2000.

- MBC01. E. Magkos, M. Burmester and V. Chrissikopoulos. Receipt-freeness in large-scale elections without untappable channels. In *1st IFIP Conference on E-Commerce / E-business / E-Government*, pp. 683-693, Kluwer Academics Publishers, 2001.

received October 30, 2002
revised November 20, 2002
accepted December 14, 2002

以智慧卡為基礎之無收據型實用電子投票系統

顧維祺* 何俊明

輔仁大學資訊工程學系

摘 要

在傳統投票系統中，買票與脅迫投票的情形即已相當普遍，此類行為將使投票的結果產生偏差；然而，在電子投票系統中買票與脅迫投票的問題將更為複雜。在本研究中，鑑於目前的電子投票系統均無法在合理的假設下有效地滿足無收據與抗脅迫的安全需求，我們運用現有技術設計出一套適用於實際環境且可有效處理買票與脅迫投票問題的安全電子投票系統，經由使用簡單的實體電子投票所，買票者與脅迫投票者將無法監控他人投票，我們並藉由智慧卡所產生的隨機值巧妙地使得投票者無法自願或非自願地提供收據給買票者或脅迫投票者。特別的是，本電子投票系統並不需要目前技術無法實現的假設，同時，在計算複雜度與傳輸複雜度上均有理想的表現，因此，本電子投票系統的實用性優於其它現有的電子投票系統。

關鍵詞：電子投票系統，無收據特性，智慧卡，投票所

A Simulation Study of Confidence Intervals of Process Capability

Index C_{pk} via Miss Rate

Sy-Mien Chen* and Hsiao-Wen Huang

Department of Mathematics

Fu-Jen University

Taipei, Taiwan 242, ROC

Abstract

Practitioners are interested in small sample property, but the exact confidence interval estimation for most models in such case is either impractical or impossible. Instead, approximate confidence intervals based on the asymptotic normal theory for the maximum likelihood estimators are widely used in applications. However, for small samples, the $z_{\alpha/2}$ in the approximate confidence intervals is no longer the right value. A good confidence interval for small samples is therefore the one that has the smallest miss rate. In capability study, the index C_{pk} is used more often than any other existing indices at the present time for measuring process capability. In this research, we compare four approximate confidence intervals for both small and large samples of the process capability index C_{pk} based on miss rate.

keywords and phrases: Confidence interval ; Process capability index ; Miss rate.

*Corresponding author. Tel.: +886-2-29031111 ext. 2451; fax: +886-2-29044509
E-mail address: math1013@mails.fju.edu.tw

1. Introduction

A process capability index is a numerical measurement that provides information on whether a production process is capable of producing items within the specification limits predetermined by the designers. Recent research and advances made in this subject are neatly summarized in Kotz and Johnson (1993), Kotz and Lovelace (1998) and Bothe (1997) among others. For statistical inferences problem, Chou *et al.* (1990), Zhang *et al.* (1990), Bissell (1990), Kushler and Hurley (1992), Franklin and Wasserman (1992), Nagata and Nagahata (1994), Chen and Hsu (2002) studied (approximate) confidence bounds for capability indices. Among others, the index C_{pk} is used more often than any other existing indices at the present time for measuring process capability.

Practitioners are interested in small sample property, but the exact confidence interval estimation for most models in such case is either impractical or impossible. Instead, approximate confidence intervals based on the asymptotic normal theory for the maximum likelihood estimators are widely used in applications. However, for small samples, the $z_{\alpha/2}$ in the approximate confidence intervals is no longer the right value. For a given confidence interval, the probability that such random interval does not cover the true value of the parameter is called the miss rate of the procedure. A good confidence interval for small samples is therefore the one that has the smallest miss rate. In this article, we do simulation study on four approximate confidence intervals of the process capability index C_{pk} via miss rate for both small and large samples.

2. Confidence Intervals of C_{pk} .

Let LSL and USL be the lower and the upper specification limits of a product characteristic X , respectively. Denote $d=(USL-LSL)/2$ as the half-length of the specification limits, $M=(USL+LSL)/2$ as the midpoint of the specification limits. Kane (1986) proposed the process capability index C_{pk} , which is defined as the following:

$$\text{Define } C_{pu} = \frac{USL - \mu}{3\sigma}$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma},$$

$$\text{then } C_{pk} = \min(C_{pu}, C_{pl})$$

$$= \frac{d - |\mu - M|}{3\sigma}.$$

From the definition, one can see that C_{pl} and C_{pu} measure how close the outer tail of the process distribution is to the specification limits. And C_{pk} measures the worst one.

When parameters μ and σ are unknown, the index must be estimated. Let $X_1,$

X_2, \dots, X_n be a sample from $N(\mu, \sigma^2)$. Define $\bar{X} = \sum_{i=1}^n X_i / n$, $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

Consider the maximum likelihood estimator $\hat{C}_{pk} = \frac{d - |\bar{X} - M|}{3S}$. Chou and Owen

(1989) had derived the exact distribution for C_{pk} for normally distributed data. However, the expression is too complicated, and is not analytically tractable.

The development of a confidence interval for C_{pk} is a tricky task, some authors have derived their own version of confidence limits for C_{pk} . Some existing approximate confidence intervals for C_{pk} are now reviewed.

(1) BISSELL (Bissell (1990))

$$\left(\hat{C}_{pk} - z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}}, \hat{C}_{pk} + z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}} \right).$$

(2) NNCI-1 (Nagata & Nagahata (1994))

$$\left(\hat{C}_{pk} - z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)} + \frac{1}{30\sqrt{n}}} , \hat{C}_{pk} + z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)} + \frac{1}{30\sqrt{n}}} \right).$$

(3) NNCI-2 (Nagata and Nagahata (1994))

$$\left(\sqrt{1 - \frac{2}{5(n-1)}} \hat{C}_{pk} - z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}} , \sqrt{1 - \frac{2}{5(n-1)}} \hat{C}_{pk} + z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}} \right).$$

It has been shown that NNCI-2 has very good numerical results and both NNCI-1 and NNCI-2 have almost the same large sample properties. (See Kotz (1998)).

(4) CHCI (Chen and Hsu (2002))

Through geometric approach, Chen & Hsu (2002) proposed the following approximate confidence interval:

(a) When the process mean μ is larger than the midpoint M , a confidence interval of C_{pk} is given by

$$\left(\frac{d - \bar{X} + M}{3S} - \frac{\chi_{n-1}^2(1-\alpha)S}{3n(\mu_0 - \bar{X})} , \frac{d - \bar{X} + M}{3S} - \frac{\chi_{n-1}^2(1-\alpha)S}{3n(\mu_1 - \bar{X})} \right).$$

(b) When the process mean μ is smaller than the midpoint M , a confidence interval of C_{pk} is given by

$$\left(\frac{d + \bar{X} - M}{3S} + \frac{\chi_{n-1}^2(1-\alpha)S}{3n(\mu_1 - \bar{X})} , \frac{d + \bar{X} - M}{3S} + \frac{\chi_{n-1}^2(1-\alpha)S}{3n(\mu_0 - \bar{X})} \right),$$

where μ_0 & μ_1 are the intersection points of the level curve of C_{pk} and the joint confidence set of \bar{X} and S , $\frac{n}{S^2}(\mu - \bar{X})^2 + \frac{2n}{S^2}(\sigma - S)^2 \leq \chi_2^2(1-\alpha)$. Without loss of generality, we assume that $\mu_0 \geq \mu_1$. Chen and Hsu (2002) showed that the coverage

rate of CHCI is as good as the coverage rate of NNCI-2.

3. Simulation Study.

In this section, we compare four approximate confidence intervals of C_{pk} , namely, BISSELL, NNCI-1, NNCI-2 and CHCI, for datum that are in statistical control and are independently normally distributed. Some important terms about confidence intervals are defined first. The asymptotic value of a miss rate (an error probability) is called the nominal error probability, whereas the miss rate (error probability) achieved in finite samples is called the actual miss rate (actual error probability). The effect of sample size n and process mean μ on miss rate under several combinations of parameters are studied. Also, the relative error of actual miss rate to nominal miss rate is studied.

IMSL is used in this simulation study to generate random number and calculations. EXCEL is used to draw diagrams. When one generates random numbers from normal distributions $N(\mu, \sigma^2)$, the values of σ are determined by μ and C_{pk} .

3.1 Parameters in Simulation

1. nominal miss rate $\alpha = 0.01 ; 0.05 ; 0.1$.
2. lower specification limit, $LSL = -3$.
3. upper specification limit, $USL = 3$.
4. mid point of the specification limits $M = 0.0$.
5. process capability index, $C_{pk} = 1.0 , 1.33 , 2.0$.

These three index values correspond to a process with 13 out of 1000, 3 out of 100000, 1 out of 100000000 parts outside tolerance limits, respectively.

6. sample size $n = 30 , 70 , 100 , 200 , 300 , 400 , 500 , 1000 , 2000$.
7. the true process mean $\mu = 0.0 (0.1) 1.5$ (To see the effect of process shift to the miss rate. Here we only consider positive shift) .

3.2 Simulation Results

The results are similar when the nominal miss rate α are 0.01, 0.05 and 0.1. To save the space, we show only the case when $\alpha = 0.05$ in detail. First we consider the effect of the process mean shift to the miss rate when the sample size is fixed. From Fig 1, we see that when the sample size is small, say $n=30$, the actual miss rate derived from CHCI is smaller than the nominal value when the process mean is at the midpoint M . When the process mean shifts away from M , the miss rate increases rapidly and becomes stable when the shift amount achieves 0.3. However, the miss

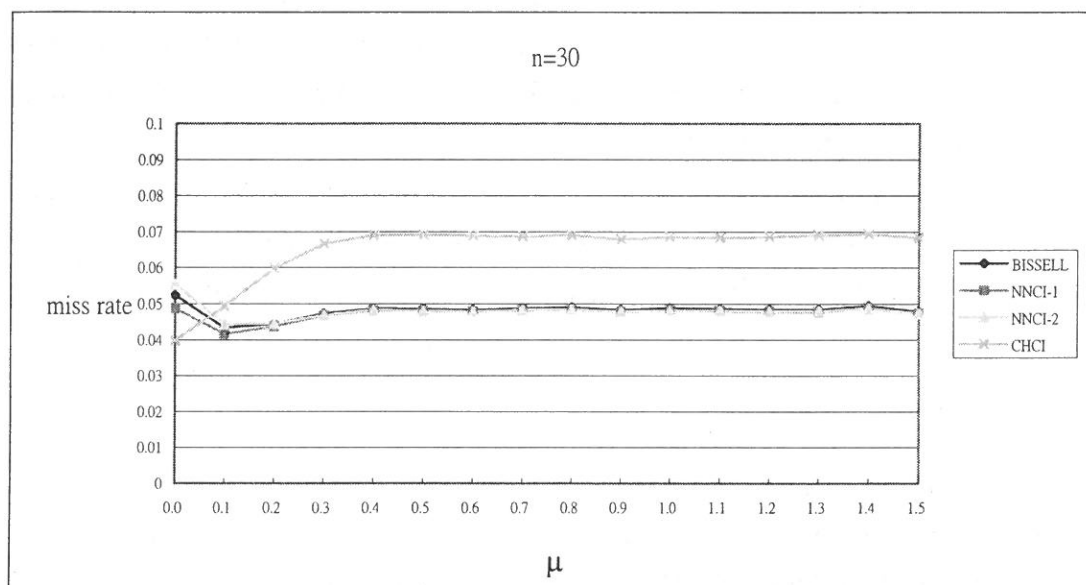


Fig 1 : miss rate of confidence intervals for $n=30$, $\alpha=0.05$, $C_{pk}=1.0$

rate is much too high (about 7%). The miss rates derived from the other three methods are not stable at the beginning, but become stable and are very close to the nominal value 5% when the process mean shift achieve 0.3. When the sample size n becomes larger, and when the process mean shifts, miss rate derived from all the methods are all very close to the nominal value (See Fig 2). Overall speaking, CHCI is recommended when the process mean is at the midpoint of the specification limits, and when the sample size is small. But if the process mean shifts even only slightly,

the NNCI-1 is recommended. If the mean shifts further, say $\mu=0.4$, then BISSEL, NNCI-1 and NNCI-2 performed equally well. When the sample size is also large, then all the four confidence methods propose the same information (See Fig 3).

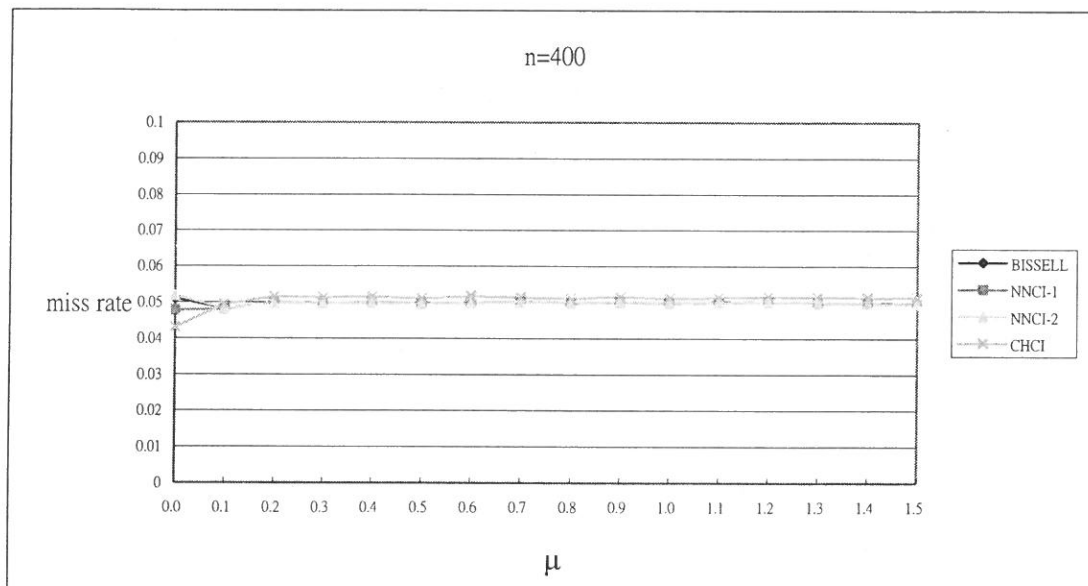


Fig 2 : miss rate of confidence intervals for $n=400$, $\alpha=0.05$, $C_{pk}=1.0$

Next, let's consider the effect of sample size when the process mean μ is fixed. Figure 4 shows the miss rate for different sample sizes when the process mean is at M .

Notice that when the sample size is small, the limiting theorem need not be valid. Hence for small sample size the best method should be the one that gives the smallest miss rate. As one can see from Fig 4, miss rate derived from CHCI is the smallest regardless the sample size. When the sample size is large enough, the miss rate is quite close to the nominal value. NNCI-1 is next. When the sample size is large enough, all four methods give about the same miss rate and are all close to the nominal miss rate.

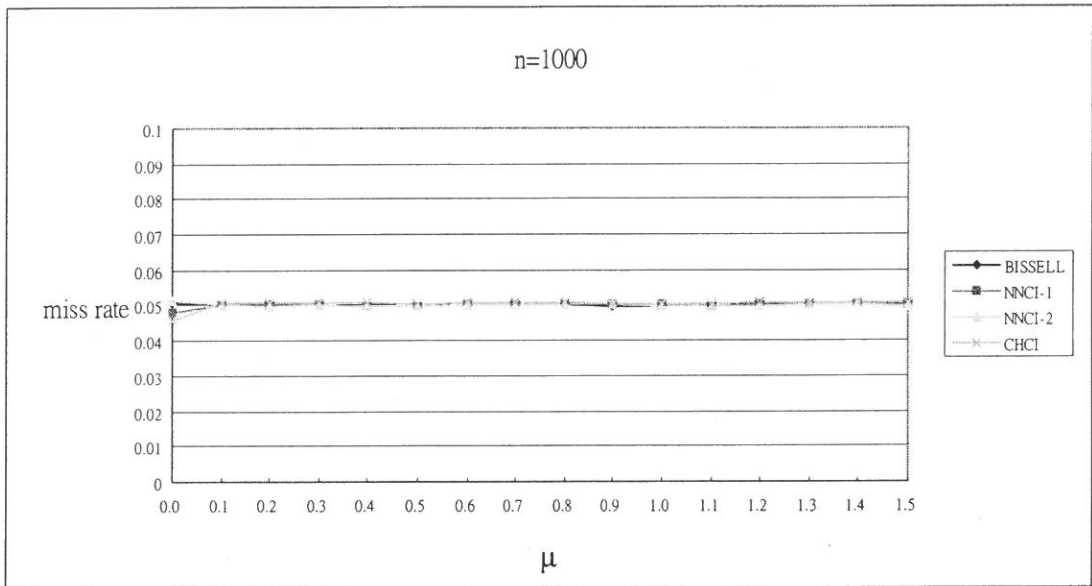


Fig 3 : miss rate of confidence interval for $n=1000$, $\alpha =0.05$, $C_{pk}=1.0$

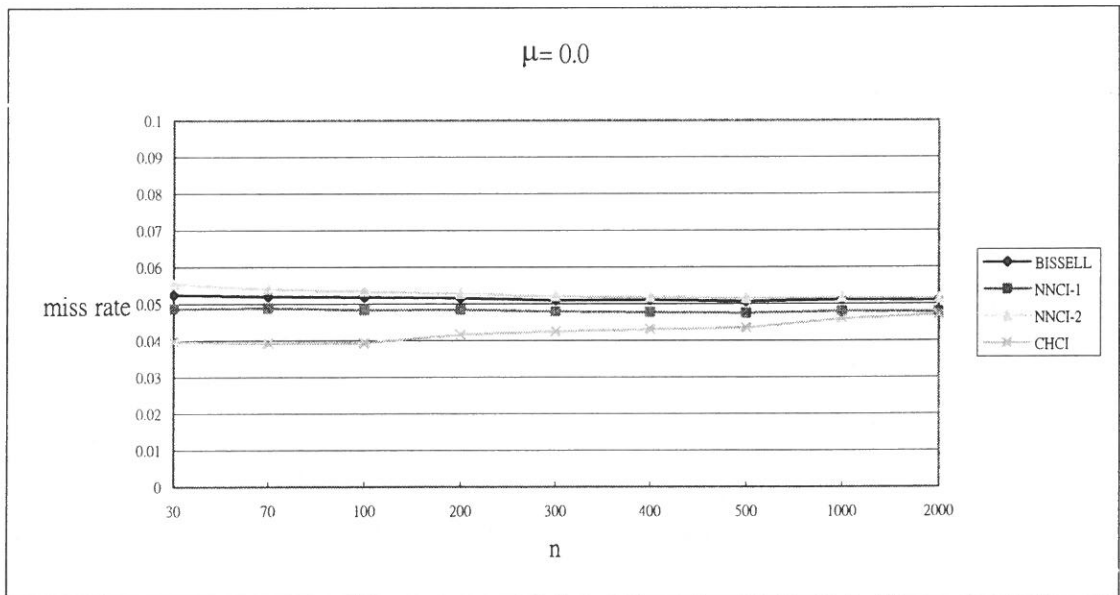


Fig 4 : miss rate of confidence intervals for $\mu =0.0$, $\alpha =0.05$, $C_{pk}=1.0$

When the process mean shifts, BISSELL, NNCI-1 and NNCI-2 perform the best (See Fig 5). When the sample size is as large as 1000, CHCI is also recommended.

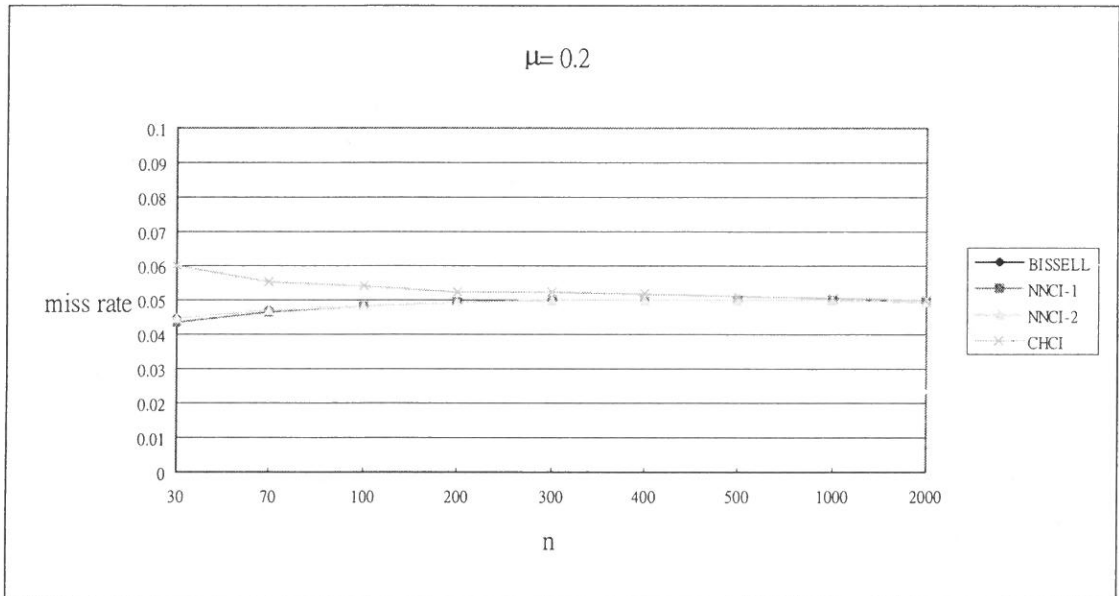


Fig 5 : miss rate of confidence intervals for $\mu = 0.2$, $\alpha = 0.05$, $C_{pk} = 1.0$

Next, we compare the four confidence intervals based on the relative error of miss rate (REMR) where

$$\text{relative error of miss rate (REMR)} = \frac{\text{actual miss rate} - \text{nominal miss rate}}{\text{nominal miss rate}}$$

Since the limiting distribution holds only for large sample, we then discuss the case when the sample size is at least 100. From Fig 6, Fig 7 and Table 1, one can see that the relative error rate of BISSELL, NNCI-1, NNCI-2 and CHCI are more sensitive to the process mean shift when the amount of shift is small. When the shift amount increases, the relative error rate becomes more stable. When the sample size is large enough, say above 1000, then no matter how serious the shift amount is, the relative error rate of BISSELL, NNCI-1, NNCI-2 and CHCI are all within 1%.

Overall speaking, BISSELL, NNCI-1 and NNCI-2 are less affected by the sample size and the true process mean μ . When the sample size is as large as 1000,

then no matter how serious the process mean shifts, CHCI can always propose good information.

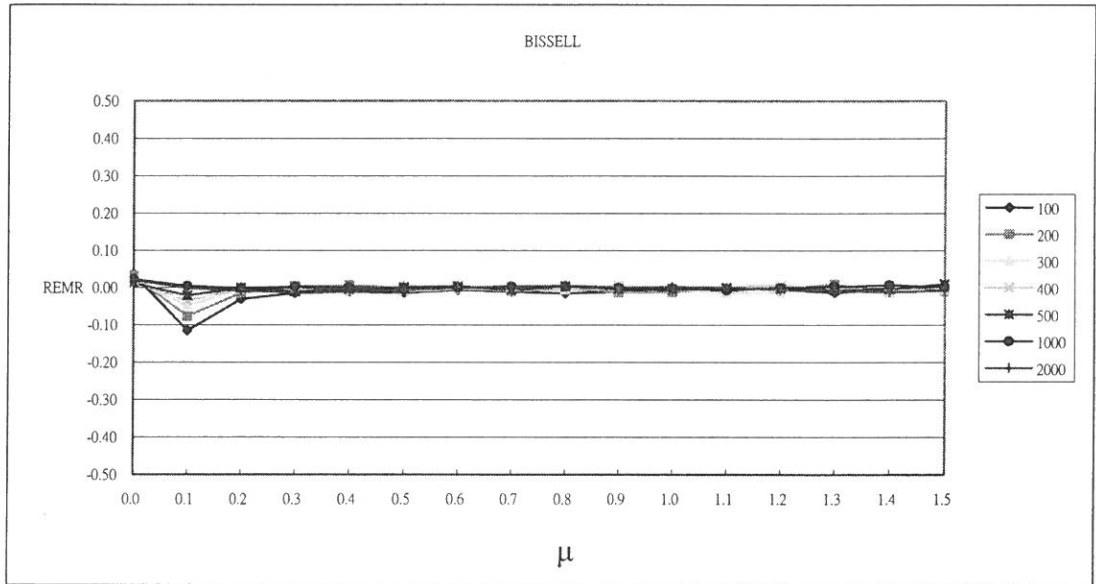


Fig 6 : REMR of BISSELL confidence interval for $\alpha = 0.05$, $C_{pk} = 1.0$

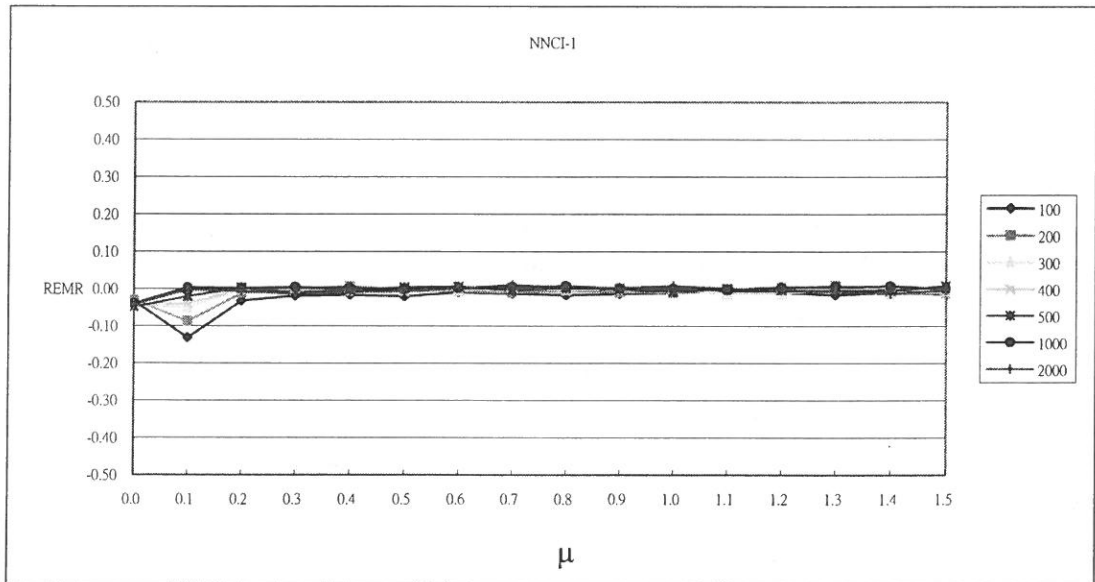


Fig 7 : REMR of NNCI-1 confidence interval for $\alpha = 0.05$, $C_{pk} = 1.0$

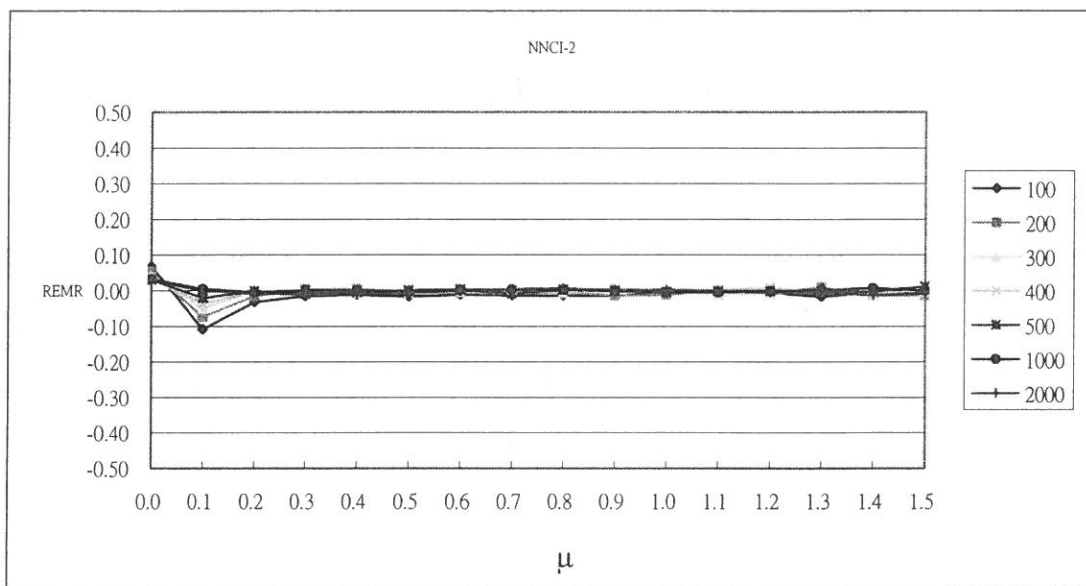


Fig 8 : REMR of NNCI-2 confidence intervals for $\alpha = 0.05$, $C_{pk} = 1.0$

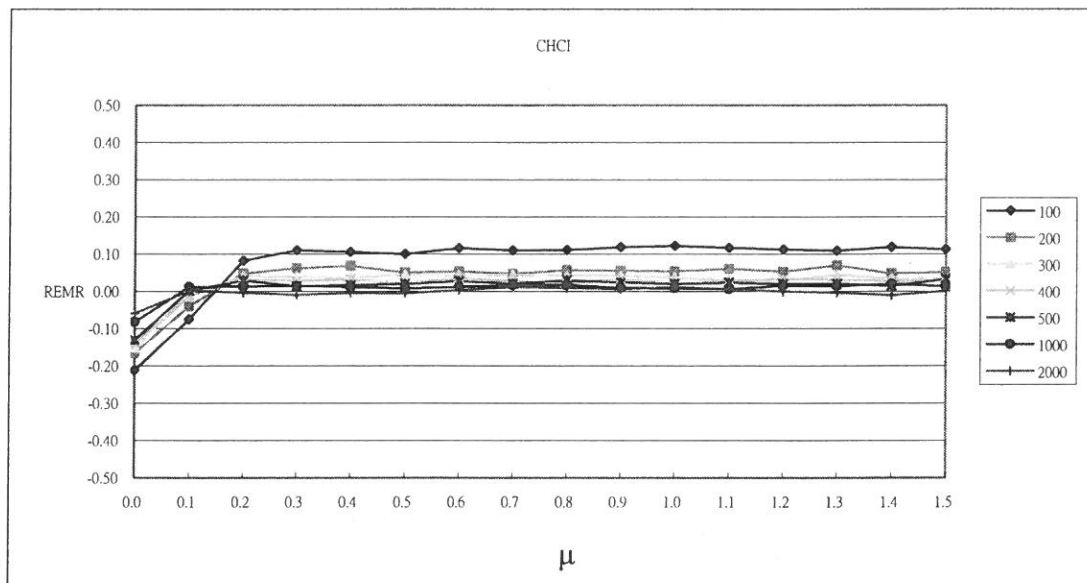


Fig 9 : REMR of CHCI confidence intervals $\alpha = 0.05$, $C_{pk} = 1.0$

method	$\mu \backslash n$	100	200	300	400	500	1000	2000
BISSELL	0.0	0.03620	0.03160	0.01980	0.02060	0.01220	0.02240	0.02020
	0.1	-0.11580	-0.07700	-0.04940	-0.03560	-0.02100	0.00440	-0.00160
	0.2	-0.03080	-0.01400	-0.00160	-0.00120	-0.00020	-0.00200	-0.00740
	0.3	-0.01460	0.00380	-0.00700	-0.00560	-0.01040	0.00260	-0.01260
	0.4	-0.01000	0.00620	-0.00700	-0.00280	-0.00420	-0.00100	-0.01140
	0.5	-0.01360	-0.00020	0.00520	-0.00980	0.00100	-0.00400	-0.00620
	0.6	-0.00620	-0.00080	-0.00100	-0.00320	0.00400	0.00100	0.00040
	0.7	-0.01020	-0.00840	0.00520	0.00100	-0.00520	0.00340	0.00400
	0.8	-0.01500	-0.00240	-0.00380	-0.00200	0.00460	0.00360	0.00580
	0.9	-0.01040	-0.01180	0.00300	0.00400	-0.00320	-0.00080	0.00060
	1.0	-0.00520	-0.01140	-0.00220	-0.00100	-0.00520	-0.00180	0.00260
	1.1	-0.00040	-0.00160	-0.01320	0.00280	-0.00080	-0.00560	0.00120
	1.2	-0.00440	-0.00240	-0.01140	0.00620	-0.00260	-0.00040	-0.00520
	1.3	-0.01320	0.00980	0.00000	-0.00280	-0.00580	0.00240	-0.00780
	1.4	-0.00420	-0.01040	-0.01080	-0.00400	-0.00300	0.00740	-0.01180
	1.5	-0.00980	-0.00940	-0.00360	-0.00620	0.01060	0.00140	-0.00620
NNCI-1	0.0	-0.03320	-0.03100	-0.04240	-0.04520	-0.04980	-0.04020	-0.04280
	0.1	-0.13200	-0.08680	-0.05020	-0.03920	-0.02040	0.00360	-0.00240
	0.2	-0.03220	-0.01220	-0.00120	-0.00040	0.00260	0.00100	-0.00700
	0.3	-0.01860	0.00000	-0.00340	-0.00700	-0.01080	0.00460	-0.01400
	0.4	-0.01600	0.00640	-0.00480	-0.00020	-0.00320	0.00000	-0.00820
	0.5	-0.02060	-0.00340	0.00340	-0.00740	0.00300	-0.00080	-0.00680
	0.6	-0.00940	-0.00600	-0.00300	-0.00180	0.00620	0.00140	0.00200
	0.7	-0.01260	-0.01000	0.00480	0.00240	-0.00380	0.00500	0.01000
	0.8	-0.01720	-0.00440	-0.00080	-0.00420	0.00280	0.00760	0.00460
	0.9	-0.01360	-0.00980	-0.00020	0.00220	-0.00100	0.00080	0.00280

Table 1 : REMR when $\alpha = 0.05$ and $C_{pk} = 1.0$

method	$\mu \backslash n$	100	200	300	400	500	1000	2000
	1.0	-0.01000	-0.01000	-0.00280	-0.00420	-0.00660	0.00080	0.00780
	1.1	-0.00400	-0.00180	-0.01200	0.00000	0.00000	-0.00340	0.00080
	1.2	-0.00960	-0.00280	-0.01160	0.00500	-0.00360	0.00440	-0.00340
	1.3	-0.01620	0.00940	0.00000	-0.00160	-0.00520	0.00500	-0.00800
	1.4	-0.00900	-0.01220	-0.01240	-0.00380	-0.00420	0.00800	-0.01260
	1.5	-0.01440	-0.01120	-0.00340	-0.00780	0.00800	0.00040	-0.00280
NNCI-2	0.0	0.06780	0.05420	0.03960	0.03720	0.03020	0.03260	0.02980
	0.1	-0.10800	-0.07320	-0.04940	-0.03560	-0.02080	0.00460	-0.00120
	0.2	-0.03240	-0.01500	-0.00220	-0.00300	-0.00080	-0.00440	-0.00740
	0.3	-0.01480	0.00240	-0.00700	-0.00660	-0.01020	0.00120	-0.01180
	0.4	-0.01160	0.00400	-0.00680	-0.00400	-0.00560	-0.00040	-0.01120
	0.5	-0.01660	-0.00240	0.00200	-0.00980	0.00060	-0.00440	-0.00660
	0.6	-0.01140	-0.00540	-0.00280	-0.00400	0.00400	0.00160	-0.00100
	0.7	-0.01300	-0.00940	0.00380	0.00080	-0.00820	0.00320	0.00460
	0.8	-0.01460	-0.00460	-0.00520	-0.00220	0.00260	0.00260	0.00600
	0.9	-0.01460	-0.01460	0.00360	0.00480	-0.00160	0.00020	0.00120
	1.0	-0.00800	-0.01160	-0.00260	-0.00020	-0.00580	-0.00320	0.00280
	1.1	-0.00180	-0.00180	-0.01240	0.00220	-0.00220	-0.00520	0.00040
	1.2	-0.00600	-0.00480	-0.01100	0.00600	-0.00160	-0.00160	-0.00620
	1.3	-0.01700	0.00920	-0.00140	-0.00280	-0.00640	0.00140	-0.00800
	1.4	-0.00500	-0.01420	-0.01360	-0.00460	-0.00180	0.00700	-0.01220
	1.5	-0.01020	-0.01220	-0.00340	-0.00680	0.01140	-0.00020	-0.00680
CHCI	0.0	-0.21200	-0.16620	-0.15000	-0.13980	-0.13080	-0.08240	-0.05860
	0.1	-0.07480	-0.04020	-0.01400	-0.01180	0.00160	0.01200	0.00160
	0.2	0.08200	0.04780	0.04220	0.03140	0.02900	0.01260	-0.00300
	0.3	0.11020	0.06240	0.03880	0.02740	0.01400	0.01480	-0.00960

Table 1 (conti.) : REMR when $\alpha = 0.05$ and $C_{pk} = 1.0$


method	$\mu \backslash n$	100	200	300	400	500	1000	2000
	0.4	0.10640	0.06860	0.03600	0.03180	0.01700	0.01160	-0.00460
	0.5	0.10000	0.05020	0.04480	0.02140	0.01980	0.00740	-0.00500
	0.6	0.11660	0.05380	0.04240	0.03620	0.02700	0.01200	0.00380
	0.7	0.11020	0.04740	0.04440	0.02680	0.02060	0.01700	0.01180
	0.8	0.11100	0.05700	0.04160	0.02220	0.02960	0.01680	0.00940
	0.9	0.11920	0.05580	0.04220	0.03160	0.02400	0.00920	0.00640
	1.0	0.12180	0.05400	0.03800	0.02220	0.02060	0.00820	0.01160
	1.1	0.11700	0.06060	0.02800	0.02740	0.02300	0.00640	0.00600
	1.2	0.11320	0.05320	0.02780	0.03340	0.02040	0.01520	0.00020
	1.3	0.10960	0.07000	0.04440	0.03100	0.02080	0.01440	-0.00380
	1.4	0.11900	0.04780	0.03360	0.02660	0.01480	0.02000	-0.01000
	1.5	0.11320	0.05240	0.03400	0.02780	0.03260	0.01380	0.00080

Table 1 (conti.) : REMR when $\alpha = 0.05$ and $C_{pk} = 1.0$

From the discussion above, we know that no matter which confidence interval we use, the miss rate becomes stable when the process mean shift achieves certain amount or when the sample size is large enough. In the following discussion, we compare the maximal miss rate of four confidence intervals when the miss rate becomes stable. The simulation results for $C_{pk}=1.0, 1.33$ and 2.0 and $\alpha=0.01, 0.05$ and 0.1 are all listed in Table 2.

Confidence Interval	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	Cpk=1.0	Cpk =1.33	Cpk =2.0	Cpk =1.0	Cpk =1.33	Cpk =2.0	Cpk =1.0	Cpk =1.33	Cpk =2.0
	n=30								
BISSELL	0.00927	0.00925	0.00933	0.04943	0.04934	0.04928	0.09995	0.09972	0.09944
NNCI-1	0.00908	0.00897	0.00914	0.04890	0.04886	0.04895	0.09947	0.09904	0.09902
NNCI-2	0.00934	0.00933	0.00949	0.04876	0.04884	0.04905	0.09863	0.09863	0.09850
CHCI	0.02678	0.02777	0.02802	0.06936	0.07083	0.07130	0.11788	0.11790	0.11825
	n=70								
BISSELL	0.00984	0.00998	0.01007	0.05011	0.04979	0.05002	0.10038	0.10024	0.10040
NNCI-1	0.00963	0.00982	0.00995	0.04981	0.04950	0.04980	0.10018	0.10008	0.10001
NNCI-2	0.00981	0.01001	0.01010	0.04994	0.04976	0.04992	0.09994	0.09993	0.09989
CHCI	0.01731	0.01791	0.01810	0.05871	0.05909	0.05952	0.10812	0.10815	0.10815
	n=100								
BISSELL	0.01003	0.01004	0.01000	0.04998	0.04996	0.05029	0.10002	0.10068	0.10041
NNCI-1	0.00993	0.00984	0.00993	0.04980	0.04968	0.05018	0.09987	0.10054	0.10028
NNCI-2	0.01003	0.01007	0.01006	0.04991	0.04982	0.05029	0.09980	0.10037	0.10014
CHCI	0.01538	0.01550	0.01573	0.05609	0.05642	0.05664	0.10541	0.10655	0.10620
	n=200								
BISSELL	0.01008	0.01008	0.01000	0.05049	0.05028	0.05020	0.10059	0.10076	0.10043
NNCI-1	0.01008	0.01001	0.00993	0.05047	0.05012	0.05010	0.10071	0.10054	0.10055
NNCI-2	0.01008	0.01009	0.01004	0.05046	0.05034	0.05015	0.10034	0.10055	0.10027
CHCI	0.01273	0.01276	0.01291	0.05350	0.05338	0.05342	0.10345	0.10360	0.10314

 Table 2 : maximal miss rate for C_{pk}

Remark :  indicates cases that $\left| \max_{0.3 \leq \mu \leq 1.5} \text{actual miss rate} - \alpha \right| < 10^{-3}$

Confidence Interval	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	Cpk=1.0	Cpk =1.33	Cpk =2.0	Cpk =1.0	Cpk =1.33	Cpk =2.0	Cpk =1.0	Cpk =1.33	Cpk =2.0
	n=300								
BISSELL	0.01010	0.01014	0.01014	0.05026	0.05051	0.05065	0.10051	0.10057	0.10070
NNCI-1	0.01001	0.01006	0.01012	0.05024	0.05030	0.05068	0.10050	0.10049	0.10081
NNCI-2	0.01009	0.01012	0.01012	0.05019	0.05053	0.05063	0.10053	0.10046	0.10065
CHCI	0.01191	0.01198	0.01216	0.05224	0.05254	0.05273	0.10211	0.10209	0.10237
	n=400								
BISSELL	0.01019	0.01008	0.01010	0.05031	0.05035	0.05040	0.10036	0.10062	0.10037
NNCI-1	0.01010	0.01003	0.01008	0.05025	0.05035	0.05043	0.10049	0.10077	0.10040
NNCI-2	0.01022	0.01009	0.01012	0.05030	0.05024	0.05040	0.10028	0.10052	0.10034
CHCI	0.01158	0.01159	0.01172	0.05181	0.05199	0.05205	0.10160	0.10194	0.10168
	n=500								
BISSELL	0.01021	0.01014	0.01013	0.05053	0.05019	0.05014	0.10085	0.10027	0.10063
NNCI-1	0.01027	0.01007	0.01013	0.05040	0.05017	0.05018	0.10086	0.10041	0.10055
NNCI-2	0.01021	0.01015	0.01010	0.05057	0.05016	0.05010	0.10084	0.10018	0.10054
CHCI	0.01131	0.01124	0.01129	0.05163	0.05146	0.05136	0.10165	0.10151	0.10173
	n=1000								
BISSELL	0.01009	0.01017	0.01012	0.05037	0.05028	0.05038	0.10063	0.10051	0.10037
NNCI-1	0.01011	0.01010	0.01009	0.05040	0.05030	0.05034	0.10092	0.10054	0.10021
NNCI-2	0.01009	0.01016	0.01011	0.05035	0.05028	0.05044	0.10056	0.10049	0.10028
CHCI	0.01064	0.01065	0.01065	0.05100	0.05091	0.05109	0.10119	0.10113	0.10074
	n=2000								
BISSELL	0.01014	0.01007	0.01008	0.05029	0.05044	0.05023	0.10051	0.10027	0.10039
NNCI-1	0.01014	0.01010	0.01009	0.05050	0.05049	0.05030	0.10056	0.10034	0.10044
NNCI-2	0.01015	0.01009	0.01011	0.05030	0.05044	0.05022	0.10046	0.10023	0.10031
CHCI	0.01032	0.01032	0.01048	0.05059	0.05077	0.05061	0.10061	0.10049	0.10064

Table 2 (conti.) : maximal miss rate for C_{pk}

Overall speaking, the maximal miss rate of confidence intervals BISSELL, NNCI-1 and NNCI-2 are less affected by the process capability. For CHCI, the maximal miss rate for more capable process is higher than the maximal miss rate for less capable process. But if the sample size is large enough, say 500, then the maximal miss rate become more stable (See Table 3). The larger the sample size, the closer the maximal miss rate of CHCI is to the nominal miss rate.

n	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	Cpk=1.0	Cpk=1.33	Cpk=2.0	Cpk=1.0	Cpk=1.33	Cpk=2.0	Cpk=1.0	Cpk=1.33	Cpk=2.0
500	0.01131	0.01124	0.01129	0.05163	0.05146	0.05136	0.10165	0.10151	0.10173
1000	0.01064	0.01065	0.01065	0.05100	0.05091	0.05109	0.10119	0.10113	0.10074
2000	0.01032	0.01032	0.01048	0.05059	0.05077	0.05061	0.10061	0.10049	0.10064
3000	0.01023	0.01028	0.01028	0.05089	0.05050	0.05066	0.10104	0.10029	0.10061
4000	0.01021	0.01027	0.01017	0.05106	0.05069	0.05056	0.10096	0.10110	0.10098
5000	0.01052	0.01051	0.01020	0.05090	0.05053	0.05044	0.10088	0.10098	0.10075

Table 3 : maximal miss rate of CHCI

The performance of four confidence intervals when the process capability index $C_{pk} = 1.33$ and 2 are the similar, for those who are interested in those results can contact the first author of this article.

4. Conclusions

In general, the easiness and accuracy are the main concern when one chooses inference methods. From the discussion in the previous section, we see that methods BISSELL, NNCI-1 and NNCI-2 are least affected by the sample size and the process mean shift, and the actual miss rate is closest to the nominal value. Meanwhile, the calculations of these three methods are easier too. Now a day, the computational

function of computer is very powerful, hence CHCI is also recommended when the process mean is at the midpoint of the specification limits. In Table 4 we list proper confidence intervals for different kind of combinations as a reference to the users of process capability index.

Parameters			Recommended Confidence Methods
Real value of C_{pk}	True Process Mean μ	Sample size n	
1.0	$\mu = 0.0$	$n \leq 100$	CHCI, NNCI-1
		$100 < n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2
	$\mu \geq 0.1$	$n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2, CHCI
1.33	$\mu = 0.0$	$n \leq 100$	CHCI, NNCI-1
		$100 < n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2
	$\mu \geq 0.1$	$n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2, CHCI
2.0	$\mu = 0.0$	$n \leq 100$	CHCI, NNCI-1
		$100 < n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2
	$\mu \geq 0.1$	$n < 1000$	BISSELL, NNCI-1, NNCI-2
		$n \geq 1000$	BISSELL, NNCI-1, NNCI-2, CHCI

Table 4 : Recommended confidence intervals for different kind of combinations of parameters

References

1. Bissell, A. F. (1990). " How Reliable is Your Capability Index? ". *Applied Statistics*, 39,331-340

2. Bothe, D. R. (1997). "Measuring Process Capability." McGraw Hill.
3. Chen, S. M and Hsu, Y. S. (2002). "Geometric Interpretations for Confidence Intervals of Process Capability Indices C_{pk} and C_{pm} ." Submitted to *Mathematical Methods of Statistics*.
4. Chou, Y. M., Owen, D. B. and Borrego A. S. A. (1990). "Lower Confidence Limits on Process Capability Indices." *Journal of Quality Technology*, 22, 3, 223-229.
5. Franklin, L.A. and Wasserman, G.S. (1992). "A note on the conservative nature of the tables of lower confidence limits for C_{pk} with a suggested correction," *Communication in Statistics – Simulation and Computation*.
6. Kane, V. E. (1986). "Process Capability Indices." *Journal of Quality Technology*, 18, 41-52.
7. Kotz, S. and Johnson, N. L. (1993). "Process Capability Indices." Chapman & Hall.
8. Kotz, S. and Lovelace C. R. (1998). "Process Capability Indices in Theory and Practice." Arnold.
9. Kushler, R. H. and Hurley P. (1992). "Confidence Bounds for Capability Indices." *Journal of Quality Technology*, 24, 4, 188-195.
10. Nagata, Y. (1992). "Interval estimation for the process capability indices." *J. Japan. Soc. Qual. Control*, 21, 109-114.
11. Nagata, Y. and Nagahata, H. (1994). "Approximation formulas for the lower confidence limits of process capability indices." *Okayama Economic Review*, 25(4), 301-314.
12. Zhang, N. F., Stenback, G.A. and Wardrop, D.M. (1990). "Interval Estimation of Process Capability Index C_{pk} ." *Communications in Statistics – Theory and Methods*, 19, 4455-4470.

received October 30, 2002

revised November 22, 2002

accepted December 15, 2002

製程能力指標 C_{pk} 之信賴區間之擬研究

陳思勉* 黃筱雯

摘 要

實業界對小樣本的理論結果較感興趣。但是對大部份的模型而言，小樣本之下的信賴區間若不是不實際便是不可能得到。因此大樣本之下所得的近似信賴區間便常被實業界引用。然而對小樣本而言，大樣本之下所得的近似信賴區間中 $z_{\alpha/2}$ 這項不見得適合。因此對小樣本而言，一個好的信賴區間應能提供小的誤判率。本文中我們將根據誤判率同時對大小樣本之下比較製程能力指標 C_{pk} 的四種近似信賴區間。

關鍵字：信賴區間；製程能力指標；誤判率

Executable Test Sequence for the Protocol Control and Data Portions¹

Wen-Huei Chen*

Department of Electronic Engineering

Fu Jen University

Taipei, Taiwan 242, R.O.C.

Abstract

A new method is proposed to generate the *executable test sequence* for testing a protocol implementation's conformance to its specification. The executable test sequence can simultaneously verify the protocol's control and data portions, which are modeled as a deterministic finite state machine (FSM) and a set of rules between parameter values respectively. The method involves converting the FSM and rules into a *SelectO digraph*, in which the tour can be used to generate the executable test sequence (i.e., one associated with feasible parameter values) that checks the transitions of the FSM and the rules of the data portion. The *Selecting Chinese Postman Algorithm* is then used to find a specific tour for minimizing the length of the test sequence that checks each transition and each rule at least once simultaneously. Experimentation on the Simple Session Protocol indicates that an optimally executable test sequence can be achieved which is 37% shorter than that achieved by testing the two portions separately.

*Corresponding author. Tel.: +886-2-29031111 ext. 3739; fax: +886-2-29042638
E-mail address: whchen@ee.fju.edu.tw

¹ A preliminary version of this paper appears on the IEEE Int'l Conference on Communication (ICC), New Orleans, Louisiana, U.S.A., June, 2000. This work is supported in part by the National Science Council, Taiwan, R.O.C. under Grant 88-2213-E-130-008.

Keywords: protocol, conformance testing, executable test sequence, SelectO digraph.

1. Introduction

The objective of protocol conformance testing is to see if a protocol implementation conforms to the protocol specification. A black box is the assumed implementation, which is tested by an externally generated sequence of inputs, and then verifying that those inputs result in a sequence of expected outputs. The sequence of input/output pairs is the *test sequence* [13]. A continuous portion of the test sequence, used for checking a specific property of the protocol, is called a *test segment*. The number of input/output pairs of the test sequence (test segment) is called its length. The protocol specification, from which test sequences are generated, contains a control and a data portion. The control portion can be considered as a *deterministic² Finite State Machine* (FSM) [14] which contains *states* and *transitions*. Initially, the FSM remains at a specific state called the *initial state*. An input (i.e., stimulus) will cause the FSM to generate output(s) (i.e., responses) and to change from the current state to a new state; this process is realized by a transition. The data portion specifies the values of supplementary parameters (e.g., quality of service in the connection). In general, the relations between messages' parameter values can be considered as data flow digraphs[17], or program segments [7], or rules between parameter values [8][9].

Various formal methods have been proposed to verify the control portion modeled as an FSM [18][19]. The typical aim of these methods is to generate a test sequence which checks whether each transition specified in the FSM is correctly implemented. In the T- method of [20], each transition is checked by an input/output pair. Experimentation of [21] suggests that checking each transition by a simple input/output cannot detect a lot of faulty implementations, which transfer to an incorrect state upon receiving the input. In the U-method of [22], each transition

² "deterministic" means that for each input there is at most one transition defined at each state.

(assumes it ends at state J) is checked by an input/output pair, followed by a Unique Input/Output (UIO) sequence of state J to ensure that the FSM transfers to the expected state J. The input/output pair and the UIO sequence forms a test segment which checks the transition. In [1], the *Rural Chinese Postman Algorithm* is proposed to connect all the (non overlapping) test segments into a “minimum-length” test sequence. In [11], Miller and Paul overlap these test segments into an even shorter test sequence. Other well-known formal methods include the W-, Wp- and Uv- methods [18][19]. However, they produce a test sequence which is significantly longer than that produced by the U-method used with the overlapping technique of [11], which has become a popular method in generating a test sequence for verifying the control portion.

Various formal methods have been proposed to verify the data portion based on different models. In the data portion modeled as data flow digraphs or program segments, a test sequence becomes *executable* if it is associated with parameter values which do not violate the data flow properties described in these models. Methods have been proposed to generate an executable test sequence that checks important data flow properties [7][16][17]; however, these methods are not completely automatic (the operator must interfere to determine the executability of the test sequence), as a limitation due to the decidability theory [16]. In [8][9], the data portion is modeled as a set of rules between parameter values. The FSM and rules are converted into a *Behavior Machine Digraph*, in which tours can be used to automatically generate executable test segments verifying the rules. In [2], Chen proposes a *Selecting Chinese Postman Algorithm* to find a specific tour of a Selecting Digraph (constructed from the Behavior Machine Digraph) for automatically generating a minimum-length executable test sequence that checks each rule at least once.

The earlier methods of testing the two portions in two different phases introduces redundant test segments. In this paper, we are going to combine Miller and Paul’s optimal method [11] (for control portion testing) and Chen’s optimal method [2] (for data portion testing) into a new method that can simultaneously

verify the two portions. In Section 2, the two methods are reviewed. In Section 3, the new method is proposed. In Section 4, the new method is analyzed. In Section 5, our conclusions are presented.

2. Review of Earlier Methods

2.1 Miller and Paul's method for finding a test sequence that checks every transition of the control portion

Consider an example protocol M. The control portion can be modeled as an FSM represented by a *strongly connected*³ digraph $G(V, E)$ of Figure 1, where the vertex set $V = \{1, 2, 3\}$ represents the states, the edge set $E = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$ represents the transitions, and vertex "1" represents the initial state.

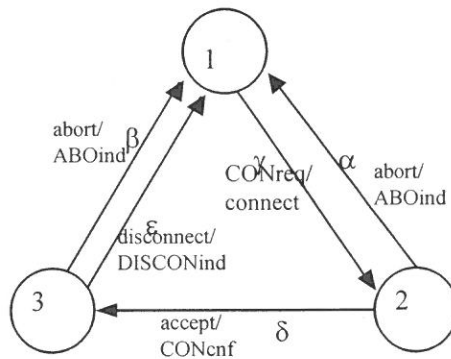


Figure 1. The digraph $G(V, E)$ of an FSM that represents the control portion of Protocol M.

Each transition $q = (J, K; i/o)$ is a transition from the current state J to the next state K , caused by an input "i" and has an output "o". In Figure 1, $\alpha = (2, 1; \text{abort}/\text{ABOind})$ means that an input "abort" will cause the FSM which is at state 2 to generate an output "ABOind" and to change from state 2 to state 1. In $G(V, E)$, a sequence of adjacent edges T_1, T_2, \dots, T_r denoted as $[T_1, T_2, \dots, T_r]$ is called a *path*. A path which starts and ends at the same vertex (in this paper, specifically the initial

³ G is strongly connected if, there is a path from any vertex to any other vertex.

vertex) is called a *tour*. Clearly, a path (or tour) of $G(V, E)$ can be used to generate a test segment. In Figure 1, the path $[\gamma, \delta, \beta]$ is used to generate a test segment $[\text{CONreq/connect}, \text{accept/CONcnf}, \text{abort/ABOind}]$. For simplicity, the path can be called the test segment.

To check a transition $q = (J, K; i/o)$ of the FSM, the implementation is put into state J , input “ i ” is applied, the output is checked to see that it is “ o ”, and the new state is checked to ensure that it is K by $\text{UIO}(K)$; that is, the transition can be checked by a test segment $[q, \text{UIO}(K)]$. $\text{UIO}(K)$, the Unique Input/Output Sequence of state K , is a sequence of input/output pairs which starts only from state K . In Figure 1, $\text{UIO}(1) = [\gamma] = [\text{CONreq/connect}]$ since the input/output pair “CONreq/connect” can start from state 1 but cannot start from state 2 or state 3. In Figure 1, the test segment $[\varepsilon, \text{UIO}(1)] = [\varepsilon, \gamma]$ can be used to check transition ε . We define $\text{Check}(x)$ as the set of transitions that can be checked by test segment x . Thus, $\text{Check}([\varepsilon, \gamma]) = \{\varepsilon\}$.

Miller and Paul prove that if $q = (J, K; i/o)$ is *nonconvergent*⁴ then $[q, \text{UIO}(K)]$ is a $\text{UIO}(J)$. Hence, $[p, q, \text{UIO}(K)] = [p, \text{UIO}(J)]$ can check transition p . That is, $[p, q, \text{UIO}(K)]$ contains overlapping test segments where test segment $[p, q, \text{UIO}(K)]$ checks transition p and test segment $[q, \text{UIO}(K)]$ checks transition q . In general, the path (i.e., test segment) $[T_1, T_2, \dots, T_{(j-1)}, T_j, \text{UIO}(K)]$, where $T_2, \dots, T_{(j-1)}, T_j$ are nonconvergent, can check transitions $T_1, T_2, \dots, T_{(j-1)}$, and T_j . Let test segment $S_1 = [\beta, \gamma, \delta, \varepsilon, \gamma]$ where β is convergent, γ, δ and ε are nonconvergent, and $\text{UIO}(1) = [\varepsilon]$. Then, $\text{Check}(S_1) = \{\beta, \gamma, \delta, \varepsilon\}$. Because $\text{Check}(S_1)$ does not cover all the transitions of the FSM, we find another test segment $S_2 = [\alpha, \text{UIO}(1)] = [\alpha, \gamma]$ (where α ends at state 1), and $\text{Check}(S_2) = [\alpha]$. The two test segments S_1 and S_2 can check all transitions.

Directly connecting these test segments may not form a continuous test sequence. Thus, redundant transitions (i.e., edges) are used to connect the test

⁴ A transition $(J, K; i/o)$ of $G(V, E)$ is convergent if there is a transition $(J', K; i/o)$ where $J' \neq J$. It is nonconvergent if it is not convergent.

segments into a continuous test sequence which starts at the initial state. In Figure 1, redundant segment $[\gamma, \delta]$ is used to connect test segments S_1 and S_2 . The final test sequence is $S = [[\gamma, \delta], S_1, S_2] = [[\gamma, \delta], [\beta, \gamma, \delta, \varepsilon, \gamma], [\alpha, \gamma]] = [\gamma, \delta, \beta, \gamma, \delta, \varepsilon, \gamma, \alpha, \gamma]$. Clearly, $\text{Check}(S) = \text{Check}(S_1) \approx \text{Check}(S_2) = \{\beta, \gamma, \delta, \varepsilon\} \approx \{\alpha\} = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$ which covers all transitions, i.e., the test sequence can check all transitions of FSM M .

2.2 Chen's Method for finding an executable test sequence that checks every rule of the data portion

The data portion of protocol M is specified by the two parameters "quality" and "reason". The parameter "quality" specifies the quality of service in the connection (with "1" indicating a better quality of service than "0"). The parameter "reason", whose value is either "abort" or "ABOind", indicates the reason for aborting the connection. The relations between parameter values are represented by a set of rules $R = \{A, B, C, D\}$. Rules A , B and D mean that for some input/output pairs, the parameter value of the input must be the same as that of its output. Rule C means that during a connection accomplished by the sequence "CONreq(x)/connect(x), accept(y)/CONcnf(y)", the response quality (indicated by the value of y) cannot be better than the request quality (indicated by the value of x). A test sequence with parameter values not violating the rules is called an executable test sequence. For example, [CONreq(0)/connect(0), accept(1)/CONcnf(1), abort(error)/ABOind(error)] is not executable since rule C is violated, because the response quality (1) is better than the requested quality (0).

Rule A: $\text{quality}(\text{CONreq}) = \text{quality}(\text{connect})$ in E_γ

Rule B: $\text{quality}(\text{accept}) = \text{quality}(\text{CONcnf})$ in E_δ

Rule C: $\text{quality}(\text{accept}) \leq \text{quality}(\text{connect})$ in the sequence $[E_\gamma, E_\delta]$

Rule D: $\text{reason}(\text{abort}) = \text{reason}(\text{ABOind})$ in E_β , and in E_α

Remarks: $g(y)$ denotes the value of the parameter g of message y .

The value of parameter quality is either "1" or "0".

The value of parameter reason is either "error" or "shutdown".

Figure 2. The rules that represent the data portion of Protocol M.

Each rule of Figure 2 is checked by a test segment, where feasible parameter values are assigned so that the rule can be verified. For example, either “CONreq(0)/connect(0)” or “CONreq(1)/connect(1)” can be used to check rule A. However, only one of them is needed. “CONreq(1) / connect(1), accept(0) / CONcnf(0)” can be used to check rules A, B, and C, because “CONreq(1)/connect(1)” checks rule A, “accept(0)/CONcnf(0)” checks rule B, and “CONreq(1)/connect(1), accept(0)/CONcnf(0)” checks rule C. Again, either “CONreq(1)/connect(1), accept(1)/CONcnf(1)” or “CONreq(0)/connect(0), accept(0)/CONcnf(0)” can check rules A, B and C; however, only one of them is needed.

Chen’s method involves three steps. First, the transition digraph $G(V, E)$ (Figure 1) and the rules (Figure 2) are converted into a *Behavior Machine Digraph* $G'(V', E')$ (Figure 3); the digraph’s paths can be used to generate test segments with feasible parameter values. Vertex 2 is converted into vertex 20 (which indicates a requested quality “0”) and vertex 21 (which indicates a requested quality “1”). Vertex 3 is converted into vertex 30 (which indicates a response quality “1”) and vertex 31 (which indicates a response quality “1”). The edges are associated with appropriate parameter values and are then mapped into new edges. For example, transition $\gamma = (1, 2; \text{CONreq/connect})$ is mapped into edge $E_{\gamma_1} = (1, 20; \text{CONreq(0)/connect(0)})$ and edge $E_{\gamma_2} = (1, 21; \text{CONreq(1)/connect(1)})$. Notice that transition δ is mapped into edges E_{δ_1} , E_{δ_2} and E_{δ_3} but there is no edges connecting vertex 20 to vertex 31, according to rule C; vertex 20 represents a requested quality “0” and vertex 31 represents a response quality “1”, but the response quality cannot be better than the requested one. Notice that the Behavior Machine does not represent all the possible behaviors of the protocol, but only the behaviors that will be used for testing. Paths of the Behavior Machine Digraph can be used to generate the test segments which check the rules (that is, these paths can check the rules). For example, either $[E_{\gamma_1}]$ or $[E_{\gamma_2}]$ can check rule A, and either $[E_{\gamma_1}, E_{\delta_1}]$ or $[E_{\gamma_2}, E_{\delta_2}]$ or $[E_{\gamma_2}, E_{\delta_3}]$ can check rules A, B and C. We define $\text{check}(x)$ as the set of rules that can be checked by path x . For example, $\text{check}([E_{\gamma_2}, E_{\delta_3}]) = \{A, B, C\}$.

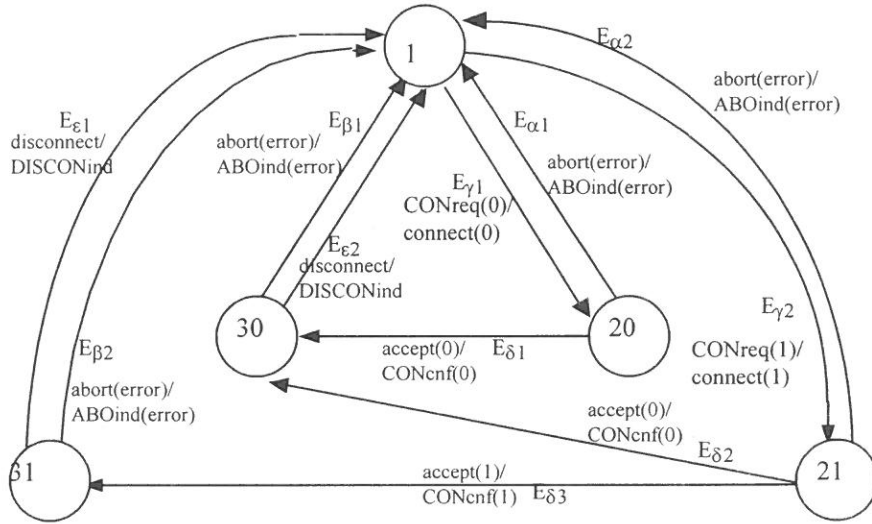


Figure 3. The behavior machine digraph $G'(V', E')$ converted from the digraph $G(V, E)$ of Figure 1 and the rules of Figure 2.

Second, Each edge x of Figure 3 (which contains at least two input/output pairs), used to check rule y , is labeled with “ y ” in Figure 4 (the Selecting Digraph $G''(V'', E'')$), indicating that it can be used to check rule y . For example, in Figure 3, edges $E_{\gamma 1}$ and $E_{\gamma 2}$, which can check rule A, are labeled with “A” in Figure 4. Similarly, edges $E_{\beta 1}$ and $E_{\beta 2}$, which can check rule D, are labeled with “D” in Figure 4.

Each multi-edged path x of Figure 3 (which can check rule y) is converted into a *pseudo edge* $(J, K; x)$ of Figure 4, where J and K are the starting and ending vertices respectively of path x . Traversing edge $(J, K; x)$ of Figure 4 is equivalent to traversing path x of Figure 3. The pseudo edge is labeled with “ y ”, indicating that it can check rule y . Consider the example path $[E_{\gamma 2}, E_{\delta 2}]$ of Figure 3. The path starts at vertex 1 and ends at vertex 30. Thus, we create a pseudo edge $(1, 30; [E_{\gamma 2}, E_{\delta 2}])$. As described earlier, $[E_{\gamma 2}, E_{\delta 2}] = \text{“CONreq(1)/connect(1), accept(0)/CONcnf(0)”}$ can check rules A, B and C. Thus, the new edge $(1, 30; [E_{\gamma 2}, E_{\delta 2}])$ is labeled with “A”, “B”, and “C”.

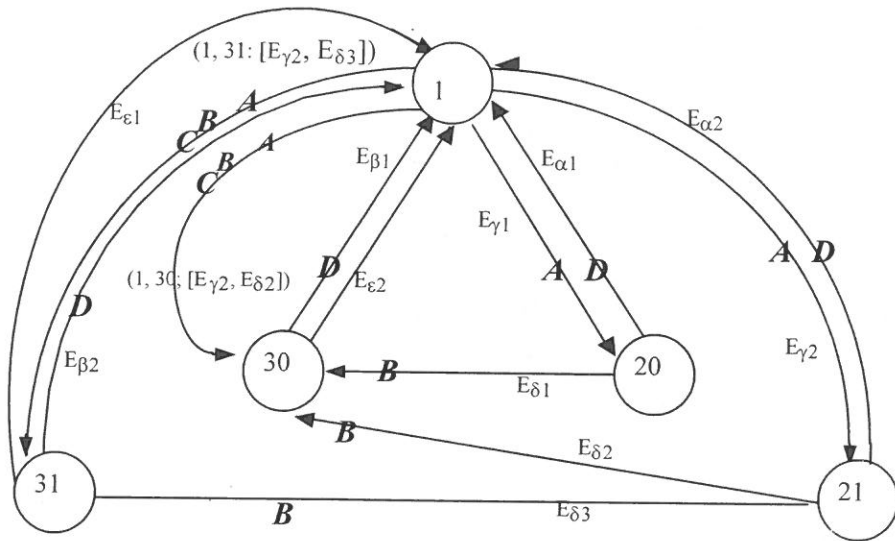


Figure 4. The Selecting Digraph $G''(V'', E'')$ constructed from the Behavior Machine Digraph $G'(V', E')$ of Figure 3 and the rules of Figure 2 (the labels are shown in bold faces.)

Third, we let these labels form a set Σ . In Figure 4, $\Sigma = \{A, B, C, D\}$ covers all the rules of the set R of Figure 2. A Selecting Chinese Postman Tour of $G''(V'', E'')$ is a minimum-length tour of $G''(V'', E'')$ where each label of Σ appears on the tour at least once. Such a tour is used to generate an executable test sequence that checks all rules. The length of the pseudo edge $(J, K; x)$ is defined as the length of the path x , and that of the other edges is 1. In Figure 4, the Selecting Chinese Postman algorithm proposed in [2] computes the Selecting Chinese Postman Tour $[(1, 31: [E_{\gamma 2}, E_{\delta 3}]), E_{\beta 2}]$, which produces the test sequence $[E_{\gamma 2}, E_{\delta 3}, E_{\beta 2}]$ where $[E_{\gamma 2}, E_{\delta 3}]$ checks rules A, B and C, and $[E_{\beta 2}]$ checks rule D.

3. Executable Test Sequence for the Control and Data Portions

In Miller and Paul's method (Section 2.1), a specific tour of the transition digraph $G(V, E)$ has been used to find a test sequence that checks all transitions. In Chen's method (Section 2.2), a Selecting Chinese Postman Tour of the Selecting Digraph $G''(V'', E'')$ has been used to find an executable test sequence that checks all

rules. In this section, we will construct a SelectOverlap (SelectO) Digraph $G^*(V^*, E^*)$, the Selecting Chinese Postman Tour of which can be used to find a test sequence that checks all the transitions and rules.

However, constructing a SelectO Digraph involves complex concepts. Thus, in Section 3.1, we first modify Miller and Paul's method. The modified one obtains similar result but helps us to see a basic concept of constructing a SelectO Digraph. Then, in Section 3.2, similar modification process is applied to convert the Selecting Digraph $G''(V'', E'')$ into a SelectO Digraph $G^*(V^*, E^*)$.

3.1 Modification of Miller and Paul's method

In order to modify Miller and Paul's method, we first observe the form of the test sequence obtained. Recall that Miller and Paul first find the test segment $[T_1, T_2, \dots, T_{(j-1)}, T_j, \text{UIO}(K)]$, where $T_2, \dots, T_{(j-1)}, T_j$ are nonconvergent for check transitions $T_1, T_2, \dots, T_{(j-1)}$, and T_j . We want to modify the transition digraph $G(V, E)$ into an Overlap Digraph $G^\wedge(V^\wedge, E^\wedge)$, a Selecting Chinese Postman Tour of which can be used to check all the transitions. Clearly, a nonconvergent transition of $G(V, E)$ does not need to be modified, because it can play the role of either " T_2 " or " T_3 ", ..., or " T_j ". We simply label the edge with the transition symbol, indicating that it can be the first edge of a test segment which checks the transition. A convergent transition of $G(V, E)$ needs to be transformed into a new edge (which is a UIO sequence) so as to play the role of $\text{UIO}(K)$. Notice that the last edge of $\text{UIO}(K)$ can be the first edge of another sequence $[T_1', T_2', \dots, T_{(j-1)}', T_j', \text{UIO}(K')]$. That is, it can play the role of " T_1 ", so we label the edge with the transition symbol.

The Transition Digraph $G(V, E)$ of Figure 1 is converted into the Overlap Digraph $G^\wedge(V^\wedge, E^\wedge)$ of Figure 5. Each transition $q = (J, K; i/o)$ is converted into a new edge as follows:

Case I. q is nonconvergent:

Transition (i.e., edge) q is labeled with " q ", indicating that it can be the first

edge of a test segment which checks transition q . For example, the nonconvergent transitions δ , γ and ϵ of Figure 1 are labeled with δ , γ and ϵ respectively.

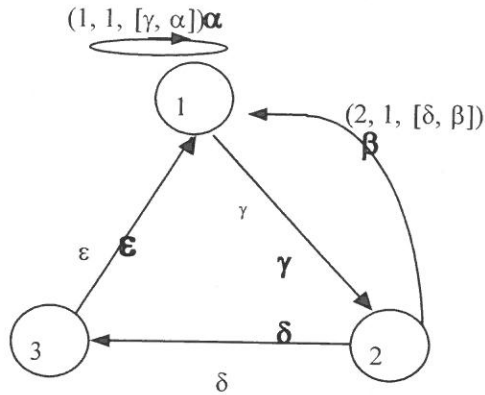


Figure 5. An Overlap Digraph $G^{(V^, E^)}$ converted from the transition digraph $G(V, E)$ of Figure 1 (labels are bold-faced.)

Case II. q is convergent.

In Figure 1, we backwardly search for a path $[q_1, q_2, \dots, q_x, q]$ which is a UIO sequence, which starts from vertex L . We then remove edge q and create a pseudo edge $(L, K; [q_1, q_2, \dots, q_x, q])$. The path must satisfies a special property⁵. Because in a tour of Figure 5, the pseudo edge will be either followed by an edge which represents a UIO sequence, or by a sequence of nonconvergent edges and then a UIO sequence (these nonconvergent edges and the UIO sequence also form a UIO sequence, as stated earlier.) That is, the last edge of the path (i.e., “ q ”) will be the first edge of a test segment which checks transition q . Thus, the pseudo edge is $(L, K; [q_1, q_2, \dots, q_x, q])$ is labeled with “ q ”. For example, $\alpha = (2, 1; \text{abort/ABOind})$ is convergent. We backwardly search to find a path $[\gamma, \alpha]$, which is a UIO sequence of state 1. We then remove edge α and create the new edge $(1, 1; [\gamma, \alpha])$, and we label the edge with “ α ”, indicating that α can be the first edge of a test segment which

⁵ The path must satisfy a property that removing edge q while creating an edge connecting vertex L to vertex K would not cause the digraph to become not strongly connected. Such a path can be found because in the worst case, we back to vertex J .

checks transition α .

We let these labels form a set Σ . We then find a Selecting Chinese Postman Tour of $G^{\wedge}(V^{\wedge}, E^{\wedge})$ with $\Sigma = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$. The tour is $[(1, 1; [\gamma, \alpha]), \gamma, ([2, 1; [\delta, \beta]), \gamma, \delta, \varepsilon] = [\gamma, \alpha, \gamma, \delta, \beta, \gamma, \delta, \varepsilon]$. Because the last transition of the tour needed to be checked, we need a UIO(1) at its end. Thus, In Figure 1, UIO(1) = $[\gamma]$ is added to the end of $[\gamma, \alpha, \gamma, \delta, \beta, \gamma, \delta, \varepsilon]$. The final test sequence is shown in Figure 6.

Tour+UIO(1)	$[(1, 1; [\gamma, \alpha]), \gamma, ([2, 1; [\delta, \beta]), \gamma, \delta, \varepsilon], \gamma$				
remarks	uio seq	nonconv	uio seq	nonconver	uio seq
Test Sequence	γ, α	$\gamma,$	$\delta, \beta,$	$\gamma, \delta, \varepsilon,$	γ
Overlapping test segments					
	α	$\gamma,$	δ, β	checks transition α	
		$\gamma,$	δ, β	checks transition γ	
			$\beta,$	$\gamma, \delta, \varepsilon,$	γ checks transition β
				$\gamma, \delta, \varepsilon,$	γ checks transition γ
				$\delta, \varepsilon,$	γ checks transition δ
				$\varepsilon,$	γ checks transition ε

Figure 6. A test sequence constructed from the Overlap Digraph $G^{\wedge}(V^{\wedge}, E^{\wedge})$ of Figure 5 for checking every transition of the FSM of Figure 1.

3.2 The proposed method

In the Selecting Digraph $G''(V'', E'')$ of Figure 4, edges are labeled with rules to indicate that these edges can be used to check these rules. In the Overlap Digraph $G^{\wedge}(V^{\wedge}, E^{\wedge})$ of Figure 5, edges are labeled with transition symbols to indicate that these edges can be used to check these transitions. We will modify the Selecting Digraph $G''(V'', E'')$ and add symbols (which represent the checking of transitions) to the edges, resulting in a SelectO digraph $G^*(V^*, E^*)$ of Figure 7. The Selecting Chinese Postman Tour of $G^*(V^*, E^*)$ (with UIO(1) added to the end) can be used to check all transitions and rules.

We first give some notation. Notice that every edge of Figure 4 is converted

from the FSM of Figure 1. An edge E_p of Figure 4 may represent a transition of Figure 1 (namely, $\text{Original}(E_p)$) and is called a *simple edge*. For example, the simple edge $E_{\delta 2}$ of Figure 4 represents transition δ of Figure 1, and $\text{Original}(E_{\delta 2}) = \delta$. Similarly, $\text{Original}(E_{\delta 1}) = \text{Original}(E_{\delta 3}) = \delta$. An edge E_p of Figure 4 may be a pseudo edge $q = (J, K; [E_1, E_2, \dots, E_x])$ which represents consecutive transitions E_1, E_2, \dots and E_x and $\text{Original}(q) = [\text{Original}(E_1), \text{Original}(E_2), \dots, \text{Original}(E_x)]$. For example, in Figure 4, edge $(1, 31; [E_{\gamma 2}, E_{\delta 3}])$ is a pseudo edge. $\text{Original}((1, 31; [E_{\gamma 2}, E_{\delta 3}])) = [\text{Original}(E_{\gamma 2}), \text{Original}(E_{\delta 3})] = [\gamma, \delta]$.

Each edge E_p of Figure 4 is converted into a new edge of Figure 7 as follows:

Case I. $\text{Original}(E_p)$ is nonconvergent:

Similar to the discussion in the earlier subsection, because $\text{Original}(E_p)$ is nonconvergent, it can be the first edge of a test segment which checks transition $\text{Original}(E_p)$. Thus, we simply add the label $\text{Original}(E_p)$ to the edge. For example, in Figure 4, edge $E_{\gamma 1}$ where $\text{Original}(E_{\gamma}) = \gamma$ is nonconvergent. So, in Figure 7, we add label “ γ ” to edge $E_{\gamma 1}$. Because on the edge $E_{\gamma 1}$ of Figure 4, the label “ A ” already exists there, so edge $E_{\gamma 1}$ has two labels: “ γ ” and “ A ”, indicating that it can be the first edge of a test segment which checks transition “ γ ”, and can be a test segment which checks rule A . That is, test segments for checking the rules and transitions are overlapped. Similarly, we add label “ γ ” to edge $E_{\gamma 2}$, add label “ δ ” to edges $E_{\delta 1}$, $E_{\delta 2}$ and $E_{\delta 3}$, and add label “ ϵ ” to edges $E_{\epsilon 1}$ and $E_{\epsilon 2}$.

Case II. $\text{Original}(E_p)$ is convergent:

We backwardly search the digraph $G''(V'', E'')$ of Figure 4 from edge E_p , looking for a UIO path $[E_1, E_2, \dots, E_x, E_p]$. Suppose that such a UIO path starts at vertex L and ends at vertex K (which must satisfy the property as described in Case II of the earlier subsection). We delete edge E_p and create a new edge $E_r = (L, K; [E_1, E_2, \dots, E_x, E_p])$ (namely, a *ghost edge*). Similar to Case II of Section 3.1, we add label $\text{Original}(E_p)$ to the edge because E_p can become the first edge of a test segment that

checks the transition $\text{Original}(E_p)$. Moreover, because $[E_1, E_2, \dots, E_x, E_p]$ can check some rules, those rules must be labeled on it as well. For example, consider the convergent edge $E_{\alpha 1} = (20, 1; \text{abort}(\text{error})/\text{ABOind}(\text{error}))$. We backwardly search from edge $E_{\alpha 1}$ to find a UIO sequence $[E_{\gamma 1}, E_{\alpha 1}]$. Edge $E_{\alpha 1}$ is deleted, and the new edge $(1, 1; [E_{\gamma 1}, E_{\alpha 1}])$ is created. We add $\text{Original}(E_{\alpha 1}) = \alpha$ to the edge, indicating that it can be first edge of a test segment which checks transition α . Because $[E_{\gamma 1}, E_{\alpha 1}]$ can check rules A and D, edge $E_{\alpha 1}$ is labeled with α, A, D , indicating that it can be used to check transition α and rules A and D. Similarly, the convergent edges $E_{\alpha 2}, E_{\beta 1}, E_{\beta 2}$ are replaced by ghost edges $(1, 1; [E_{\gamma 2}, E_{\alpha 2}]), (20, 1; [E_{\delta 1}, E_{\beta 1}])$ and $(21, 1; [E_{\delta 3}, E_{\beta 2}])$ respectively. They are labeled with $\{\alpha, A, D\}, \{\beta, B, D\}$ and $\{\delta, B\}$ respectively.

Case III. E_p is a pseudo edge $(J, K; E_1, E_2, \dots, E_x)$ and $\text{Original}(E_1), \text{Original}(E_2), \dots, \text{Original}(E_x)$ are nonconvergent.

This is similar to Case I, in that E_1, E_2, \dots, E_x can serve as the first edge of test segments which check transition $\text{Original}(E_1), \text{Original}(E_2), \dots$, and $\text{Original}(E_x)$ respectively. For example, in Figure 3, edge $(1, 31; [E_{\gamma 2}, E_{\delta 3}])$ is a pseudo edge where both edges $E_{\gamma 2}$ and $E_{\delta 3}$ are nonconvergent. Hence, we add labels γ, δ to the edge, forming a new set of labels label-set $\{\gamma, \delta, A, B, C\}$ in Figure 7.

Case IV. E_p is a pseudo edge $(J, K; [E_1, E_2, \dots, E_x])$ and either $\text{Original}(E_1)$ or $\text{Original}(E_2), \dots$, or $\text{Original}(E_x)$ is convergent.

This is similar to Case II. Thus, we delete edge E_p and create a *pseudo-ghost* edge $E_q = (L, K; [E_1', E_2', \dots, E_{\gamma'}, E_1, E_2, \dots, E_x])$. Because E_x can be the first edge of a test segment which verifies transition $\text{Original}(E_x)$, we add the symbol $\text{Original}(E_x)$ to the edge. No edges in Figure 3 conform to this case.

We let these labels of $G^*(V^*, E^*)$ form a set $\Sigma = \{\alpha, \beta, \gamma, \delta, \varepsilon, A, B, C, D\}$. Then, we compute the Selecting Chinese Postman tour of $G^{\wedge}(V^{\wedge}, E^{\wedge})$ where each label of Σ appeared on the tour at least once. In Figure 7, $[E_{\gamma 2}, (21, 1; [E_{\delta 3}, E_{\beta 2}]), (1,$

$1: [E_{\gamma_2}, E_{\alpha_2}]), (1, 31: [E_{\gamma_1}, E_{\delta_3}]), E_{\epsilon_1}]$ is such a tour which starts from vertex 1. By appending $UIO(1)$ at the end of the tour, we have a path $[E_{\gamma_2}, (21, 1: [E_{\delta_3}, E_{\beta_2}]), (1, 1: [E_{\gamma_2}, E_{\alpha_2}]), (1, 31: [E_{\gamma_1}, E_{\delta_3}]), E_{\epsilon_1}, E_{\gamma_1}]$ (where E_{γ_1} is the UIO path), which is used to generate the executable test sequence $[E_{\gamma_2}, E_{\delta_3}, E_{\beta_2}, E_{\gamma_2}, E_{\alpha_2}, E_{\gamma_1}, E_{\delta_3}, E_{\epsilon_1}, E_{\gamma_1}] = [\text{CONreq}(1) / \text{connect}(1), \text{accept}(1) / \text{CONcnf}(1), \text{abort} / \text{ABOind}, \text{CONreq}(1) / \text{accept}(1), \text{abort}(\text{error}) / \text{ABOind}(\text{error}), \text{CONreq}(0) / \text{connect}(0), \text{accept}(1) / \text{CONcnf}(1), \text{disconnect} / \text{DISCONind}, \text{CONreq}(0) / \text{connect}(0)]$. In Figure 8, each edge labeled with rule x is used to generate a test segment which checks rule x . For example, edge $(21, 1: [E_{\delta_3}, E_{\beta_2}])$ which is labeled with “B, D” is used to generate a test segment $[E_{\delta_3}, E_{\beta_2}]$ which checks rules B and D. Each path $[y, z]$ where y is labeled with transition symbol “ y ” is used to generate a test segment which checks transition y . For example, the path $[E_{\gamma_2}, (21, 1: [E_{\delta_3}, E_{\beta_2}])]$ where $[E_{\gamma_2}]$ labeled with “ γ ” is used to generate a test segment $[E_{\gamma_2}, E_{\delta_3}, E_{\beta_2}]$ which checks transition γ .

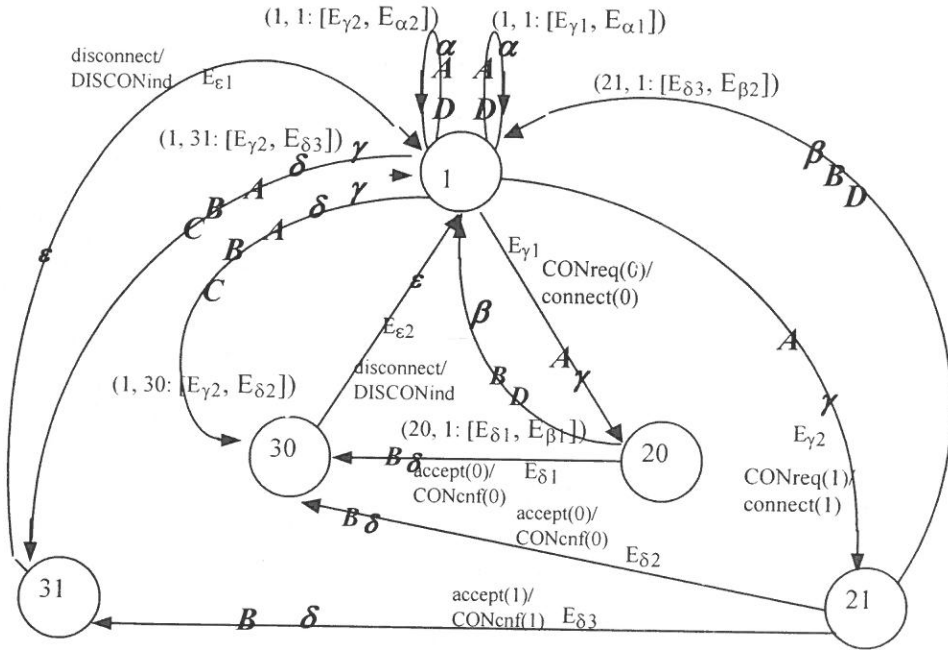


Figure 7. The SelectO Digraph $G^*(V^*, E^*)$ constructed from the Selecting digraph $G'(V', E')$ of Figure 3.

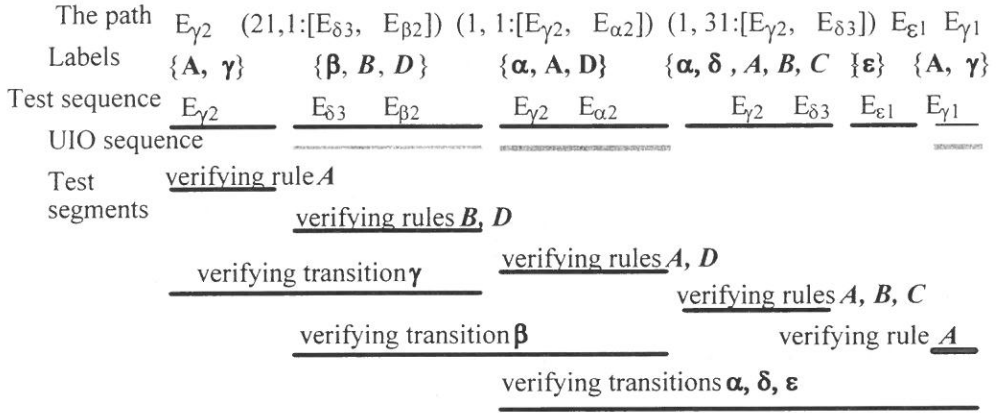


Figure 8. Overlapping test segments for the executable test sequence generated from the Selecting Chinese Postman tour of the SelectO digraph in Figure 7.

4. Analysis of the Proposed Method

4.1 Correctness proof

In Section 3, we have obtained a Selecting Chinese Postman Tour $[E_1, E_2, \dots, E_r, E_{r+1}, \dots, E_{s-1}, E_s]$ from the SelectO digraph $G^*(V^*, E^*)$, which then becomes a path $Z = [E_1, E_2, \dots, E_r, E_{r+1}, \dots, E_{s-1}, E_s, \text{UIO}(1)]$, for generating an executable test sequence that checks all the rules and transitions specified in the set Σ . We are going to prove that the set covers all the transition and rule symbols, so as to prove that the test sequence generated from such a tour can check all the transitions and rules.

Lemma 1 $\Sigma \supset R$.

Proof

Consider the conversion process described in Section 3.2. In Case I and Case III, each edge and its labels of $G'(V', E')$ are directly copied from the Selecting Digraph to the SelectO Digraph $G^*(V^*, E^*)$. In Case II and Case IV, each edge E_p of $G'(V', E')$ is mapped into a ghost edge E_r of $G^*(V, E^*)$ which represents a path $[E_1, E_2, \dots, E_x, E_p]$ of $G'(V', E')$; labels that originally appear on edge of E_p is then copied to the edge of E_r . That is, the labels appeared on $G'(V', E')$ form a set which is a subset of

the labels that appeared on $G^*(V^*, E^*)$. Because the labels appeared on $G'(V', E')$ form a set R , R is a subset of the labels appeared in $G^*(V^*, E^*)$. We thus have: $\Sigma \supset R$.

Lemma 2 $\Sigma \supset E$.

Proof

During the conversion from transition digraph $G(V, E)$ to the Selecting Digraph $G^*(V^*, E^*)$, no edges are deleted but only new edges are added. Thus, each edge E_x corresponds to an edge E_y , where $\text{Original}(E_y)$ is either nonconvergent or convergent. Consider the conversion from $G'(V', E')$ to $G^*(V^*, E^*)$ (Section 3.2). If $\text{Original}(E_y)$ is nonconvergent, it will be labeled on the same edge. And If $\text{Original}(E_y)$ is convergent, a new edge is created, but $\text{Original}(E_y)$ will be labeled on it too. That is, all the transitions of $G(V, E)$ will be labeled on the edges of $G^*(V^*, E^*)$. We thus have: $\Sigma \supset E$.

From Lemma 1 and Lemma 2, we know $\Sigma \supset E \cup R$. Because no other labels other than those in $E \approx R$ have been used to label $G^*(V^*, E^*)$, we have $E \approx R \supset \Sigma$. We thus have:

Theorem 1 $\Sigma = E \approx R$.

4.2 Complexity analysis

The Selecting Chinese Postman tour of the SelectO digraph can be used to generate an executable test sequence verifying the transitions and rules. The complexity of converting the Selecting digraph to the SelectO digraph is dominated by the process in computing the backward breadth-first-search tree for the convergent edge and the bold edge which includes convergent edges (assume there are j such edges). The breadth-first-search needs $O(2^j)$, checking the property of UIO sequence needs $O(n)$ (assume the FSM has n states), and checking the property of strongly-connectness needs $O(m)$ (assume the FSM has m edges) [10]. Hence, the process needs $O(j * (2^j * (n+m)))$. Fortunately, the UIO sequence can be computed

mostly in 2 to 3 input/output pairs [12], and hence l is not very large.

The Selecting Chinese Postman algorithm involves three phases [2]. The first phase replicates or deletes edges of $G''(V'', E'')$ to result in a new digraph where each vertex has the same number of incoming and outgoing edges, and each label of Σ appears in the new digraph at least once. The second phase involves using a tour to connect the new digraph into an Euler digraph, if new digraph is a set of connected components. The third phase computes the Euler tour of the Euler digraph. If no tours is needed in the second phase, then the tour in the third phase is the optimal tour. Otherwise, the cost of the tour is larger than the optimal cost by the extra cost used in the second phase. The second and the third phase need $O(|E''|)$ and the complexity is dominated by the first phase. The first phase uses an integer programming approach to find a solution. If the problem size is not very large, the LINDO package can compute the optimal solution very quickly [15]. Otherwise, we can construct an initial solution based on Kwast's result for the data portion [8][9] and Miller's result [11] for the control portion. The branch and bound algorithm for the integer programming will iterate to improve the initial solution, which guarantee an improved result within an acceptable time limit.

4.3 Experimentation

We experiment our method on the Simple Session Protocol [8][9] (Figure 9). In [8], the protocol has 8 rules, but in [2] only 7 rules are used. We use the 8 rules of [8] and extend it into the 16 rule of Figure 8 that can completely test the data portion properties. For example, for the rule that the value of the parameter of "Abort/SUABin" must be the same, we convert into the 8 rules which considers the pair's starting state. By using the method of [11][14], we obtain for the control portion an executable test sequence that involves 37 input/output pairs. By using the method of [2], we obtain for the data portion an executable test sequence that involves 28 input/output pairs. By using the method of this paper, we obtain for both portions an executable test sequence that involves only 41 input/output pairs and a 37% saving is

achieved. The integer programming equations have been solved by the Lingo on a PC in less than one second [15].

- R1: $\text{qual-of-serv}(\text{Connect}) = \text{qual-of-serv}(\text{SCONreq})$ in the transition (1, 2; SCONreq/connect)
- R2: If $\text{qual-of-serv}(\text{Connect}) = 0$ or \wedge , then $\text{qual-of-serv}(\text{Accept}) = 0$ in the path [SCONreq/connect Accept/SCONcnf(+)] from state 1.
- R3: $\text{qual-of-serv}(\text{SCONcnf}) = \text{qual-of-serv}(\text{Accept})$ in (2, 1; Accept/SCONcnf(+))
- R4A: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (2, 1; Abort/SUABind)
- R4B: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (2, 1; SUABind/Abort)
- R4C: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (4, 1; Abort/SUABind)
- R4D: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (4, 1; SUABind/Abort)
- R4E: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (6, 1; Abort/SUABind)
- R4F: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (6, 1; SUABind/Abort)
- R4G: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (1, 1; SUABind/Abort)
- R4H: $\text{reason}(\text{Abort}) = \text{reason}(\text{SUABind})$ in (1, 1; Abort/SUABind)
- R5A: If $\text{qual-of-serv}(\text{Accept}) = 0$ then $\text{reason}(\text{Nonfinish}) = \wedge$, and $\text{reason}(\text{SRELcnf}) = \wedge$ in the path [Accept/SCONcnf(+), SRELreq/Finish, Nofinish/SRELcnf(-)] starting from state 2.
- R5B: If $\text{qual-of-serv}(\text{Accept}) = 1$ then $\text{reason}(\text{Nonfinish}) \neq \wedge$ in the path [Accept/SCONcnf(+), SRELreq/Finish, Nofinish/SRELcnf(-)] starting from state 2.
- R6: $\text{reason}(\text{SRELcnf}) = \text{reason}(\text{Nofinish})$ in (6, 4; Nofinish/SRELcnf(-))
- R7: If $\text{result}(\text{SCONcnf}) = ??$ then $\text{qual-of-serv}(\text{SCONcnf}) = \wedge$ in (2, 1; Refuse/SCONcnf(-))
- R8: If $\text{result}(\text{SRELcnf}) = ??$ then $\text{reason}(\text{SRELcnf}) = \wedge$ in (6, 4; Nofinish/SRELcnf(-)).

Figure 9. The rules for the data portion of the Simple Session Protocol.

5. Conclusions

In this paper, we have proposed a method to automatically generate the executable test sequence for both the control and the data portions of the protocol. In contrast to testing data portion using the optimal method of [2] and testing the control portion using the optimal method of [11] separately, using our combined approach achieves a 37% saving in test sequence length.

In protocol testing, the test sequence length may not completely reflect the testing time, because some operations may take longer testing time than others. However, our method of minimizing the test sequence length can be easily extended to minimizing the test sequence cost, by assigning cost to the edges of the digraph. In a remote testing system that involves the upper and lower testers, a synchronization problem may be encountered. And the current proposed method can be combined with the duplexE digraph method to generate a synchronizable and executable test sequence [4].

References

- [1] A. V. Aho, A. T. Dahbura, D. Lee and M. U. Uyar, "An optimization technique for protocol conformance test generation based on UIO sequences and Rural Chinese Postman tour," *Proc. IFIP 8th Int'l Workshop on Protocol Specification, Testing and Verification*, 1988.
- [2] W. H. Chen, "Test sequence generation from the protocol data portion based on the Selecting Chinese Postman algorithm," *Information Processing Letters*, Vol. 6, 1998.
- [3] W. H. Chen, C. S. Lu, E. R. Brozovsky and J. T. Wang, "An optimization technique for protocol conformance testing using multiple UIO sequences," *Information Processing Letters*, Vol. 36, 1990.
- [4] W. H. Chen and H. Ural, "Synchronizable test sequence based on multiple UIO sequences," *IEEE/ACM Trans. on Networking*, Vol. 3, No. 2, 1995, pp. 152- 157.
- [5] K. T. Cheng and A. S. Krishnakumar, "Automatic functional test generation using the extended finite state machine model," *Proc. IEEE Design Automation Conference*, 1993.
- [6] S. T. Chanson and J. Zhu, "A unified approach for protocol test sequence generation," *INFOCOM, IEEE*, 1993.
- [7] C. M. Huang, M. S. Chiang, M. Y. Jang, "UIO_E: a protocol test sequence generation method using the transition executability analysis (TEA)," *Computer Communications*, Vol. 21, 1998.
- [8] E. Kwast, "Toward automatic test generation for protocol data aspects," *Proc. IFIP Int'l Symp. on Protocol Specification, Testing, and Verification*, 1991.
- [9] E. Kwast, "Automatic test generation for protocol data aspects," *Proc. IFIP Int'l Symp. on Protocol Specification, Testing, and Verification*, 1992.
- [10] J. A. Mcbugh, *Algorithmic Graph Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1990.
- [11] R. E. Miller and S. Paul, "On the generation of minimal-length conformance tests for communication protocols," *IEEE/ACM Trans. on Networking*, Vol. 1, No. 1, 1993.

- [12] K. K. Sabnani and A. T. Dahbura, "A protocol test generation procedure," *Computer Networks and ISDN Systems*, Vol. 15, 1988.
- [13] B. Sarikaya and G. v. Bochmann, "Synchronization and specification issues in protocol testing," *IEEE Trans. on Communications*, Vol. 40, 1984.
- [14] B. Sarikaya, G. v. Bochmann and E. Ceny, "A test design methodology for protocol testing," *IEEE Trans. on Software Engineering*, Vol. 13, pp. 518-531, 1987.
- [15] L. Schrage, *LINDO 5.0 User's Manual*, Scientific Press., 1991.
- [16] H. Ural, "Test sequence selection based on static data flow analysis," *Computer Communications*, Vol. 10, No. 5, 1987.
- [17] H. Ural and B. Yang, "A test sequence selection method for protocol testing," *IEEE Trans. on Communications*, Vol. 9, No. 4, 1991.
- [18] R. Dssouli, K. Saleh, E. Aboulhamid, A. En-Nouaary and C. Bourhfir, "Test development for communication protocols: towards automation," *Computer Networks*, Vol. 31, 1999.
- [19] D. Lee and M. Yannakakis, "Principles and methods of testing finite state machines -- a survey," *Proceedings of the IEEE*, Vol. 84, No. 8, pp. 1089-1123, 1996.
- [20] M. U. Uyar and A. T. Dahbura, "Optimal test sequence generation for protocols: the Chinese postman algorithm applied to Q.931," *Proc. IEEE Global Telecommunication Conference*, 1987.
- [21] D. Sidhu and T. Leung, "Experience with test generation for real protocols," *Computer Communications*, Vol. 10, No. 2, 1987.
- [22] Z. Kohavi, *Switching and Finite Automata Theory*, New York, McGraw-Hill, 1978.

received October 31, 2002
revised November 23, 2002
accepted December 15, 2002

通訊協定控制與資料部分可執行測試序列產生方法之研究

陳文輝*

輔仁大學電子工程系

本篇論文提出一個新的可執行測試序列產生方法來測試通訊協定的實作產品是否與它的規格相符合。這個測試序列同時測試通訊協定（表現為一個有限狀態機）和資料部分（表現為一些規則）。這個方法是將有限狀態機跟這些規則合成為一個叫做 SelectO 的圖，圖上的任何路徑都可以產生可執行測試序列。選擇式中國郵差演算法最後被運用來找出圖上的一條最佳路徑，這個路徑產生的可執行測試序列能夠測試有限狀態機的每一個轉移（transition）和每一條規則至少一次。針對一個簡單會議層通訊協定的實驗顯示本方法能夠降低測試序列的長度 37%。