

ISSN: 0128-7680

Pertanika Journal of

SCIENCE & TECHNOLOGY

VOLUME 13 NO.2
JULY 2005



A scientific journal published by Universiti Putra Malaysia Press

Pertanika Journal of Science & Technology

■ About the Journal

Pertanika, the pioneer journal of UPM, began publication in 1978. Since then, it has established itself as one of the leading multidisciplinary journals in the tropics. In 1992, a decision was made to streamline Pertanika into three journals to meet the need for specialised journals in areas of study aligned with the strengths of the university. These are (i) **Pertanika Journal of Tropical Agricultural Science** (ii) **Pertanika Journal of Science & Technology** (iii) **Pertanika Journal of Social Science & Humanities**.

■ Aims and Scope

Pertanika Journal of Science & Technology welcomes full papers and short communications in English or Bahasa Melayu in the fields of chemistry, physics, mathematics, and statistics, engineering, environmental control and management, ecology and computer science. It is published **twice a year in January and July**.

Articles must be reports of research not previously or simultaneously published in other scientific or technical journals.

Communications are notes of a significant finding intended spaced typewritten pages and must be accompanied by a letter from the author justifying its publication as a communication.

Reviews are critical appraisals of literature in areas that are of interest to a broad spectrum of scientist and researchers. Review papers will be published upon invitation.

■ Submission of Manuscript

Three complete clear copies of the manuscript are to be submitted to

The Chief Editor
Pertanika Journal of Science & Technology
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor Darul Ehsan
MALAYSIA
Tel: 03-89468854; Fax: 03-89416172

■ Proofs and Offprints

Page proofs, illustration proofs and the copy-edited manuscript will be sent to the author. Proofs must be checked very carefully within the specified time as they will not be proofread by the Press editors.

Authors will receive 20 offprints of each article and a copy of the journal. Additional copies can be ordered from the Secretary of the Editorial Board.

EDITORIAL BOARD

Prof. Ir. Abang Abdullah Abang Ali- *Chief Editor*
Faculty of Engineering

Assoc. Prof. Ir. Dr. Norman Mariun
Faculty of Engineering

Assoc. Prof. Ir. Dr. Mohd. Saleh Jaafar
Faculty of Engineering

Assoc. Prof. Dr. Gwendoline Ee Cheng Lian
Faculty of Science

Prof. Dr. Abu Bakar Salleh
Faculty of Biotechnology & Biomolecular Sciences

Prof. Dr. W. Mahmood Mat Yunus
Faculty of Science

Assoc. Prof. Dr. Noor Akma Ibrahim
Faculty of Environmental Studies

Assoc. Prof. Dr. Hamidah Ibrahim
Faculty of Information Technology & Science
Computer

Rosta Harun
Faculty of Environmental Studies

Sumangala Pillai - *Secretary*
Universiti Putra Malaysia Press

Published by Universiti Putra Malaysia Press
ISSN No. 0128-7680

INTERNATIONAL PANEL MEMBERS

Prof. D.J Evans
Parallel Algorithms Research Centre

Prof. F. Halsall
University College of Swansea

Prof. S.B Palmer
University of Warrick

Prof. Dr. Jerry L. Mc Laughlin
Purdue University

Prof. Dr. John Loxton
MaxQuarie University

Prof. U.A Th. Brinkman
Vrije Universiteit

Prof. A.P. Cracknell
University of Dundee

Prof. A.J. Saul
University of Sheffield

Prof. Robert M. Peat
University of Florida

Prof. J.N Bell
Imperial College of Science, Technology and Medicine

Prof. Yadolah Dodge
University De Neuchatel

Prof. W.E Jones
University of Windsor

Prof. A.K. Kochar
UMIST

Pertanika Journal of Science & Technology

Volume 13 No. 2, 2005

Contents

- The Application of Water Cooling on Reducing NO_x from a Gas Burner System – *Mohammad Nazri Mohd. Jaafar, G.E. Andrews & M.C. Mkpadi* 139
- An Experimental Study on Solar Flat Plate Collector Using an Alternative Working Fluid – *A. Manickavasagan, A. Sampathrajan, P.R. Manigandan, S. Manivannan, F. Magdaline Eljeeva Emerald, R. Malarkodi & S. Mangalam* 147
- Thermal Diffusivity Measurement of La Doped SrSnO_3 ($\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$) Ceramics by Photoflash Method – *Josephine L.Y.C., W. Mahmood Mat Yunus, Iftetan A. Taha & Abdul Halim Shaari* 163
- Location Management Cost Reduction Using Adaptive Velocity-movement Based Scheme – *M.S. Gembari, M.H. Habaebi, N.K. Noordin, B.M. Ali & V. Prakash* 173
- A New Congestion Control Algorithm for Active Networks – *M.I. Buhari, M.H. Habaebi & B.M. Ali* 187
- Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer Overlaid by Dense Sand Layer – *Abdulhafiz O. Al-Shenawy & Awad A. Al-Karni* 213
- Modelling of Changes in Evapotranspiration for an Area in Peninsular Malaysia – *Lee Teang Shui, M.M.M. Najim, M. Aminul Haque & Huang Yuk Feng* 237
- Bioaccumulation of Carbofuran and Endosulfan in the African Catfish *Clarias gariepinus* – *Isam Eldin Mohamed Abu Zeid, Mohd Arif Syed, Johari Ramli, Juzu Hayati Arshad, Ismail Omar & Nor Aripin Shamaan* 249
- Consensus of Criteria Ranking in Women's Decision-making – *Puzziawati Ab. Ghani & Abdul Aziz Jemain* 257
- Estimation of Evapotranspiration in a Rice Irrigation Scheme in Peninsular Malaysia – *T.S. Lee, M.M.M. Najim, M.H. Aminul & Y.F. Huang* 271

The Application of Water Cooling on Reducing NO_x from a Gas Burner System

¹Mohammad Nazri Mohd. Jaafar, ²G. E. Andrews & ²M. C. MkPadi

¹*Department of Aeronautics & Automotive,*

Mechanical Engineering Faculty

Universiti Teknologi Malaysia

81310 UTM Skudai, Johor, Malaysia

²*Department of Fuel & Energy, Leeds University*

Leeds LS2 9JT United Kingdom

Received: 15 November 2001

ABSTRAK

Sebuah sistem pembakar yang mengaplikasikan kaedah penyejukan air telah dikaji menggunakan kebuk pembakar dengan garis pusat dalam 140-mm dan panjang 294-mm. Kebuk pembakar diletakkan secara menegak. Kesemua ujian dijalankan menggunakan gas asli sahaja. Pemusar udara bilah lurus yang tetap dengan garis pusat keluaran 76 mm telah dipasangkan di satah masukan ke kebuk pembakar. Sebuah plat orifis dengan garis pusat 59 mm telah dipasangkan di satah keluaran pemusar untuk menggalakkan pembentukan gelora dan membantu dalam percampuran bahan api dengan udara. Bahan api dipancarkan di plat belakang pemusar menggunakan pemancit bahan api pusat dengan lapan lubang bahan api yang mengarah keluar secara jejarian. Semua ujian dilakukan pada kejatuhan tekanan 5-mm H_2O . Pengurangan emisi NO_x sebanyak 21.53 telah dicapai pada nisbah setara hampir stoikiometri (0.88) dan pengurangan sebanyak 35.7% telah dicapai pada nisbah setara 0.42. Emisi lain seperti karbon monoksida adalah di bawah aras 100 ppm kecuali pada nisbah setara di bawah 0.5 yang disebabkan oleh kesan penyejukan. Emisi hidrokarbon tak terbakar juga adalah di bawah 10 ppm kecuali pada nisbah setara di bawah 0.5.

ABSTRACT

A gas burner system applying water cooling has been investigated using a 140-mm inside diameter combustor of 294-mm length. The combustor was placed vertically upwards. All tests were conducted using natural gas only. A fixed straight blade radial swirler with 76-mm outlet diameter was placed at the inlet plane of the combustor. An orifice plate of 59 mm was inserted at the exit plane of the swirler to enhance turbulence and help in mixing of the fuel and air. Fuel was injected at the back plate of the swirler using central fuel injector with eight fuel holes pointed radially outward. Tests were conducted at 5-mm H_2O pressure loss. A reduction of about 21.53% on NO_x emissions was achieved at equivalence ratio of near stoichiometric (0.88) and a reduction of 35.7% was achieved at equivalence ratio of 0.42. Other emissions such as carbon monoxide were well under 100 ppm except below the equivalence ratio of 0.5 due to cooling effect. Unburned hydrocarbon emissions were well below 10 ppm except below the equivalence ratio of 0.5.

Keywords: Oxides of nitrogen, carbon monoxide, unburned hydrocarbons, water cooling

INTRODUCTION

The effects of increased levels of NO_x (oxides of nitrogen) in the atmosphere are far reaching. In the atmosphere nitric oxide (NO) is rapidly oxidised to nitrogen dioxide (NO_2) and in this form plays an essential role in the formation of tropospheric ozone and photochemical smog. NO_2 is further oxidised to form nitric acid that may then be deposited as acid rain (Harrison 1990). At ground level, increased concentrations (above 0.06 ppm) of NO_2 can cause respiratory problem (World Health Organisation 1987).

NO_x emission limits legislation in many parts of the world has substantially complicated the burner design process. Attempts at lowering NO_x emissions by reducing the flame temperature will lead to reduced flame stability or increased carbon monoxide (CO) emissions. Unacceptable stability problems or CO emissions always limit the lowest NO_x emission obtained in a given configuration. Therefore, the process of burner design has become a trial-and-error, multi-parameter optimisation process (Van Der Meij *et al.* 1994).

Generally there are two techniques of controlling NO_x : those which prevent the formation of nitric oxide, NO and those which destroy NO from the products of combustion.

The former approach has been widely applied. The main aim is to reduce the flame temperature thus reducing the formation of NO_x . There are several ways of accomplishing this.

One method is the modification of combustion processes, i.e. either to burn fuel-rich or fuel-lean. The operation of lean burning is to introduce additional air in order to reduce the flame temperature. This would generally cause a significant decrease in the production of NO_x . A reduction of the primary-zone flame temperature on the other hand may increase the emissions of carbon monoxide and unburned hydrocarbon (UHC). The problem with fuel-rich combustion is the formation of soot and CO, even though the stability margin is widened.

Another method, i.e. steam or water injection has been shown to be very effective in accomplishing the above goal (Shaw 1974). A typical NO_x reduction curve as a function of rate of water injection is shown in *Fig. 1* (Correa 1991). These data were obtained in an aeroderivative inductive gas turbine at full power. Fox and Schlein (1992), testing the FT8 gas turbine combustor, also demonstrated the same effect in their final test run. The FT8 engine is an industrial/marine gas turbine engine that is a derivative of the widely used JT8D aircraft jet engine. However, to avoid detrimental effects on turbine durability, the water has to be purified to a maximum of 2.5 ppm of dissolved solids (Sarofim and Flagan 1976; Correa 1991). Furthermore, there are other complications such as incorporating the water injection system into the combustor design. Another disadvantage of water injection is the undesirable side effects of quenching CO burnout. These drawbacks render water injection method unattractive for smaller gas turbines or where availability of sizeable water supply is difficult. However, it is a feasible technique for burner NO_x control in water heater or steam generator.

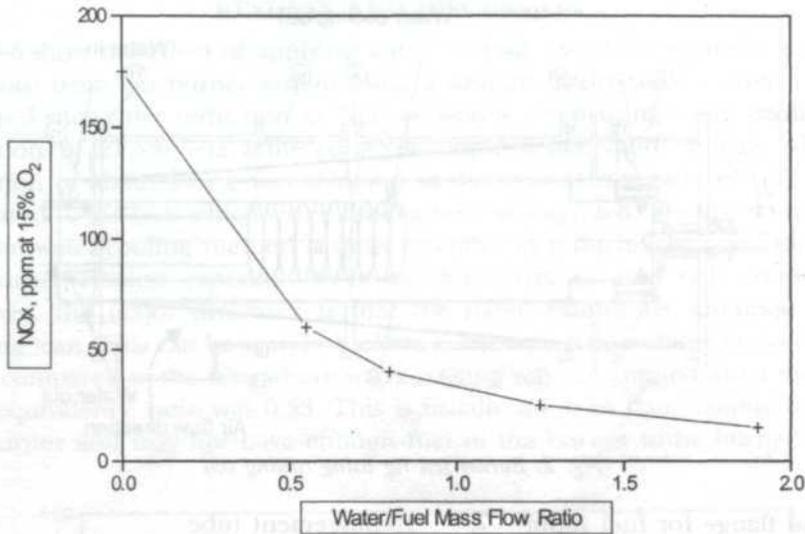


Fig. 1: NO_x as a function of water addition in a gas-turbine combustor running on natural gas at a pressure ratio of 30 (after Correa 1991)

Other methods of NO_x control involve staged combustion, variable geometry combustion, lean premixed prevaporised combustion and catalytic combustion. In the present work, a water cooling comprising of a 6-mm stainless steel tube coil was used to reduce emissions by supplying constant amount of cooling water through it. The cooling coil was placed at a particular distance from the flame.

METHOD

The general rig set-up for burner tests is shown in Fig. 2. The air is introduced through the inlet pipe, designated (13) on the diagram and flows downward before entering the combustor through the fixed straight blade radial swirler of 76-mm outlet diameter.

The rig is equipped to be fuelled either by wall or central fuel injector. The central fuel injector is of a simple type with eight holes pointing radially outward. The inside diameter of the combustor is 140 mm and the length is 294 mm. The combustor is enclosed by a 250 mm inside diameter pipe through which air flows. Thus, the air is actually preheated by the combustion inside the burner as it flows through this pipe before entering the combustor through the swirler. However, this preheat cannot be controlled externally. The exhaust sampling probe is mounted at the top of the end pipe to measure all the major exhaust gases such as oxides of nitrogen (NO and NO_2), carbon monoxide and unburned hydrocarbon.

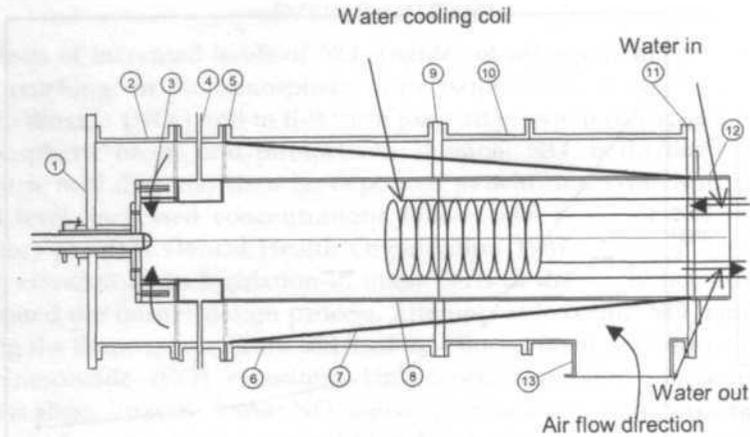


Fig. 2: Burner test rig using cooling coil

- | | |
|---|---|
| 1. End flange for fuel supply & swirler support | 7. Impingement tube |
| 2. Swirler case pipe | 8. 140mm-dia. case pipe |
| 3. Variable swirler | 9. Joint flange |
| 4. Wall fuel injector | 10. Reverse air passage pipe |
| 5. Flange | 11. Recess flange |
| 6. 250mm-dia. case pipe | 12. End pipe |
| | 13. Reverse air inlet pipe inlet. Inlet P & T tapping |

Test Conditions

Tests were carried out at around 400 K inlet temperature simulating domestic central heating unit. The air was preheated by combustion in the main combustion chamber. However, to maintain the inlet temperature at this value is almost impossible since the air preheated temperature cannot be controlled externally. At some point in these tests, the air inlet temperature will exceed 400 K. Natural gas was used as fuel throughout the entire investigation. Temperature of the cooling water was not measured in this experiment.

Domestic central heating boilers operate with a fan air supply at below 0.5 percent (5 mb, 50 mmH₂O) pressure loss compared to 2-5% pressure loss for gas turbine combustion. These pressure losses of 2-5% when converted to burner pressure loss are equivalent to 200-500 mmH₂O.

Tests were conducted using an orifice plate of 59 mm diameter that was inserted at the exit plane of the outlet of the wall injector section (refer to Fig. 2). This was to enhance flame stabilisation and to provide better mixing of air and fuel prior to ignition. The orifice plate also helped to prevent fuel from entraining into the corner recirculation zone that will create a local rich zone hence resulting in higher local NO_x emissions which ultimately contributes to high total NO_x emissions.

RESULTS AND DISCUSSION

Figs. 3-6 show the effect of applying water cooling on reducing mean exhaust emissions from gas burner system using a straight blade radial swirler.

Fig. 3 shows the reduction in NO_x emissions when using water cooling. A reduction of 21.5% was achieved at the equivalence ratio of 0.88, while a reduction of about 35.7% was achieved at the equivalence ratio of 0.42 when compared to the tests which water cooling were not applied. These demonstrate that the water-cooling method is quite effective in reducing NO_x emissions in gas burner systems especially near stoichiometric or fuel rich conditions. However, the major drawback is that the flame cannot be sustained when burning lean. This can be seen by looking at the lean flammability limits of 0.42 when compared to the tests where water-cooling was not applied when the lean limit equivalence ratio was 0.33. This is usually the lean flammability limit of this burner and may not have enough fuel in the burner to be burned.

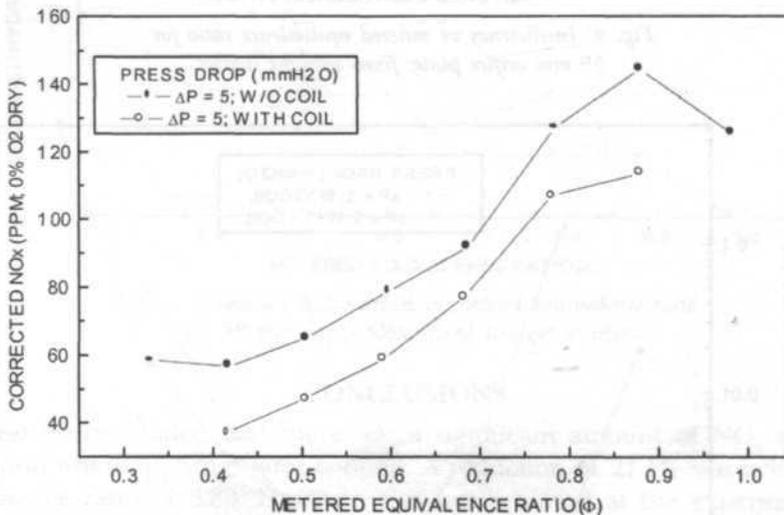


Fig. 3: Corrected NO_x vs metered equivalence ratio for 59 mm orifice plate; fixed straight swirler

Fig. 4 shows the combustion inefficiency plotted against equivalence ratio for both test conditions. It could be seen that combustion inefficiency was increased when using water-cooling. Combustion inefficiency increases drastically near lean condition to around 1% from about 0.03% at equivalence ratio of near 0.9. This could be due to flame quenching when using water-cooling where the emission of carbon monoxide increases and causes the combustion inefficiency to rise.

Fig. 5 shows the carbon monoxide emissions plotted against equivalence ratio. It shows the same trend as a combustion inefficiency curve that implies that the combustion efficiency was influenced by carbon monoxide oxidation. However, it could be seen that carbon monoxide emissions increase when

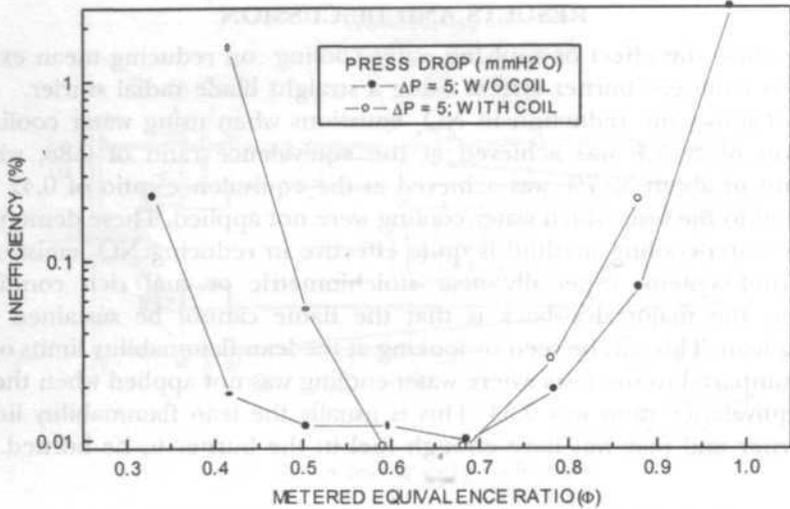


Fig. 4: Inefficiency vs metered equivalence ratio for 59 mm orifice plate; fixed straight swirler

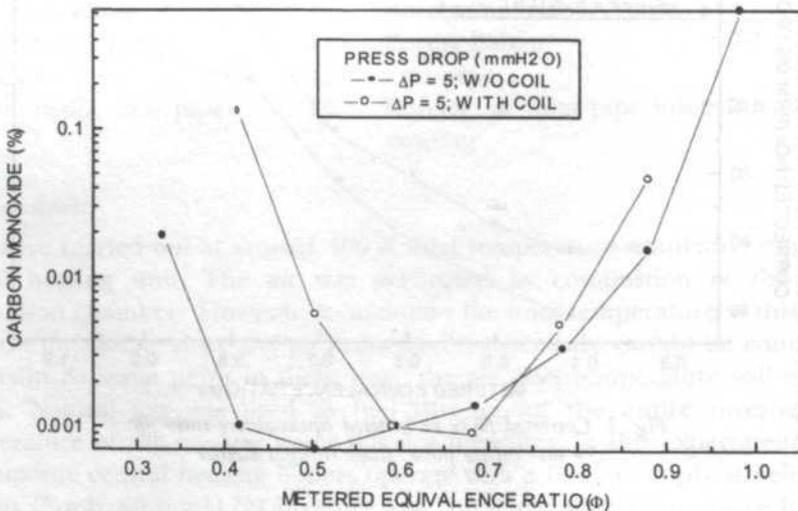


Fig. 5: Carbon monoxide vs metered equivalence ratio for 59 mm orifice plate; fixed straight swirler

water cooling is applied. There was some reduction in carbon monoxide emissions between equivalence ratios of 0.6 to 0.73. The carbon monoxide curves show the same curvature for both conditions. There is a significant increase in the lean region that is due to insufficient residence time for carbon monoxide burnout at the low flame temperatures of these lean mixtures. The significant increase in carbon monoxide emissions on the rich side is due to high equilibrium carbon monoxide at these rich equivalence ratios.

Unburned hydrocarbon emissions of less than 10 ppm can be achieved for both conditions over a wide range of operating equivalence ratios and this is shown in Fig. 6. However, water cooling actually increases unburned hydrocarbon emissions rather than reduce them, except over a small range of equivalence ratios of 0.53 to 0.65. This may be due to several reasons. One of them is the chilling effect of water cooling and the other may be inadequate burning rates at lean conditions.

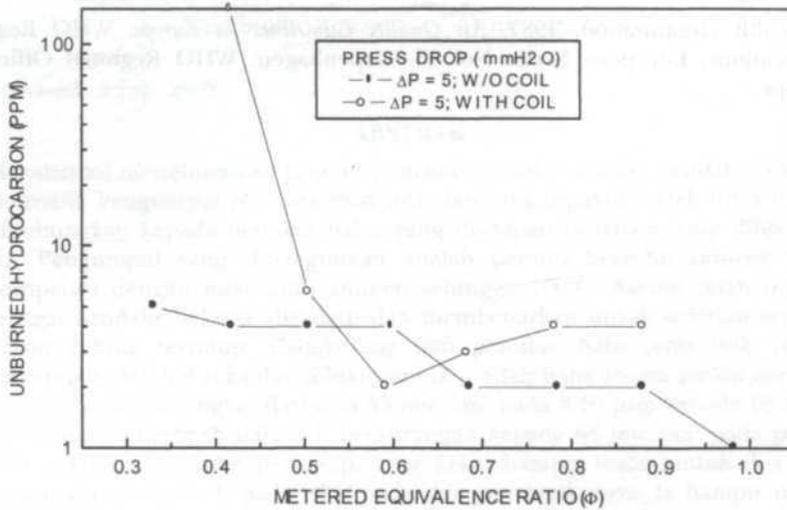


Fig. 6: Unburned hydrocarbon vs metered equivalence ratio for 59 mm orifice plate; fixed straight swirler

CONCLUSIONS

It could be concluded that there was a significant amount of NO_x emission reduction when applying water cooling. A reduction of 21.5% was achieved at equivalence ratio of 0.88. However, this was achieved at the expense of an increase in other emissions such as carbon monoxide and unburned hydrocarbon by 200 and 100%, respectively. These increases in carbon monoxide and unburned hydrocarbon are even larger for lean burn since at this condition, the chilling effect is greater due to lower temperature caused by water cooling and also by burning less fuel, thus producing lower temperatures.

REFERENCES

- CORREA, S.M. 1991. A review of NO_x formation under gas-turbine combustion conditions. *Combust. Sci. and Tech.* **87**: 329-362.
- FOX, T.G. and B.C. SCHLEIN. 1992. Full annular rig development of the FT8 gas turbine combustor. *Journal of Engineering for Gas Turbine and Power* **114**: 27-32.
- HARRISON, R.M. 1990. *Important Air Pollutants and Their Chemical Analysis. Pollution: Causes, Effects and Control*. pp. 156-175. London: Royal Society of Chemistry.

- SAROFIM, A.F. and C. FLAGAN. 1974. NO_x control for stationary combustion sources, *Prog. Energy Combustion Science* **2**: 1-25.
- SHAW, H. 1974. The effects of water, pressure, and equivalence ratio on nitric oxide production in gas turbines. *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power* July: 240-246.
- VAN DER MEIJ, C.E. *et al.* 1994. On the effects of fuel leakage on CO production from household burners as revealed by LIF and CARS. *Twenty Fifth Symposium (International) on Combustion*, pp. 243-250. The Combustion Institute.
- World Health Organisation. 1987. *Air Quality Guidelines for Europe*. WHO Regional Publications, European Series No. 23. Copenhagen: WHO Regional Office for Europe.

An Experimental Study on Solar Flat Plate Collector Using an Alternative Working Fluid

¹A. Manickavasagan, ²A. Sampathrajan, ²P.R. Manigandan, ²S. Manivannan,
²F. Magdaline Eljeeva Emerald, ²R. Malarkodi & ²S. Mangalam

¹*Department of Biosystems Engineering*

Faculty of Agricultural and Food Sciences, University of Manitoba, Canada

²*Department of Bio Energy, College of Agricultural Engineering
Tamilnadu Agricultural University, India*

Received: 4 July 2002

ABSTRAK

Makalah ini menghuraikan prestasi pemanas air solar dengan bendalir bekerja alternatif. Pengumpul plat rata solar orientasi yang dipasang telah dibikin dan dihubungkan kepada penukar haba, yang disimpan di dalam tong simpanan air. Pengumpul yang dibangunkan adalah peranti bersuhu rendah yang beroperasi dengan julat suhu ambien sehingga 100°C. Aseton telah dipilih sebagai bendalir bekerja alternatif dan membenarkan untuk sirkulasi semula dalam gelung tertutup. Cangkerang dan penukar haba jenis tiub (tidak bercampur) telah direka dan dibikin untuk pindah haba antara aseton dan air. Intensiti solar meningkat daripada 55 mw/cm² pada 8.30 pagi kepada 85 mw/cm² pada 1.30 tengah hari dan berkurangan kepada 64 mw/cm² pada pukul 5.00 petang. Suhu plat penyerap, yang kekal hampir malar untuk dua jam pertama, meningkat kepada 90°C pada dua jam berikutnya. Ia hampir malar dari pukul 11.00 pagi hingga 1.30 tengah hari dan kemudian turun kepada pada tengah hari. Bendalir panas didapati melepaskan haba kepada air di sekeliling pada kadar malar lebih atau kurang dari pukul 11.30 pagi hingga 5.00 petang. Suhu aseton paling tinggi (72 °C) diperhatikan semasa pukul 1.00 tengah hari hingga 2.00 petang. Suhu air meningkat daripada 30°C kepada maksimum 62°C. Keseluruhan keberkesanan sistem tersebut didapati adalah 45% sementara mempertimbangkan input solar kepada haba yang diperolehi oleh air. Bendalir bekerja alternatif seperti aseton, metanol atau etanol mungkin dipertimbangkan untuk menggantikan air dalam pengumpul plat rata memandangkan kepada fakta bahawa bendalir tersebut mempunyai titik didih rendah dikaitkan dengan haba pendam tinggi penyejatan.

ABSTRACT

This paper describes the performance of a solar water heater with an alternative working fluid. A solar flat plate collector of fixed orientation was fabricated and connected to a heat exchanger, which was kept inside the water storage drum. The developed collector was a low temperature device operating with the temperature range of ambient to 100°C. Acetone was selected as an alternative working fluid and allowed to re-circulate in a closed loop. A shell and tube type heat exchanger (non-mixing) was designed and fabricated for the heat transfer

* Corresponding Author:
E-mail: umannama@cc.umanitoba.ca

between acetone and water. The solar intensity increased from 55 mw/cm² at 8.30 am to 85 mw/cm² at 1.30 pm and decreased to 64 mw/cm² at 5.00 pm. The absorber plate temperature, which remained nearly constant for the first two hours, went up to 90°C in the next two hours. It was nearly constant from 11.00 am to 1.30 pm and then dropped down to 72°C in the afternoon. The hot fluid was found to loose heat to the surrounding water at more or less constant rate from 11.30 am to 5.00 pm. The highest temperature of acetone (70°C) was observed during 1.00 pm to 2.00 pm. The temperature of water increased from 30°C to a maximum of 62°C. Overall efficiency of the system was found to be 45% while considering the solar input to the heat gained by the water. Alternative working fluids such as acetone, methanol or ethanol may be considered for substituting water in the flat plate collector in view of the fact that the fluids have low boiling point coupled with high latent heat of evaporation.

Keywords: Solar water heater, alternate working fluid, shell and tube heat exchanger

INTRODUCTION

A solar water heater utilises helio thermal process method to extract the solar energy. It represents a popular use of solar energy in a mechanically simple system. Solar water heaters are becoming popular because they are affordable and a cheap substitute for other conventional water heating methods. Saiful-Bari (2001) studied the orientations of domestic solar water heaters in Malaysia. The solar collectors in Malaysia are installed on the roof and the slope is almost same as the roof (30 to 40°). It was reported that most collectors in Malaysia were incorrectly installed and were receiving 10-35% less radiation than a properly installed collector, while in some extreme cases, they were receiving as little as 50%. Reddy (1995) conducted a case study on the economical aspects of solar water heaters and electrical water heaters in India. Based on certain conditions, either alternative might be economically advantageous. Hot water at moderate temperature (36 to 58°C) could be obtained in buildings during day time in winter by using reinforced cement concrete slabs or by modifying the roof structure and laying down a net work of aluminum pipes over it, which can offer a low cost passive solar water heating system in the building itself (Chaurasia 2000). In solar dryers using flat plate collectors, the air temperature was in the range of 45.5 to 55.5°C which was suitable for drying most of the agricultural commodities (EL-Sebaili *et al.* 2002).

The performance of the collector increased with the number of glass covers (Felske 1977). Agarwal and Pillai (1982) determined that high conductivity metal powders with suitable binders as bond material between the tubes and absorber plate showed a considerable improvement in the collector efficiency. According to Norton and Probert (1984), single and multi-pass water heaters exhibited similar diurnal heat gain efficiencies at high insolation levels. For lower insolation levels a multi pass unit was found to be more effective. Sorour (1985) studied various designs of continuous and intermittent solar water

heaters and found that a continuous water heater with flow through two transparent perspex sheets and perforated copper plate produced a simple and more efficient collector than the classical design. Qudais *et al.* (2002) compared the performance of a flat plate collector with a convex type collector and found that an additional 45% energy gain was achieved from the convex type over the flat plate collector. A well made solar collector could operate satisfactorily for more than 15 years with minimum maintenance (Rudnich *et al.* 1986).

Phase Changing Fluids

Generally, water is used as a working fluid for a solar flat plate collector. In a two-phase heat transfer process of a flat plate solar collector with a phase changing fluid, the collector efficiency increased with the fluid mass flow rate (Chandra *et al.* 1983). Bhargava (1983) stated that a water heater based on phase changing materials could provide hot water through out the day if the water pipes were placed near the surface. The efficiency of the system was directly related to the thermal conductivity of the liquid and solid phase of the phase changing material. Prakash *et al.* (1985) investigated the performance of a novel built in storage type water heater containing a layer of paraffin wax as the phase changing material. Water heaters with phase changing material storage at the bottom were more suitable for hot water requirements during off sunshine hours. Yilmaz and Ogulata (1990) used a two-phase solar collector in a water heating system with refrigerant as working fluid. It was claimed that refrigerants could be used with out the risk of freezing, corrosion and fouling. The two-phase collectors could work satisfactorily, if the tank height was greater than half of the collector height. According to Kaushika *et al.* (1982), collector efficiency increased with the increase in liquid length until a point is reached, when the region of super heating causes the vapour to disappear. They also found that the efficiency was high when a heat removal fluid of high latent heat of vaporization was used in the collector. An increase in the saturation temperature of the working fluid in the collector reduced its efficiency. Chun *et al.* (1999) studied the utilisation of heat pipes for solar water heaters and determined that the system performance for different types of working fluid was relatively insensitive to the selection of working fluid. The objective of this study was to identify a suitable alternative working fluid for a solar flat plate collector, which extracts more heat from solar energy, and to study the performance of the water heating system with that fluid.

MATERIALS AND METHODS

This study was conducted at the department of Bio Energy, College of Agricultural Engineering, Tamilnadu Agricultural University, India. A black painted copper sheet (26 G thickness) with 30 mm insulation (glass wool) was used as an absorber for the solar flat plate collector. The collector outer box was made of plywood. A 3 mm thickness glass cover plate (tuffened glass) was placed on the top of the collector. The specifications of the collector are given

in Table 1 and the side and top views of the over all experimental set up are given in *Figs. 1* and *2*. A heat exchanger was kept inside the water storage drum and connected to the headers in the collector. The outlet coming from the heat exchanger was connected to the bottom header and the top header was connected to the inlet tube of the exchanger. Top and bottom headers were connected by riser pipes. The copper pipes between the collector and water drum were insulated by asbestos rope to avoid heat losses to atmosphere.

TABLE 1
Dimensions of the flat plate collector

Component	Specification
Collector	
Area	1.7 m ²
Length	1.7 m
Width	1.0 m
Bottom header	
Material	Copper
Diameter	20 mm
Length	1.0 m
Riser pipes	
Material	Copper
Diameter	13 mm
Length	164 mm
No. of pipes	8
Top header	
Material	Copper
Diameter	20 mm
Length	1.0 m
Inclination of flat plate collector	20° facing south
Absorber plate	Corrugated, dull black
Number of glass covers	One (3 mm thickness)

Heat Exchanger

Various types of heat exchangers are available for heat transfer from liquid to liquid. Shell and tube, plate type, U tube and spiral tube are the widely used heat exchangers for commercial applications. The flow path configuration and the mode of heat transfer are different for each model. Yadav and Tiwari (1987) investigated the performance of one pass, double pipe heat exchanger coupled with a flat plate solar collector in both co-current and counter current modes of flow. They found that the temperature of the hot fluid stream decreased with increasing space coordinate along the length of the collectors, but that of the cold fluid was increased, in the parallel flow mode. In the counter flow mode, the temperatures of both fluid streams decreased with the increasing space

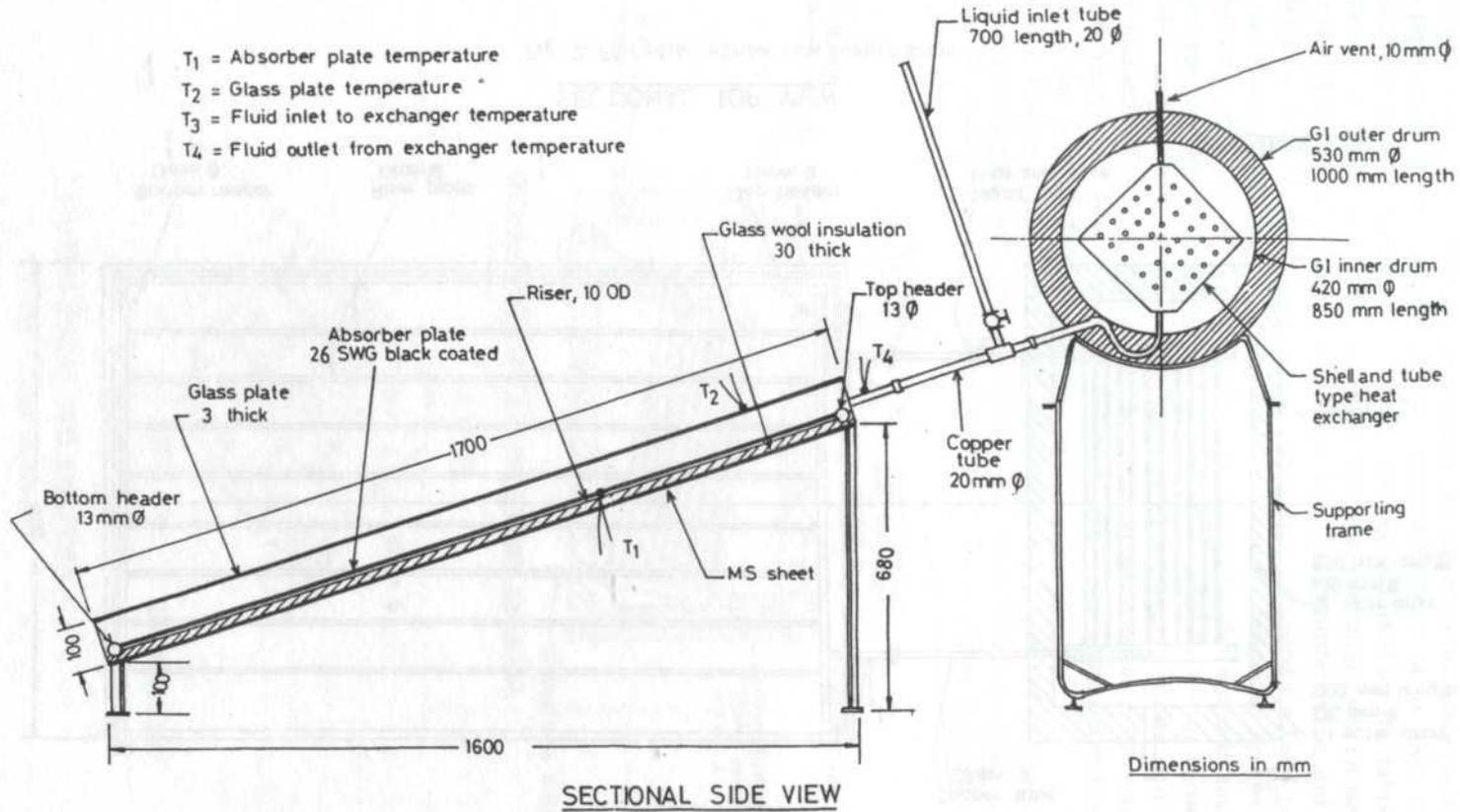


Fig. 1: Flat plate collector cum storage drum

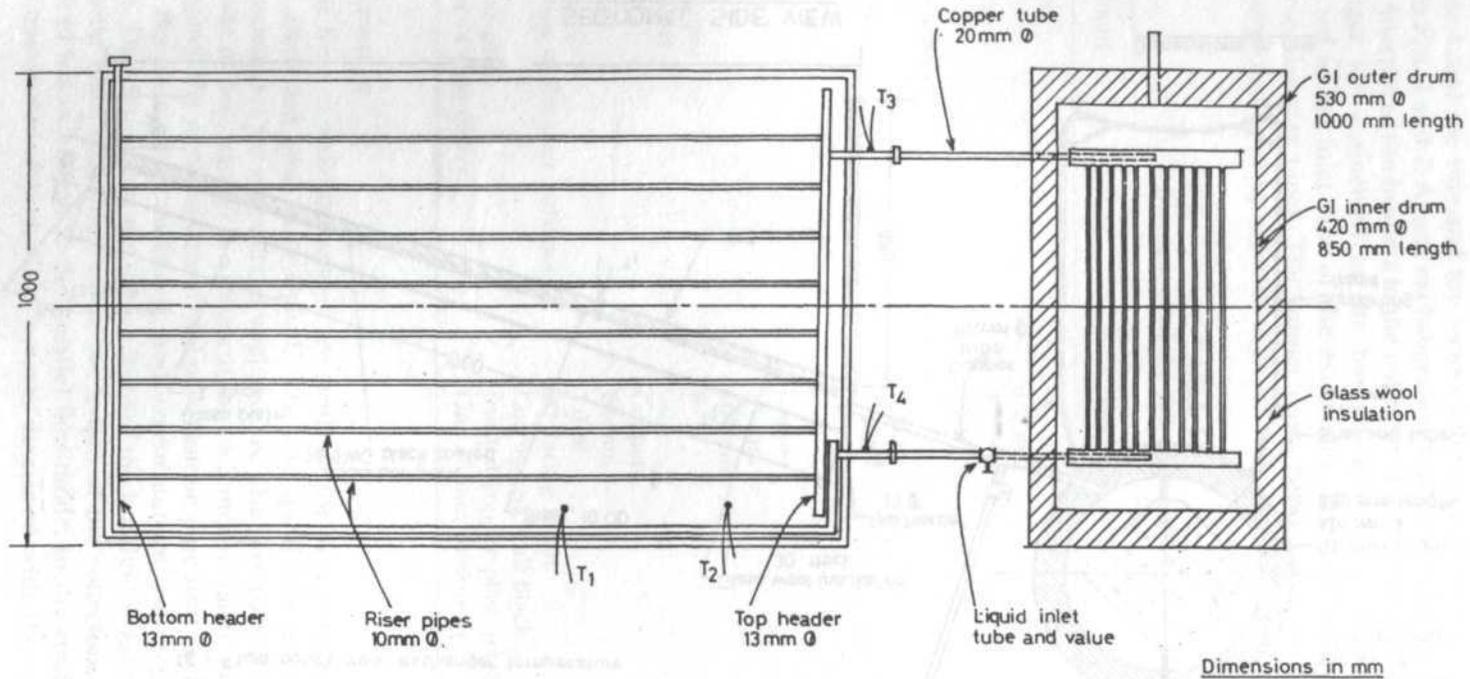
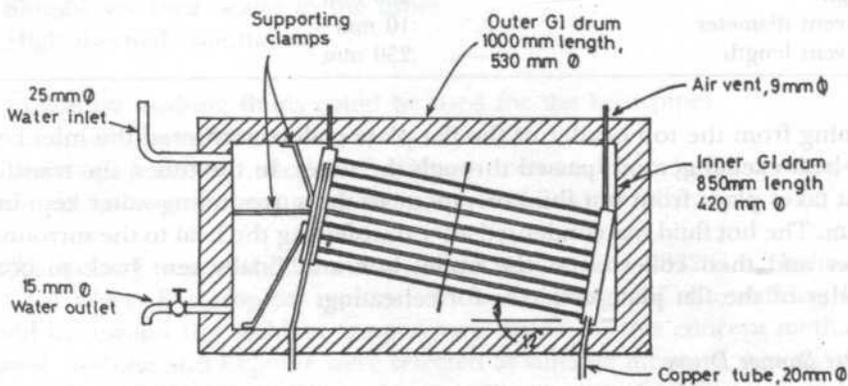


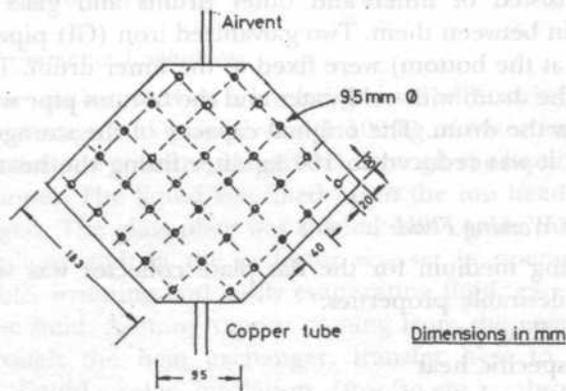
Fig. 2: Flat plate collector cum storage drum

coordinate along the heat exchanger length. As copper is normally used as the riser pipe material in commercial water heaters because of its high thermal conductivity, in our experiment the same material was used for the fabrication of the rest of the parts. A shell and tube type heat exchanger was chosen and it contained three major parts: inlet header box, heat transfer tubes and outlet header box (Fig. 3). Thirty-four copper tubes were provided for fluid passage and the spacing between tubes was kept at 40 mm (centre to centre). The copper tubes were connected to the inlet and outlet header boxes by means of copper gas welding. At the top of the inlet and outlet boxes air vents were provided and the size of the air vent was 9.5 mm outer diameter and 6 mm inner diameter. The exchanger was placed at an inclination of 12° inside an inner drum for easy flow of the fluid from the inlet box to the outlet box. The details of the heat exchanger are given in the Table 2.

A solar flat plate collector of fixed orientation was fabricated and connected to a heat exchanger. The working medium was isolated from the water and was allowed to re-circulate in a closed loop. The hot working fluid or vapour



SECTIONAL SIDE VIEW OF THE DRUM WITH EXCHANGER



SECTIONAL VIEW OF HEAT EXCHANGER

Fig. 3: Heat exchanger

TABLE 2
Specifications of the developed heat exchanger

Component	Specification
Type	:Shell and tube
Number of tubes	:34
Tube material	:Copper
Diameter	:9.5 mm (OD) 6 mm (ID)
Length	:600 mm
Total surface area	:18000 mm ²
Capacity of the inlet and outlet boxes with heat transfer tubes	:4.5 lit
Inclination with in the water drum	:12°
Overall size	
Length	:660 mm
Height	:250 mm
Air vent diameter	:10 mm
Air vent length	:250 mm

coming from the top header of the flat plate collector entered the inlet box of the heat exchanger and passed through the tubes. In the tubes, the transfer of heat takes place from hot fluid or vapour to the surrounding water kept in the drum. The hot fluid was condensed after transferring the heat to the surrounding water and then collected at the outlet box and finally sent back to bottom header of the flat plate collector for reheating.

Water Storage Drum

The specifications of the water storage drum are shown in Table 3. The storage drum consisted of inner and outer drums and glass wool insulation was provided in between them. Two galvanized iron (GI) pipes (one at the top and the other at the bottom) were fixed to the inner drum. The top pipe was used to fill up the drum with cold water and the bottom pipe was used to tap the hot water from the drum. The original capacity of the storage drum was 117 kg of water and it was reduced to 100 kg after fitting the heat exchanger.

Selection of Working Fluid

The working medium for the flat plate collector was selected based on the following desirable properties:

- i. Low boiling point
- ii. High specific heat
- iii. High latent heat
- iv. Non-corrosiveness for most of the fabrication materials
- v. Wide useful range

TABLE 3
Dimensions of the water storage drum

Component	Specification
Inner drum	
Material	G.I.Sheet
Length	0.85 m
Diameter	0.42 m
Outer drum	
Material	G.I.Sheet
Length	1.0 m
Diameter	0.53 m

- vi. Easily available in the market and low in cost
- vii. Excellent stability in the working range
- viii. Low freezing temperature
- ix. Should not form scales in the tubes
- x. High thermal conductivity

The following working fluids could be used for the heat pipes:

1. Gases such as helium, nitrogen
2. Liquids such as water, methanol, ethanol, acetone, freon, pentane, ammonia
3. Metals such as mercury, potassium, calcium, lithium, silver.

The melting points of different materials ranged from -272°C for helium to 96°C for silver. For tropical regions, the useful range of the working fluid should be around the ambient temperature. Based on this concept methanol, ethanol, acetone and heptane were selected as suitable medium. Among these four working fluids, acetone has a lower boiling point of 57°C at atmospheric pressure and hence this was chosen as the working medium.

Testing Method and Performance Evaluation

The flat plate collector was installed on a leveled ground with an inclination of 20° facing south. The storage tank was filled with 100 kg of water and the fluid tubes were filled with acetone (commercial grade) through the liquid inlet pipe with the help of a funnel. The liquid was filled up to the top header and the filling tube was plugged. The glass plate was shaded with canvas during filling then exposed to sunlight and the collector was set in operation. Since acetone is inflammable, irritating and easily evaporating fluid, care should be taken while filling the fluid. Acetone vapour coming from the riser tubes was allowed to pass through the heat exchanger, transfer heat to water and condensed back to liquid. Solar insolation ($\text{mw}/\text{sq.cm.}$), absorber plate temperature ($^{\circ}\text{C}$), glass plate temperature ($^{\circ}\text{C}$), acetone inlet to exchanger temperature ($^{\circ}\text{C}$), acetone outlet from exchanger temperature ($^{\circ}\text{C}$), water temperature ($^{\circ}\text{C}$) and ambient temperature ($^{\circ}\text{C}$) were recorded at 30-minute

intervals. Solar insolation was measured by surya mapi (commercially available instrument to measure solar insolation) and water and ambient temperatures were measured by mercury thermometers. All other temperatures were measured by copper constant thermocouples fixed at the respective spots. An overview of the unit with all observation instruments is shown in Fig. 4.



Fig. 4: Overview of the experimental unit with observation instruments

Efficiency Calculation

The efficiency of the flat plate collector with an alternative working fluid was calculated using the heat gained by water with respect to the actual solar energy received by the flat plate collector (Eqn. 1).

$$\text{Overall efficiency of the system} = (\text{heat gained by water} / \text{input solar energy}) \quad (1)$$

$$\text{Heat gained by the water} = m c_p \Delta T \quad (2)$$

$$\text{Input solar energy (solar energy falling on the collector)} = q A T \quad (3)$$

where

m = weight of water (gram)

c_p = specific heat of water (joule/gram °C)

- ΔT = temperature difference (outlet temperature–inlet temperature, °C)
 q = solar insolation (joule/ hour . m²)
 A = area of the collector (m²)
 T = time (hour)

RESULTS AND DISCUSSION

The flat plate collector developed for this study was a low temperature device operating with the temperature range of ambient to 100°C. It absorbed solar energy consisting of both beam and diffused radiation and transferred heat to the working fluid. The flat plate collector of fixed orientation was selected in view of the fact that the construction is easier, cost is low, requires no tracking, is easy to maintain and is widely available compared to concentrating collectors. Similar type of flat plate collectors are commercially used with water as working fluid for heating water of about 100 litres within a day. This system was modified for indirect heating of water using acetone as working media through a heat exchanger. The performance of the collector with acetone as working fluid was monitored by observing the variations of temperature at different locations and recorded at 30-minute intervals (Table 4).

Variation of Ambient Parameters

Fig. 5 indicates the changes in solar intensity, absorber plate temperature and glass plate temperature of the solar collector with respect to time. It was observed that the solar intensity increased from 55 mw/cm² at 8.30 AM to 85 mw/cm² at 1.30 PM and decreased to 64 mw/cm² at about 5.00 PM. The glass plate temperature was in the range of 38 to 60°C during sunset. It was also noted that the glass plate temperatures closely followed the trend of solar intensity with uniform variation. The absorber plate temperature, which remained nearly constant for the first two hours, went up to 90°C in the next two hours. It remained nearly constant from 11.00 AM to 1.30 PM for a period of three hours and dropped down to 72°C in the afternoon. A difference in temperature of about 24°C was observed over a period of 3½ hours between the absorber plate and glass plate. The increase in the absorber plate temperature from 11.00 AM may be due to the increased solar intensity and the trapping of infrared flux.

Variation of Working Fluid Temperature

Fig. 6 describes the variations in the inlet and outlet temperature of the working fluid. The fluid temperature at inlet to the heat exchanger varied from 38°C at 8.30 AM to about 70°C at 2.00 PM and 52°C at 5.00 PM. The temperature difference between inlet and outlet fluid was 6°C during the start of the trial and it gradually went up to 30°C. The hot fluid was found to lose heat to the surrounding water at more or less constant rate from 11.30 AM till 5.00 PM. The boiling point of acetone was reached after a lapse of about three hours from the start of the trial. It may be inferred that after three hours only acetone

TABLE 4
Observations on solar flat plate collector with alternate working fluid

Time	Solar insolation mw/cm ²	Temperature °C					
		Absorber plate	Glass plate	Acetone- inlet	Acetone- outlet	Water- inlet	Water- outlet
8.00	56	40	38	38	32	26	26
8.30	56	40	38	38	32	26	26
9.00	56	40	38	38	34	26	26
9.30	60	52	40	42	36	26	28
10.00	68	60	45	48	36	26	30
10.30	72	65	48	54	36	26	32
11.00	78	88	52	60	36	26	38
11.30	78	90	56	62	40	26	42
12.00	82	90	58	68	40	26	48
12.30	84	90	58	68	42	26	52
1.00	85	93	60	70	43	26	56
1.30	85	93	60	70	43	26	56
2.00	82	93	60	70	43	26	56
2.30	82	90	58	66	38	26	56
3.00	80	88	58	66	36	26	56
3.30	78	85	58	63	36	26	56
4.00	75	80	56	60	32	26	56
4.30	70	80	50	55	30	26	56
5.00	64	72	46	52	30	26	54

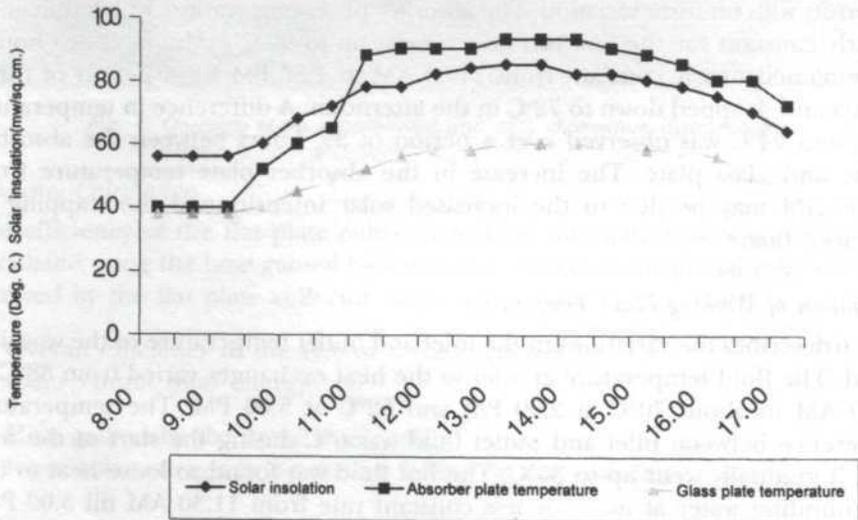


Fig. 5: Variation of solar intensity and temperatures of collector

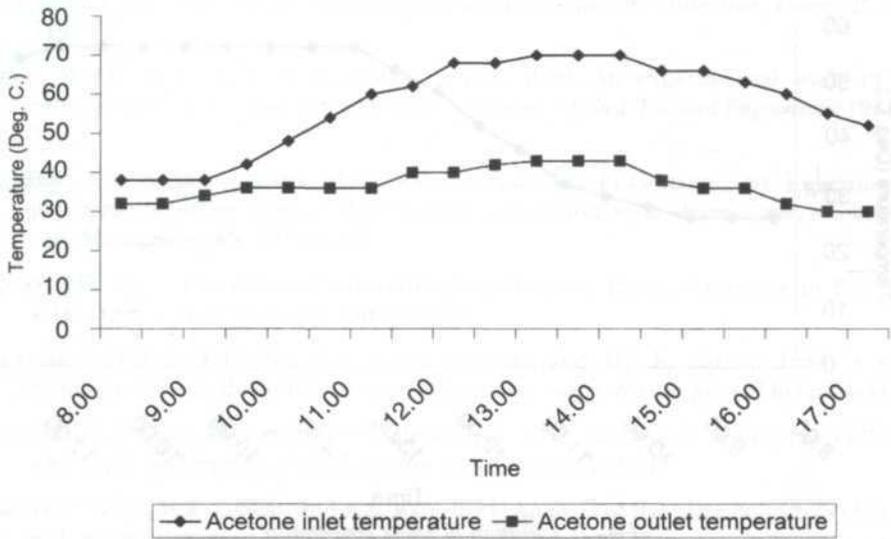


Fig. 6: Variation of working fluid temperature

heated evaporately and the advantage of latent heat of vaporization was realised. This trend is explicitly shown in the temperature variation in Fig. 6. Kaushika *et al.* (1982) also reported that the efficiency of the system was high when a heat removal fluid of high latent heat of vaporisation was used in the collector. It is also inferred that flux of the vaporisation stopped after 4.30 PM due to insufficient solar insolation as evident in the buff fluid inlet temperature. Hence the effective time was only 5½ hours when it was exposed to nearly 9 hours giving an effective exposure time of nearly 50%. The temperature of acetone was maximum (70°C) during 1.00 PM to 2.00 PM and it was not possible to workout the evaporation efficiency due to lack of proper instrumentation facility. Kishore *et al.* (1984) determined that the liquid level was one of the important factors affecting the efficiency of the collector while using phase changing fluids. From the constant temperature profile of both inlet and outlet fluids, the rate of evaporation remained nearly constant due to constant solar insolation. The initial water temperature was 26°C and went up to a maximum of 56°C at 1.00 PM. It remained constant till 4.30 PM and then decreased to 54°C at 5.00 PM (Fig. 7). Hot water with this temperature may be sufficient for domestic use, but for catering applications it may require slightly higher temperature (around 70°C) (Hawlander *et al.* 1984). The temperature of water followed closely the trend of acetone inlet temperature. Bhargava (1983) reported that the water temperature during the evening hours increased with the increase in thermal conductivity solid-liquid phases of the material in a solar water heater with phase changing material.

The overall efficiency of the system was 45% while considering the solar insolation falling on the collector to the heat gained by the water in the storage tank. Chun *et al.* (1999) also achieved the same efficiency while using heat pipes

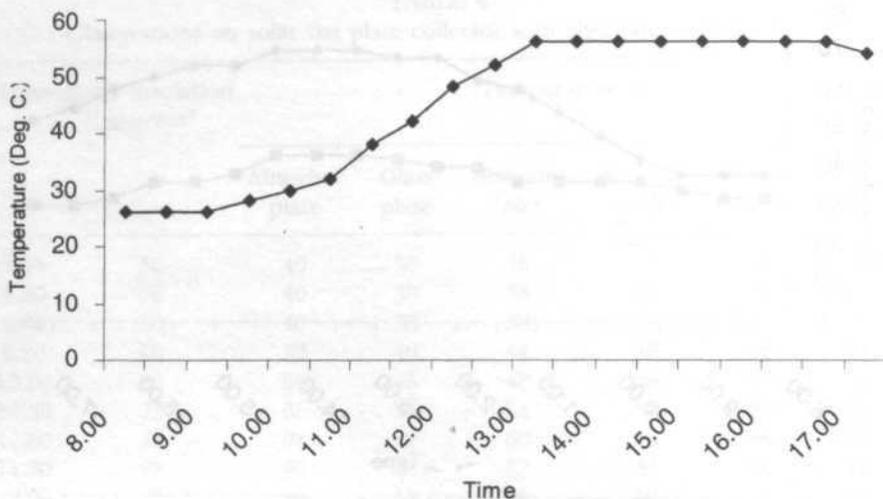


Fig. 7: Variation of water temperature

for solar water heaters. While evaluating the performance of solar water heater with steam cooker, Mahar (1985) achieved an efficiency of 51.5% for solar water heater and 16.1% for solar steam cooker.

CONCLUSION

Alternative working fluids such as acetone, methanol and ethanol can be considered for substituting water in the flat plate collector in view of the fact that the fluid has low boiling point coupled with high latent heat. However, problems with the handling of the fluid, corrosion of the components and the environment effects have to be kept in mind before choosing an alternative fluid. Careful consideration has to be made in cases where the hot water is used for edible purposes which may be contaminated with diffused working fluid in the heat exchanger due to hairline cracks, pressure build up or due to corroded components. These alternate fluids have potential for further exploration as working fluid instead of water and it is also possible to de-link the flat plate collector from the storage tank.

REFERENCES

- AGARWAL, R.C. and P.K.C. PILLAI. 1982. Bond studies on a flat plate solar collector. *Sun World* 6(4): 110-111.
- BHARGAVA, A.K. 1983. A solar water heater based on phase changing material. *Applied Energy* 14: 197-209.
- CHANDRA, S., S. KAUSHIK, J. SINGH and N. KAUSHIK. 1983. Solar collector with heat removal fluid undergoing phase change-Basic analysis. *Solar World Congress* 2: 762-768.

- CHAURASIA, P.B.L. 2000. Solar water heaters based on concrete collectors. *Energy* **25**: 703-716.
- CHUN, W.Y.H. KANG, H.Y. KWAK and Y. S. LEE. 1999. An experimental study of the utilization of heat pipes for solar water heaters. *Applied Thermal Engineering* **19**: 807-817.
- EL-SEBAIL, A.A., S. ABOUL -ENEIN, M.R.I. RAMADAN and H.G. EL-GOHARY. 2002. Experimental investigation of an indirect type natural convection solar dryer. *Energy Conversion and Management* **43**: 2251-2266.
- FELSKE, J.D. 1977. The effect of off-south orientation on the performance of flat plate solar collectors. *Solar Energy* **20**(1): 29-36.
- HAWLADER, M.N.A., K.C. NG, T.T. CHANDRATILLEKA and H.L.K. KELVIN. 1984. A solar installation for industrial hot water. *Renewable Energy Review Journal* **6**(1): 40-47.
- KAUSHIKA, N.D., S.C. KAUSHIK and S.C. BHARADWAJ. 1982. Analysis of a flat plate collector with fluid undergoing phase. *Applied Energy* **11**(3): 233-242.
- KISHORE, V.V.N., M.R. GANDHI and K.S. RAO. 1984. Analysis of flat plate collectors charged with phase changing fluids. *Applied Energy* **17**(2): 113-149.
- MAHAR, N.M. 1985. Performance and testing of an improved solar water heater cum steam cooker. *Energy Research* **9**: 113-116.
- NORTON, B. and S.D. PROBERT. 1984. Achieving thermal rectification in natural circulation solar water heaters. *Applied Energy* **14**(3): 211-225.
- PRAKASH, J., H.P. GARG and G. DATTA. 1985. A solar water heater with a built in latent heat storage. *Energy Conversion Management* **25**(1): 51-56.
- QUDAIS, M.A., T. TAMIMI and F. Al MOMANI. 2002. Experimental study of a tubeless convex type solar collector. *Energy Conversion and Management* **43**: 791-797.
- REDDY, B.S. 1995. Electrical vs solar water heater: A case study. *Energy Conversion and Management* **36**(11): 1097-1106.
- RUDINCH, A., Y. KAPLAN, A.I. KUDISH and D. WOLY. 1986. A study of solar collector-aging, installation and materials problem. *Solar Energy* **36**(3): 227-240.
- SAIFUL-BARI. 2001. Optimum orientation of domestic solar water heaters for the low latitude countries. *Energy Conversion and Management* **42**: 1205-1214.
- SOROUR, M.M. 1985. Developments in solar water heaters. *Energy Conversion Management* **25**(3): 365-372.
- YADAV, Y.P. and G.N. TIWARI. 1987. Transient analysis of double pipe heat exchanger coupled to flat plate solar collector. *Energy Conversion Management* **27**: 15-19.
- YILMAZ, T. and R.T. OGULATA. 1990. Influence of the tank height on the efficiency of a two-phase solar water heater. *Energy and the Environment* **2**: 749-755.

Thermal Diffusivity Measurement of La Doped SrSnO₃ (Sr_{1-x}La_xSnO₃) Ceramics by Photoflash Method

Josephine L.Y.C, *W. Mahmood Mat Yunus,
Iftetan A. Taha & Abdul Halim Shaari

Department of Physics
Faculty of Science and Environmental Studies,
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia

Received: 10 February 2003

ABSTRAK

Keresapan terma, a satu medium adalah sifat fizikal terma yang menentukan kelajuan haba merambat melalui proses kekonduksian semasa wujudnya perubahan suhu dengan masa. Pengukuran keresapan terma boleh dilakukan pada suhu bilik dengan kaedah cahaya kilat, dengan menggunakan sinar kilat kamera berkeamatan tinggi sebagai punca pengujaan. Dalam kertas ini, seramik Sr_{1-x}La_xSnO₃ didop dengan komposisi La yang berbeza (x = 0.1, 0.4, 0.6, 0.8 dan 1.0) telah dikaji sifat keresapan termalnya. Nilai keresapan terma berubah daripada 0.44 x 10⁻² - 1.15 x 10⁻³ cm² /s bersepadanan dengan x = 0.1 hingga x = 1.0. Morfologi dan struktur seramik Sr_{1-x}La_xSnO₃ telah dianalisis daripada pengukuran SEM dan XRD untuk menyokong pelakuan keresapan termal dalam kajian ini.

ABSTRACT

Thermal diffusivity, α of a medium is the thermophysical property which determines the speed of heat propagation by conduction during changes of temperature with time. The measurement of thermal diffusivity can be carried out at room temperature by photoflash method, which employs a high intensity camera flash as the excitation source. In this paper, Sr_{1-x}La_xSnO₃ ceramics doped with different compositions of La (x = 0.1, 0.4, 0.6, 0.8, 1.0) was investigated for their thermal diffusivity. The thermal diffusivity values varied from 0.44 x 10⁻² - 1.15 x 10⁻² cm²/s corresponding to x = 0.1 to x = 1.0. The morphology and structure of Sr_{1-x}La_xSnO₃ ceramics were analysed from SEM and XRD measurements to support the thermal diffusivity behaviour.

Keywords: Ceramic, doping effect, thermal diffusivity, XRD, SEM

INTRODUCTION

Alkaline earth stannates have received more attention as a component of ceramic dielectric elements (Jaffe *et al.* 1971; Buchanan 1986). The phase equilibria in the SrO-SnO₂ system have been studied by several authors, and the existence of the two stable phases have been reported (Pfaff 2000). A number of investigations has been carried out for humidity sensors utilising porous ceramic material. For application in the fields of drying processes and

* Corresponding Author

combustion control systems, there is a need for humidity sensor, which operates above 100. From the standpoint of the temperature capability of the sensing element, ceramic materials appear to be the most suitable candidates. It has been reported that CaSnO_3 , SrSnO_3 and BaSnO_3 are sensitive to humidity and oxygen gas (Shimzu *et al.* 1985; Upadhyay *et al.* 2001).

Despite the recent interest in the use of stannates as ceramic dielectric bodies, humidity sensing, gas sensing, cell devices, and structural similarity with well understood alkaline earth titanates (Upadhyay *et al.* 2001), it is important that the effect of dopant on crystal structure, microstructure and thermal properties is studied. In this paper, we report the effect of substituting La^{3+} ions in SrSnO_3 element on the thermal diffusivity values. Characterisation of the ceramic samples have also been carried out using X-Ray Diffraction (XRD) and Scanning Electron Microscope (SEM). The purpose of the current investigation is to fill the information gap on literature about the doped- SrSnO_3 systems. Therefore, a series of solid solutions ($\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$) having different compositions of La^{3+} ($x = 0.1, 0.4, 0.6, 0.8$ and 1) were investigated.

THEORY OF FLASH METHOD

Thermal diffusivity, α is of direct importance in heat-flow studies, as it determines the rate of periodic or transient heat propagation through a medium. Because of its controlling effect and common occurrence in heat flow problems, its determination is often necessary, and the knowledge of the thermal diffusivity can in turn be used to calculate the thermal conductivity (Rosencwaig 1980). Thermal diffusivity is related to thermal conductivity (k) by a simple relation,

i.e. $\alpha = \frac{k}{\rho C_p}$, where ρ is the apparent density and C_p is the heat capacity at

constant pressure. The photoflash method is one of the most common ways to measure thermal diffusivity of solid sample. It is based on depositing a very short energy pulse on the front surface of a small disc-shaped sample, and calculating its thermal diffusivity from the characteristic curve (thermogram) of the temperature excursion of its rear surface (Parker *et al.* 1961).

If the initial temperature distribution within a thermally insulated solid of uniform thickness L is $T(x,0)$, the temperature distribution at any later time t is given by Carslaw and Jaeger (1959) as

$$T(x,t) = \frac{1}{L} \int_0^L T(x,0) dx + \frac{2}{L} \sum_{n=1}^{\infty} \exp\left(\frac{-n^2 \pi^2 \alpha t}{L^2}\right) \cos \frac{n\pi x}{L} \int_0^L T(x,0) \cos \frac{n\pi x}{L} dx \quad (1)$$

where α is the thermal diffusivity in cm^2/s .

For an opaque material, the temperature at the rear surface where $x = L$ can be expressed as (Parker *et al.* 1961).

$$T(L,t) = \frac{Q}{\rho C_p L} \left[1 + 2 \sum_{n=1}^{\infty} (-1)^n \exp\left(-\frac{n^2 \pi^2 \alpha t}{L^2}\right) \right] \quad (2)$$

where Q is the energy of the light pulse, ρ is the density, C_p is the heat capacity, L is the thickness of a sample, and t is time. The term in the square bracket can be expressed as

$$V = \left[1 + 2 \sum_{n=1}^{\infty} (-1)^n \exp\left(-\frac{n^2 \pi^2 \alpha t}{L^2}\right) \right] \quad (3)$$

which represents a function of temperature rise at the rear surface and varies from 0 to 1. The maximum temperature of the rear surface can be written as

$$T(L,t)_{\max} = \frac{Q}{\rho C_p L} \quad (4)$$

Hence, two dimensionless parameters, V and ω can be defined as

$$V(L,t) = \frac{T(L,t)}{T(L,t)_{\max}} \quad (5)$$

$$\omega = \frac{\pi^2 \alpha t}{L^2} \quad (6)$$

Plotting V against ω yields a dimensionless plot of the rear face temperature history which represents the theoretical ideal response of a material.

The value of V can vary from 0 to 1 and in principle, any fractional rise on the curve may be used to calculate thermal diffusivity. But the half rise time, $t_{1/2}$, ($V=0.5$) is usually employed for convenience since this point lies within the linear region of the curve enabling the fractional rise time to be determined more easily. Therefore when $V=0.5$, $\omega=1.37$ and the thermal diffusivity, α can be calculated as

$$\alpha = \frac{1.37L^2}{\pi^2 t_{1/2}} \quad (7)$$

where $t_{1/2}$ is the time required for the back surface of the specimen to reach half the maximum temperature rise.

Numerous corrections have been introduced to account for radiative heat losses during the process, the finite width of the energy pulse, and other factors interfering with the experiment. The effect of radiation heat loss at the rear surface can be calculated from the transient response curve by employing the

ratio technique. Heat loss corrections are based on Clark and Taylor rise-curve data or Cowan cooling-curve effect (Maglic and Taylor 1992).

MATERIALS AND METHOD

Thermal diffusivity analysis was carried out by a photoflash technique. The photoflash technique is a well-established method for measuring the thermal diffusivity with very simple procedures in a wide temperature range. The photoflash technique setup used in the present experiment is schematically shown in *Fig 1*.

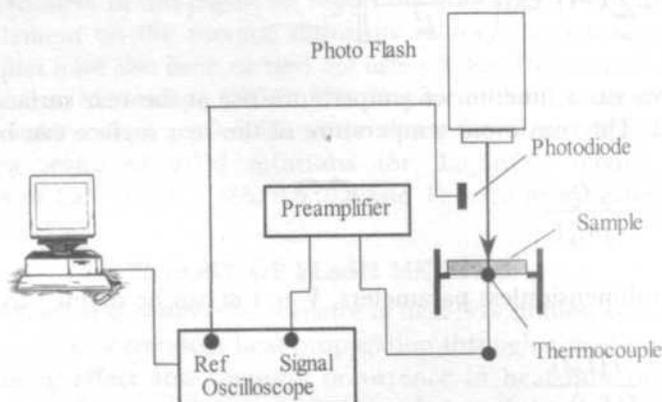


Fig. 1: Schematic experiment setup of the photoflash technique

A camera flash (Maxxum flash 5400HS) was used as an energy pulse source. It was located 2 cm in front of the sample as shown in *Fig. 1*. A photodiode was used to detect the pulse from the camera flash and subsequently trigger the oscilloscope. In this setup, a K-type thermocouple was used to monitor the temperature at the rear surface of the sample. The signal was then amplified by a preamplifier (SR560) and was transferred to a digital oscilloscope (Tektronix TDS 220). The signal was then analysed for the thermal diffusivity calculation using a personal computer. Every measurement was made at an interval of 10 minutes to allow the sample to be thermally equilibrated with room temperature. All measurements were carried out at room temperature.

Samples based on the formula (where $x = 0.1, 0.4, 0.6, 0.8$ and 1) were synthesised via conventional solid-state reaction. The starting compounds for preparation were LaO_3 (99.5%), from Sterm Chemicals, SrCO_3 and SnO_2 (purity 99.99 %) were from Aldrich Chemical. Stoichiometric amounts of respective compounds were weighed and milled in a ball mill for 24 hours with zirconia balls. Acetone was used as a mixing medium. The mixtures were dried and calcined at 1100°C for 24 hours in the air. The calcined powders were ground and mixed with 2-wt. % of water solution of polyvinyl alcohol (PVA) aqueous solution as a binder. The mixtures were dried overnight in the oven at 110°C , and was ground again, then finally pressed into pellets of 12 mm diameter, under a pressure of 5 tons. These pellets were kept in an alumina

crucible and sintered at 1400°C for 12 hours. Initially, the pellets were heated slowly at a rate of 1.5°C/min to 500 to burn off the binder; then the temperature was increased to 1400°C for 12 hours at a heating and cooling rate of 3°C/min. The dimension (thickness and diameter) of each sample was measured to be 0.1770-0.1902 cm for thickness and 0.12 cm for diameter. For the thermal diffusivity measurement, the Sr_{1-x}La_xSnO₃ samples were coated with a layer of carbon soot to enhance the absorption of flash energy.

RESULTS AND DISCUSSION

Fig. 2 (a), (b) and (c) show the thermogram for sample Sr_{1-x}La_xSnO₃ at x = 0.1, 0.4 and 1.0 respectively. From the curve, we could easily determine the $t_{1/2}$ for each of the samples. Here, we notice that the response time for Sr_{1-x}La_xSnO₃ at x = 1.0 is the fastest. It is known that the faster the propagation of heat into the medium, the higher the thermal diffusivity value. Fig. 2 also shows that heat loss occurred and it was corrected using Clark and Taylor rise-curve (Maglic and Taylor 1992). Therefore, the ratio $t_{3/4}/t_{1/4}$ is determined from the experimental data where $t_{3/4}$ and $t_{1/4}$ are the times to reach 75% and 25% of the maximum respectively. The correction factor, K_R will be calculated using:

$$K_R = -0.3461467 + 0.361578 \left(\frac{t_{3/4}}{t_{1/2}} \right) - 0.06520543 \left(\frac{t_{3/4}}{t_{1/2}} \right)^2 \quad (8)$$

Then the corrected value of thermal diffusivity at half rise time will be

$$\alpha_{corrected} = \frac{\alpha_{1/2} K_R}{0.13885} \quad (9)$$

The calculated ratios τ/t_c (pulse duration, $\tau = 5$ ms in our case) in the present experiment are much smaller than 1 ($\tau/t_c \ll 1$), therefore the finite pulse time in the present case is negligible (Cape and Lehman 1963). This characteristic rise time and the corrected value of thermal diffusivity of Sr_{1-x}La_xSnO₃ at different composition parameter x is tabulated in Table 1.

Fig. 3 shows the thermal diffusivity as a function of composition parameter x for Sr_{1-x}La_xSnO₃ (x = 0.1, 0.4, 0.6, 0.8, 1.0) samples. The thermal diffusivity values increase as more lanthanum is added to the sample and reaching a maximum value at x = 1.0. As stated by Shimizu *et al.* (1989), the concentration of electrons available for conduction in Sr_{1-x}La_xSnO₃ increased by the partial substitution of La³⁺ ions for Sr²⁺. This subsequently caused the thermal diffusivity values to increase on further increasing mole fraction x of La. Besides, the increase in thermal diffusivity should be caused by the changes in its interfacial structure, such as ordering of atomic arrangements and annihilation of interfacial defects (Qin *et al.* 1996). This is due to the thermal wave propagation in the solid materials being influenced by microstructure characteristics, size and

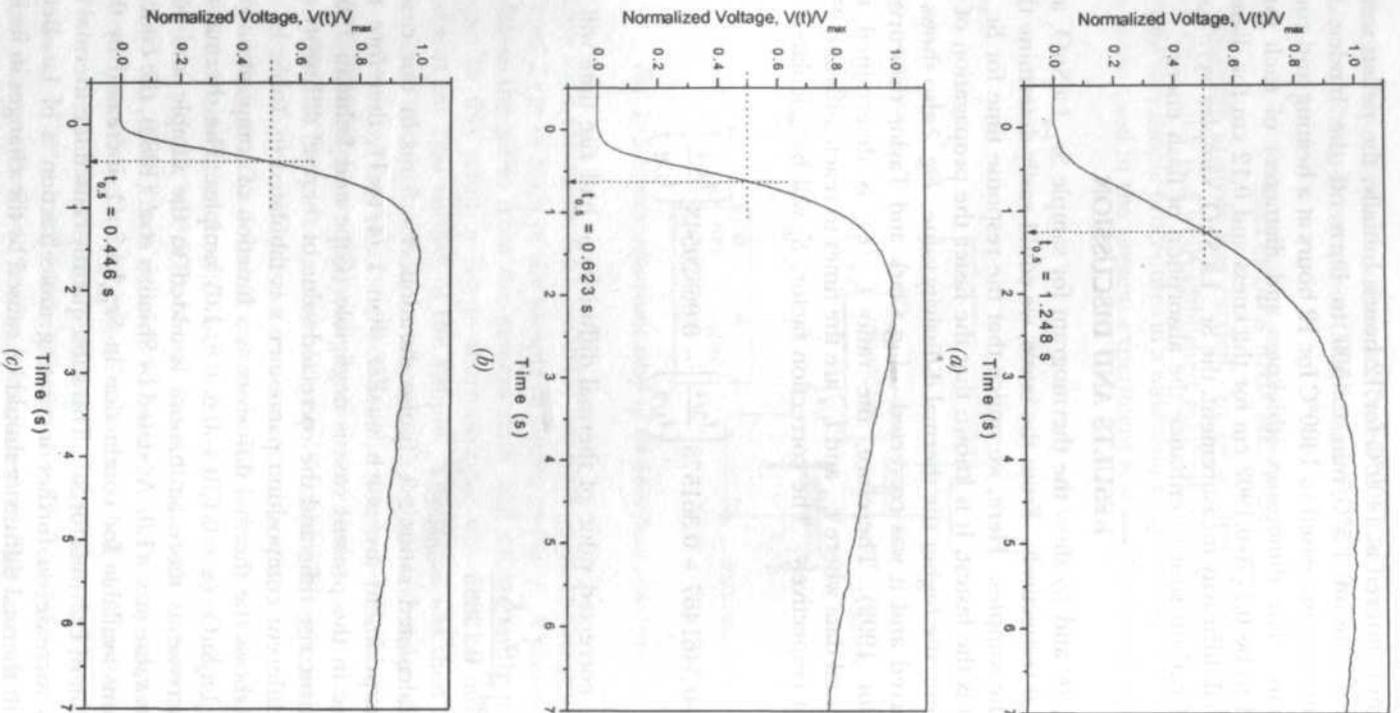


Fig. 2: (a) Thermogram for sample $\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$ at $x = 0.1$; (b) Thermogram for sample $\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$ at $x = 0.4$; (c) Thermogram for sample $\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$ at $x = 1.0$

TABLE 1
Characteristic rise time and the corrected value of thermal diffusivity
for Sr_{1-x}La_xSnO₃ at composition ranging from x = 0.1 to x = 1.0

Composition x in Sr _{1-x} La _x SnO ₃ sample	Thickness, l (cm)	Characteristic rise time, $t_c = l^2 / \alpha\pi^2$	τ / t_c	Corrected thermal diffusivity, $\alpha_{corrected}$ (cm ² /s) × 10 ⁻³
x = 0.1	0.1902	0.84	0.011	4.37 ± 0.01
x = 0.4	0.1841	0.48	0.019	7.15 ± 0.02
x = 0.6	0.1780	0.41	0.022	7.86 ± 0.04
x = 0.8	0.1770	0.37	0.024	8.65 ± 0.05
x = 1.0	0.1873	0.31	0.029	11.53 ± 0.06

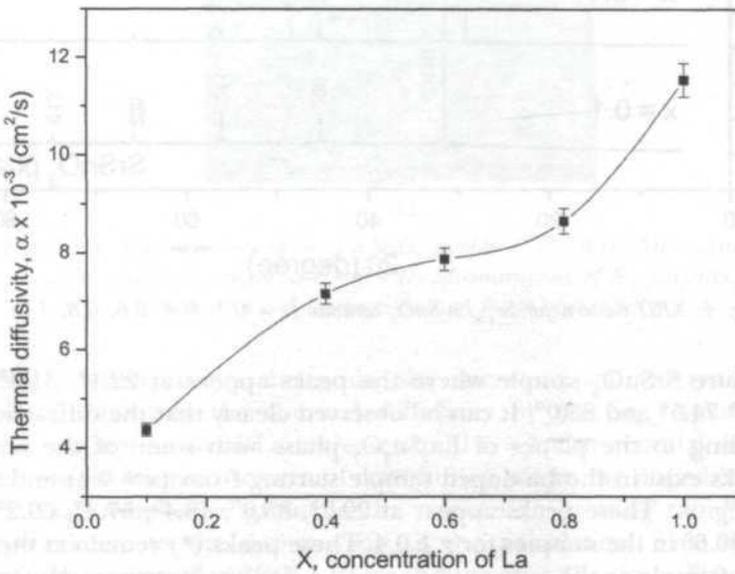


Fig. 3: Thermal diffusivity of Sr_{1-x}La_xSnO₃ at different compositions of x (sintered at 1400°C for 12 hours)

grain boundary, impurities and imperfection, where all of them contribute to the energy dispersion of the carriers like phonons and electrons inside the material (Lan *et al.* 1997).

The Sr_{1-x}La_xSnO₃ (x = 0.1, 0.4, 0.6, 0.8, 1.0) samples are characterised by a computer controlled x-ray diffractometer, XRD (Philips model 7602 EA Almelo) using Cu K α ($\lambda = 1.5418 \text{ \AA}$) irradiate with a scanning speed of 2°min⁻¹. The diffraction patterns of Sr_{1-x}La_xSnO₃ (x = 0.1, 0.4, 0.6, 0.8, 1.0) samples were obtained by scanning the samples with an interval of scanning angle (2 θ) from 5° to 85°. The XRD spectra of the Sr_{1-x}La_xSnO₃ (x = 0.1, 0.4, 0.6, 0.8, 1.0) are shown in Fig. 4. The peaks (2 $\theta = 22.3^\circ, 31.6^\circ, 45.2^\circ, 56.1^\circ, 65.8^\circ, 74.7^\circ$ and 83.3°) appearing in the (x = 0.1) sample remain unchanged if compared to the

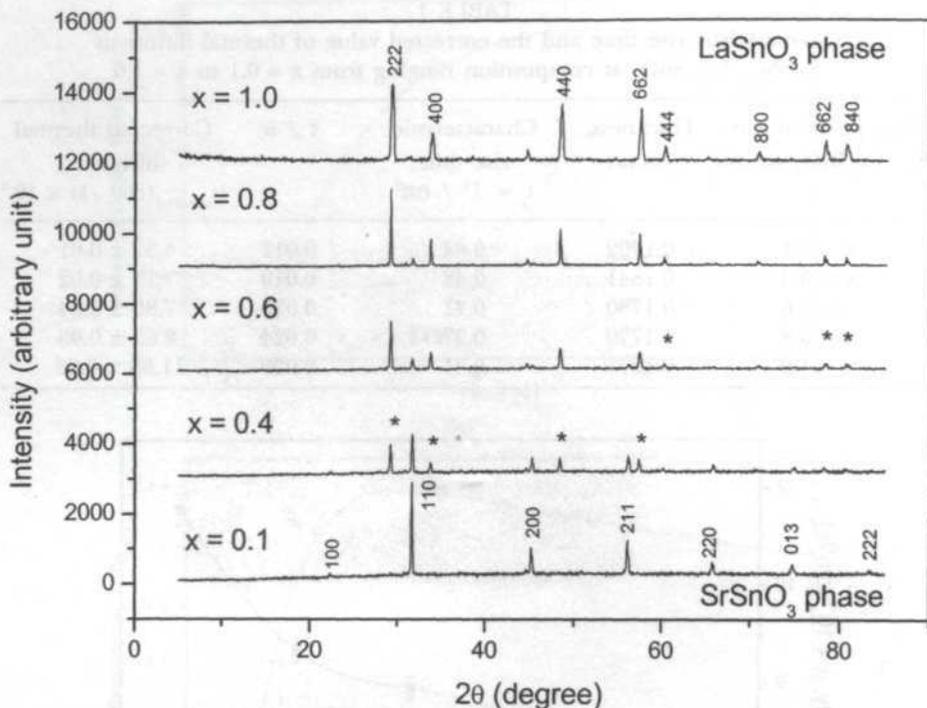


Fig. 4: XRD pattern for $Sr_{1-x}La_xSnO_3$ ceramic ($x = 0.1, 0.4, 0.6, 0.8, 1.0$)

standard pure $SrSnO_3$ sample where the peaks appear at $22.1^\circ, 31.4^\circ, 45.0^\circ, 55.9^\circ, 65.5^\circ, 74.5^\circ$ and 83.0° . It can be observed clearly that the diffraction lines corresponding to the planes of $La_2Sn_2O_7$ phase with some of the SnO_2 and La_2O_3 peaks exist in the La doped sample starting from ($x = 0.4$) and marked as (*) in Fig. 4. These peaks appear at $29.2^\circ, 33.8^\circ, 48.4^\circ, 57.4^\circ, 60.2^\circ, 70.8^\circ, 78.2^\circ$ and 80.6° in the samples for $x \geq 0.4$. These peaks (*) remain at the similar values of 2θ angle as the concentration of La doping increases. However, the intensity of the $La_2Sn_2O_7$ phase was found to increase as the lanthanum (La) concentration increased. It is also observed that the peak belonging to the $SrSnO_3$ diminishes as the La concentration increases. This changes of phase in $Sr_{1-x}La_xSnO_3$ ceramic samples affect the thermal diffusivity of a sample, where the sample with higher doping of La gives higher thermal diffusivity value.

Surface morphology of the $Sr_{1-x}La_xSnO_3$ ($x = 0.1, 0.4$ and 1.0) was investigated using a scanning electron microscope, SEM (JEOL model 6400). A uniform thin film of gold was evaporated on the exposed surfaces by a polaron coating machine to avoid electrostatic charging during microscopic viewing. The morphologies of the sample surface at magnification of 4,000X are shown in Fig. 5(a), (b) and (c). It indicates that the morphology of the $Sr_{1-x}La_xSnO_3$ at $x = 0.1$ and $x = 0.4$ is somehow different with the $Sr_{1-x}La_xSnO_3$ at $x = 1$. This SEM micrograph revealed that the materials were uniform, with spherical grains and have a small (sub- μm) grain size for $x = 0.1$ and 0.4 . The microstructure of the

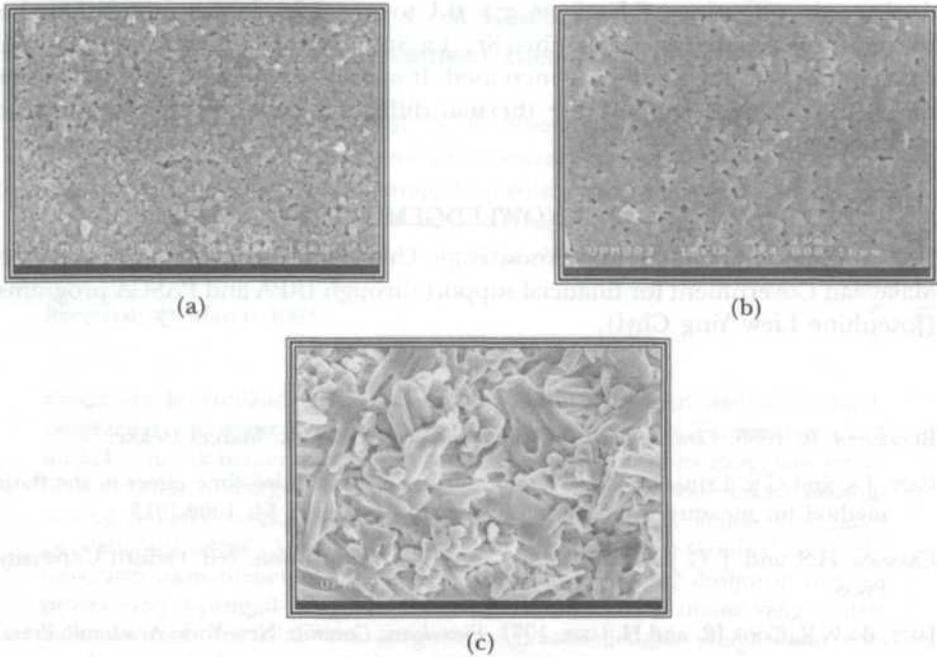


Fig. 5: (a) Microstructure of Sr_{1-x}La_xSnO₃ sample at x = 0 (b) Microstructure of Sr_{1-x}La_xSnO₃ sample at x = 0.4 (c) Microstructure of Sr_{1-x}La_xSnO₃ sample at x = 1 at magnification factor of 4,000X

Sr_{1-x}La_xSnO₃ system shows the grain size increases with increasing x and subsequently increases the thermal diffusivity of the Sr_{1-x}La_xSnO₃ sample. When the concentration of La increases until x = 1, the material underwent a drastic change in its microstructural constitution where the doped Sr_{1-x}La_xSnO₃ reveals a small number of grains in the population growing rapidly to a very large size relative to the average size of the population. This is called an abnormal grain growth (AGG) or secondary crystallisation. This is due to atomic diffusion across the grain boundaries and high curvature at the boundaries between the grains of different size. Finally, the individual grains occasionally take on highly unusual shapes (Loehman 1993; Kingery *et al.* 1976). This appearance shows a morphological change from the normal globular behavior of Sr_{1-x}La_xSnO₃ which explains why the thermal diffusivity value of Sr_{1-x}La_xSnO₃ at x = 1 increases drastically from x = 0.8.

CONCLUSION

The thermal diffusivity value of Sr_{1-x}La_xSnO₃ ceramic samples prepared at different doping concentrations of La (x = 0.1, 0.4, 0.6, 0.8 and 1.0) were investigated using a simple and inexpensive photoflash method at room temperature. The thermal diffusivity value of samples Sr_{1-x}La_xSnO₃ were determined to be $0.44 \times 10^{-2} \text{cm}^2/\text{s}$ to $1.15 \times 10^{-2} \text{cm}^2/\text{s}$ which correspond to the

doping concentration of La from $x = 0.1$ to $x = 1.0$. It was found that the thermal diffusivity value of the $\text{Sr}_{1-x}\text{La}_x\text{SnO}_3$ samples increased as the concentration of the La doping increased. It was also observed that results from the XRD and SEM support the thermal diffusivity behaviour in the present measurements.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Universiti Putra Malaysia and the Malaysian Government for financial support through IRPA and PASCA programs (Josephine Liew Ying Chyi).

REFERENCES

- BUCHANAN, R. 1986. *Ceramic Materials for Electronics*. New York: Marcel Dekker.
- CAPE, J.A. and G.W. LEHMAN. 1963. Temperature and finite pulse-time effect in the flash method for measuring thermal diffusivity. *J. Appl. Phys.* **34**: 1909-1913.
- CARSLAW, H.S. and J. C. JAEGER. 1959. *Conduction of Heat in Solids*. 2ed. Oxford University Press.
- JAFFE, B., W.R. COOK JR. and H. JAFFE. 1971. *Piezoelectric Ceramics*. New York: Academic Press.
- KINGERY, W.D., H.K. BOWEN and D. R. UHLMANN. 1976. *Introduction to Ceramics*. New York: John Wiley and Sons.
- LAN, T. T. N., H. G. WALTHER and D. T. SO. 1997. Photothermal estimation of thermal conductivity depth profiles in steel. *High Temp. High Pressures* **29**: 165-169.
- LOEHMAN, R. E. 1993. *Characterization of Ceramics*. London: Butterworth-Heinmann.
- MAGLIC, K.D. and R.E. TAYLOR. 1992. The apparatus for thermal diffusivity measurement by the laser pulse method. In *Compendium of Thermophysical Property Measurement Methods 2*, p. 281-314. New York, London: Plenum Press.
- PARKER, W.J., R.J. JENKINS, C.P. BUTLER and G.L. ABBOTT. 1961. Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. *J. Appl. Phys.* **32**: 1679-1684.
- PFAFF, G. 2000. Preparation and characterization of the strontium stannates SrSnO_3 and Sr_2SnO_4 . *J. Mater. Sci.* **35**: 3017-3021.
- QIN, X. Y., B. M. WU, Y. I. DU, L. D. ZHANG and H. X. TANG. 1996. An experimental study on thermal diffusivity of nanocrystalline Ag. *Nanostruc. Mater.* **7**: 383-391.
- ROSENCAWIG, A. 1980. *Photoacoustics and Photoacoustic Spectroscopy*. p. 265. New York, Chichester, Brisbane and Toronto: Wiley Interscience Publication.
- SHIMIZU, Y., M. SHIMABUKURO, H. ARAI and T. SEIYAMA. 1985. Enhancement of humidity sensitivity for perovskite-type oxides having semiconductivity. *Chem. Lett.* 917-920.
- SHIMIZU, Y., M. SHIMABUKURO, H. ARAI and T. SEIYAMA. 1989. Humidity-sensitive characteristics of La^{3+} -doped and undoped SrSnO_3 . *J. Electrochem. Soc.* **136**: 1206-1210.
- UPADHYAY, S., O. PARKASH and D. KUMAR. 2001. Solubility of lanthanum, nickel and chromium in barium stannate. *Mater. Lett.* **49**: 251-255.

Location Management Cost Reduction Using Adaptive Velocity-movement Based Scheme

M. S. Gembari, M. H. Habaebi, N. K. Noordin, B. M. Ali & V. Prakash

Department of Computer and Communications System Engineering,

Faculty of Engineering, Universiti Putra Malaysia,

43400 UPM Serdang, Selangor, Malaysia.

E-mail: nknordin@eng.upm.edu.my

Received: 21 March 2003

ABSTRAK

Rangkaian komunikasi peribadi tanpa wayar (PCN) terdiri daripada sebuah rangkaian tanpa wayar tetap dan sejumlah terminal bergerak. Terminal-terminal ini bebas untuk bergerak dalam kawasan PCN tersebut tanpa gangguan servis. Setiap terminal dengan tempoh tertentu akan melaporkan lokasi masing-masing kepada rangkaian melalui proses yang dipanggil kemas kini (atau pendaftaran) lokasi. Apabila panggilan tiba kepada sesuatu terminal bergerak, rangkaian akan menentukan lokasi sebenar terminal yang ditujukan melalui proses yang dipanggil pengeluaran terminal. Satu masalah utama yang timbul dalam senario ini ialah kos yang bersangkutan dengan pengeluaran dan pendaftaran. Beberapa kertas kerja cuba mengurangkan kos tersebut melalui skema-skema baru bagi pengeluaran dan pendaftaran. Salah satu skema yang menarik telah dibentangkan oleh Wan dan Lin (1998) yang mempertimbangkan skema kelua dinamik berdasarkan pada informasi kelajuan masa-separa nyata bagi setiap pengguna bergerak. Ini membolehkan jangkaan yang lebih tepat dibuat terhadap lokasi pengguna apabila sesuatu panggilan diterima. Dalam kertas kerja ini, penulis mengubah skema yang dibentangkan oleh Wan dan Lin dengan mencipta sebuah jangka kelajuan mudah suai yang akan berubah mengikut kelajuan terminal bergerak dan menggunakan analisis yang sama kepada skema berdasarkan pergerakan. Kajian ini menunjukkan bahawa kaedah yang dicadangkan oleh Wan dan Lin memberikan keputusan yang lebih baik daripada yang dilaporkan, dan pendekatan yang dicadangkan penulis membantu mengurangkan jumlah kos secara drastik berbanding skema asal. Keputusan juga menunjukkan nilai ambang pergerakan dan unit kelajuan masa mudah suai, memberikan penjimatan kos yang signifikan bagi saiz-saiz sel yang berbeza pada kelajuan sama ada tinggi dan rendah.

ABSTRACT

Wireless personal communication networks (PCNs) consist of a fixed wireless network and a large number of mobile terminals. These terminals are free to travel within the PCN coverage area without service interruption. Each terminal periodically reports its location to the network by a process called location update (or registration). When a call arrives for a particular mobile terminal, the network will determine the exact location of the destination terminal by a process called terminal paging. One major problem that arises in this scenario is the cost associated with paging and registration. Several papers in the literature attempt to reduce the cost by devising new schemes for paging and registration. One of the many interesting schemes was presented by Wan and

Lin (1998) that considers a dynamic paging scheme based on the semi-real time velocity information of an individual mobile user, which allows a more accurate prediction of the user location when a call arrives. In this paper, we modified the scheme presented by Wan and Lin by creating an adaptive velocity timer that changes according to the speed of the mobile and applies the same analysis to the movement-based scheme. The investigation shows that the proposed approach of Wan and Lin has better results than what was reported therein and our new approach helps reduce the total cost drastically compared to the original scheme. Results also show that the movement threshold and the adaptive velocity time unit, when they are adaptive, provide significant savings of cost under different cell sizes and velocities in high and low mobility systems.

Keywords: Paging cost, registration cost, location management, velocity, movement

INTRODUCTION

Current cellular networks such as Global Special Communications (GSM) partition their coverage area into a number of location areas (LAs). Each LA consists of a group of cells and each terminal reports its location to the network whenever it enters an LA. This reporting process is called location update. When an incoming call arrives, the network locates the mobile terminal by simultaneously polling all cells within the LA. This polling process is called terminal paging. Both the location update and the terminal paging processes require a certain amount of wireless bandwidth in addition to power consumption. As a result, costs are associated with both the location update and the terminal paging processes. It is clear that if each LA consists of only one cell, the network knows exactly the location of each terminal. In this case the cost of terminal paging is minimal. However, the cost of location update will be very high, as the terminal has to report its location whenever it enters a cell. A trade-off, therefore, exists between the location update cost and the terminal paging policy that can minimise the total cost. In this paper, it is proposed that the mobile terminal velocity is estimated in terms of velocity classes as opposed to the actual speed. Velocity classes' estimation greatly simplifies the computations and provides important robustness against the change of velocity (Wan and Lin 1998). It is also proposed that the network estimates to which velocity classes each mobile terminal belongs each time it updates. When a mobile terminal receives a call, the network checks its velocity classes, estimates how far it could go, and pages the candidate cells.

Most of the research work performed on paging and registration algorithms can be categorised into two approaches. The first one, called the group mobility approach, builds the paging and registration schemes on top of the system users' collective mobility pattern. The pattern is typically derived from the summary of statistical data collected by the system over time. A good example of this approach is the location area scheme used in current PCN systems (Rose 1996). These algorithms essentially used the mobility information about a

group of users to estimate an individual user's movement (Narayanan and Widom 1995; Kim *et al.* 1996; Vudali 1996). It is not surprising that these algorithms lack accuracy in their paging prediction and result in a rather large paging zone. The second approach, called the profile approach, recognizes the inherent problem with the group mobility approach and tries to use individual user profile to solve mobility (Silva and Brazio 1996; Rose and Yates 1997; Jannink *et al.* 1996). However it brings in significant management overhead into the system. These schemes require the collection of individual user movement statistics, user profile compilation, periodic profile update, and large profile storage. They are fairly complex to implement and put a great burden on system resources and system operators. The paging process relies on the location information provided by the registration process (Lin *et al.* 2000; Lin 1997; Akyildiz *et al.* 1996; Kim *et al.* 1996). The cost of paging also depends on the cooperating registration scheme. Generally, the more accurate location information the registration provides, the less cells the system needs to page. However, accurate location information does incur higher overhead in the form of registration cost. Therefore, a balance between the paging and registration schemes is necessary to achieve cost reduction. There are several registration methods such as location area registration, time based registration, distance based registration, movement based registration, predictive distance based update scheme, selective location update scheme and profile based scheme. The paging schemes are classified into two major types; delay bound and non-delay bound schemes. Delay constrained strategies are further classified as blanket polling and sequential group paging. Non-delay constrained strategies are sequential paging and shortest-distance-first. Examples of these paging methods are velocity-paging scheme, blanket polling paging scheme, location area paging, selective paging, shortest distance first scheme, sequential paging scheme, and velocity paging scheme.

BASIC VELOCITY PAGING AND VELOCITY CLASSES

The velocity paging approach is based on the method used in Wan and Lin (1998) and captures the semi-real time mobility characteristics of the individual mobile to provide more accurate prediction of the paging area. The scheme in Wan and Lin (1998) avoids the overhead of maintaining complex profiles for mobile terminals but still reduces paging cost significantly. Our scheme extends the velocity-movement based model and extends it by making the velocity information adaptive and a function of each mobile speed class instead of a fixed number for all. Similar proposals that minimise the paging and registration cost by utilising motion information to provide an optimal based solution have been reported in the literature (Rose 1995; 1996; Rose and Yates 1995; 1997). The objective of all these proposals is, due to the large number of users in a wireless system, to make computation for each registration/paging fairly simple. In addition, the procedures of the scheme should be easy enough to allow actual implementation. Therefore, our proposed approach focuses on the reduction of the computation intensity and the algorithm complexity.

In this section we will briefly discuss the velocity paging scheme with the movement based registration method and define the velocity classes and the relation between the velocity classes index and the adaptive velocity time unit based on velocity classes and the movement registration schemes.

Estimation of Mobile Velocity and Classes

Once in movement based registration schemes, only a counter is required to keep track of the number of movement made. The velocity classes then can be defined by the number of movements made during a predetermined time period, namely Velocity Time Unit (VTU). For a basic velocity-paging scheme, only the speed of a mobile terminal is considered. Paging prediction is sensitive to velocity changes or measurement errors. For example, if the VTU is 10 minutes, the movement threshold (given number of cells crossed by mobile before registering again), m , is 3, and a mobile terminal registers in time, t equals 15 minutes, with velocity class index, V_{cl} of 2 (Wan and Lin 1998) where

$$V_{cl} = (m/t) * VTU \quad (1)$$

In our modified scheme, we propose making the VTU timer adaptive and according to the velocity of the mobile itself. One way of doing this is as follows:

- a) The parameter t is basically the interval between two successive registrations by the mobile and therefore, it relates directly to the velocity of the mobile as a function of the distance traveled regardless of the exact movement pattern that the mobile has followed. If t decreases, it indicates that the mobile has increased its velocity and therefore it crossed m cells in time t . Therefore, t is basically the ratio between the distance traveled in terms of cells and the mobile velocity as follows:

$$t = \frac{2rm}{V_i} \quad (2)$$

where r is the cell radius, m is the movement threshold and V_i is the mobile actual speed. Knowing t would allow us to determine the velocity of the mobile and therefore we can determine its velocity class using Eq. (1).

- b) By predetermining the velocity classes of the mobiles and by replacing t in Eq. (1) by Eq. (2), the VTU timer can be made adaptive and specific to each mobile velocity. Therefore, VTU is found as follows:

$$VTU = \frac{2rV_{cl}}{V_i} \quad (3)$$

ALGORITHM

In a movement based registration scheme, each mobile terminal keeps a counter of movements and increases it every time it moves to another cell. The

mobile terminal registers its current location with the serving Visitor Location Registration (VLR). In order to cooperate with the modified velocity-paging scheme proposed, the following attributes are needed in the mobile's VLR record:

- 1) Last known location
- 2) The velocity class index
- 3) Last registration time

Accordingly, the following additional steps should be performed in the movement based registration procedure:

- a) Calculate the time span, t , since the last registration.
- b) When t is known, estimate the V_{cp} , V_i and VTU_i of the mobile terminal using Eq. (1), (2) and (3).
- c) Record the estimated V_{cp} , V_i and VTU_i in the VLR record, and update the last known location and the registration time.

When a call arrives to a standby mobile, the system will invoke the modified velocity paging mechanism:

- 1) The system asks the VLR for the mobile's velocity class index, its estimated velocity, its velocity time unit, its last known location, and its last registration time.
- 2) Assuming the mobile terminal remains in the same velocity class, the system calculates the maximum distance the mobile terminal could have traveled since the last registration. A candidate cell should be within this distance from the mobile's last known location.
- 3) The system identifies all the cells in the circular area by checking a system cell map and pages them.

ANALYTICAL MODEL

To evaluate our system analytically and to produce relevant cost measures, it is necessary to evaluate another system as a benchmark and compare it with our system. The most commonly used system in current cellular networks is the Location Area (LA) scheme where the cellular network is divided into clusters known as LAs and are made of several tiers of hexagonal cells. An example of a two-tier system for our proposed scheme is shown in Fig. 1. Note that in the figure the size and the number of tiers in the network are determined adaptively. The movement threshold m in our system will decide the number of tiers and therefore the size of the location areas. Some necessary assumptions are needed in order to evaluate the two systems such as the following:

- All cells are hexagons and of the same size.
- Registrations are performed when a mobile terminal is powered up and powered down.
- Movement classes similar to those described in Tabbane (1995) and Lam *et al.* (1996) are used to describe the mobility characteristic of user groups. Each movement class consists of a normal distribution of velocity and the percentage of total users in the class as given in Table 1.

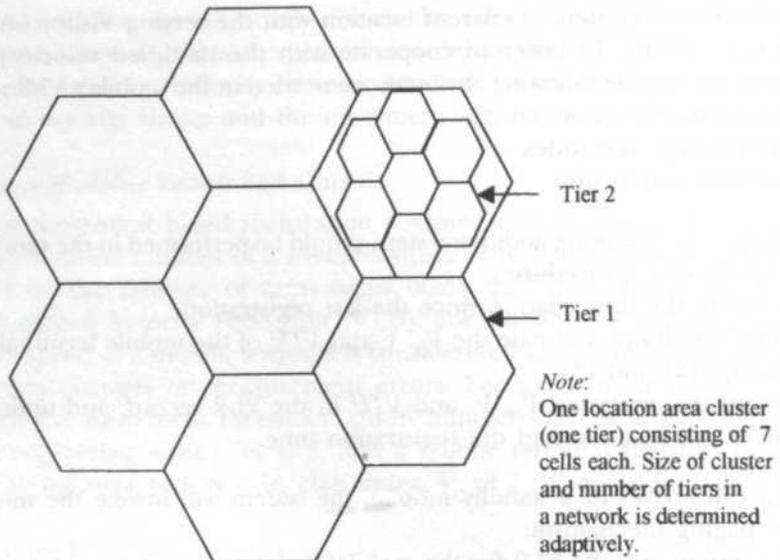


Fig. 1: Example of 2-tier model of the proposed modified velocity paging and restration scheme

TABLE 1
Movement classes for a high mobility system
in confidence interval presentation

Movement Class	Velocity (km/hr) (95%CI)	User Percentage
MC_1	0 - 10	25%
MC_2	35 - 80	74%
MC_3	110 - 150	1%

Location Area Scheme (LA)

Now given N movement classes, MC_1, MC_2, \dots, MC_N , each movement class, MC_i has its normal distribution of velocity, $N(V_p, \delta_i)$, and its user percentage, PMC_i . Basically a circular area covers approximately m tiers of cells with radius R_m . It is easy to prove that the m^{th} tier contains $6m$ cells. Therefore, the number of cells within and including the m^{th} tier is (Tabbane 1995):

$$N_m = 3m(m+1)+1 \tag{4}$$

Let the number of moves necessary for mobile terminal to travel from one cell to another be m . Since the radius of equivalent circular area, which has the same area as the hexagonal cells, is (Tabbane 1995).

$$R_m = \sqrt{N_m r_{eq}} \tag{5}$$

Then the radius of an equivalent circle, which has the same area as a hexagonal cell of side length r , is (Tabbane 1995):

$$r_{eq} = \sqrt{\frac{3\sqrt{3}}{2\pi}} r \tag{6}$$

- Paging cost

The total system paging cost is given by Wan and Lin (1998)

$$C_{pageTotal} = S\alpha N_m C_{page} \tag{7}$$

where

S is the total number of calls attempted per hour in the system

α is the percentage of mobile terminating calls

N_m is the average number of cells per location area

C_{page} is the paging unit cost for the LA scheme

- Registration cost

The total registration cost is given by Wan and Lin (1998) which was derived from Tabbane (1995) and Thomas *et al.* (1998) is:

$$C_{regTotal} = N_{Sub} \left(\beta - \frac{S}{N_{Sub}} T_{call} \right) \frac{8 \sum_{i=1}^N (V_i PMC_i)}{3\pi r \sqrt{N_m}} C_{reg} \tag{8}$$

where

S is the total number of calls attempted per hour in the system

N_{Sub} is the number of subscribers in the system

β is the power up probability for a mobile terminal

T_{call} is the average call duration

C_{reg} is the unit registration cost for the LA scheme

Modified velocity paging scheme (MVP)

- Paging cost of MVP

The total system paging cost for the modified velocity-paging scheme is (Wan and Lin 1998)

$$C_{pageTotal} = C'_{page} S\alpha \sum_{i=1}^N (N_{E[d_i]} PMC_i) \tag{9}$$

where

α is the percentage of mobile terminating calls

PMC_i is the user percentage of mobility class i

C'_{page} is the paging unit cost for the MPV scheme
 S is the total number of calls attempted per hour in the system
 $N_{E[d_i]}$ is the average number of moves made by a mobile MC_i of class i , in time t_p , and is given by (Wan and Lin 1998)

$$N_{E[d_i]} = \min \left(\frac{VTU_i \sqrt{12 \frac{V_i^2 + \delta_i^2}{r_{eq}^2} - 3 - 3}}{6} \frac{t_i}{VTU_i}, m \right) \quad (10)$$

where

m is the movement threshold

VTU_i is the velocity time unit for mobile MC_i found using Eq. (3)

V_i and δ_i is the mean and variance of the normal distribution for MC_i , respectively

- Registration Cost

The total registration cost of the MPV scheme is (Lin *et al.* 2000):

$$C'_{regTotal} = N_{Sub} \left(\beta - \frac{S}{N_{Sub}} T_{call} \right) C'_{reg} \sum_{i=1}^N \frac{\sqrt{12 \frac{V_i^2 + \delta_i^2}{r_{eq}^2} PMC_i}}{6} \quad (11)$$

where

C'_{reg} is the unit cost for registration in the MPV scheme.

NUMERICAL RESULTS

The velocity classes similar to those used in Wan and Lin (1998) are assumed for our results in order to facilitate comparison. The call arrivals are assumed to be Poisson processes and the registration activities will depend on the movements of the mobile terminals and the registration method used. Due to the extra computation incurred by the modified velocity-paging scheme, we used the conservative assumption of 10% cost increase for our scheme over the BVP and LA schemes as done in Wan and Lin (1998). Two systems, one with high mobility and the other with low mobility, were used in the study. The different mobility pattern in each system is modeled by a different set of movement classes. Tables 1 and 2 show the movement classes for the high and low mobility systems used in Wan and Lin (1998), respectively. The rest of the parameters are given in Table 3.

TABLE 2
Movement classes for a low mobility system
in confidence interval presentation

Movement class	Velocity (km/hr) (95%CI)	User percentage
MC1	0 - 10	50%
MC2	15 - 35	48%
MC3	40 - 80	2%

TABLE 3
Parameter values for performance comparison study

Parameter	Value	Comment
S	100000 calls per hour	Calls attempted in the system
α	30%	Mobile terminating call percentage
N_{Sub}	150,000	Number of subscribers
β	80%	Power up probability
T_{call}	2 minutes	Average call duration
VTU	10 minutes	Velocity time unite
C'_{page}	1	Unit paging cost for LA Scheme
C'_{reg}	2	Unit registration cost for LA Scheme
C_{page}	1.1	Unit paging cost for BVP Scheme
C_{reg}	2.2	Unit registration cost for BVP Scheme
m	3	Movement threshold

We have evaluated our modified velocity paging (*MVP*) scheme against the basic velocity paging (*BVP*) and *LA* schemes presented in Wan and Lin (1998). All the results of the combined costs of paging and registration for the two schemes are plotted against the cell radius (r) in km in *Figs. 2* and *3* for both mobility systems, high and low, respectively. It was highlighted in Wan and Lin (1998) that the *BVP* scheme tends to have a better performance compared to *LA* scheme except when $1.5 \text{ km} < r < 2.4 \text{ km}$. This is actually not true as illustrated by our results in *Fig. 2*. In fact the *BVP* scheme has a better performance in terms of cost savings compared to *LA* scheme for any value of $r > 2 \text{ km}$. These results are more accurate than those presented in Wan and Lin (1998) because the average paging zone size in the *BVP* scheme comes close to the average location area size in this range ($r < 2 \text{ km}$). The reduction in paging zone size cannot compensate the assumed 10% increase in unit costs and, thus, results in higher cost for the *BVP* scheme. What is more interesting is that our *MVP* scheme has a better performance record compared to *LA* and *BVP* schemes even for a cell radius size of $r > 1 \text{ km}$. As for the bigger paging zone sizes, our *MVP* scheme maintains the lowest total cost.

Our modified scheme concentrates on making the velocity time unit (*VTU*) timer adaptive and, thus as our analysis shows, the reduction is in the paging

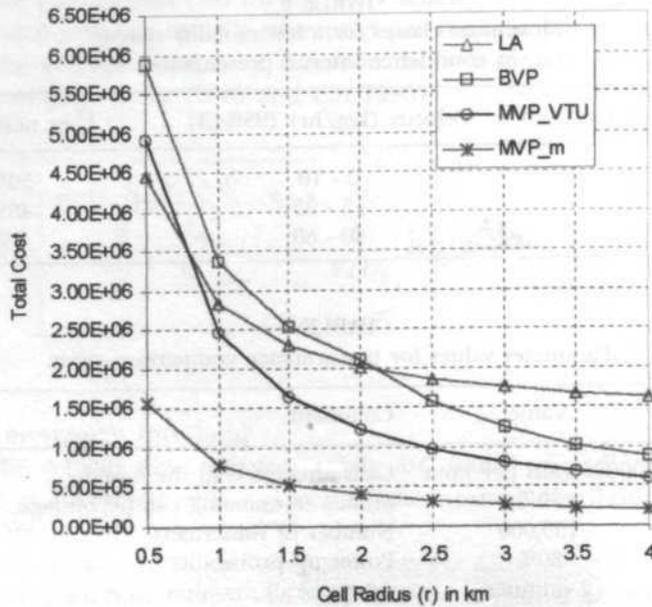


Fig. 2: Total system cost for high mobility class

cost only. Nevertheless, the improvement is very significant with no extra computational complexity on the implementation algorithm. The only difference between the *BVP* algorithm and the *MVP* algorithm is that with the *MVP*, the system will have to calculate the *VTU* parameter for every to-be-paged mobile, where as in the *BVP* scheme it considers a fixed value for the *VTU* parameter before it carries on the rest of the computations. Similar results are obtained for the low mobility system illustrated in Fig. 3 although in this case, our *MVP* system outperforms both *LA* and *BVP* systems for any paging zone size.

To further improve our system, one should consider making the movement threshold, m , adaptive as well as a function of the mobile velocity. We carried out a simulation work based on our theoretical study to understand the effect of the movement threshold on the total cost of the system and the results are presented in Fig. 4. The figure depicts the results of the total cost of the *MVP* scheme for several movement thresholds, m . It is interesting to note that for low to average estimated mobile velocities, V_i , an $m = 4$ value tends to give the lowest total cost for the system. And for the higher velocities, an $m = 10$ value keeps the *MVP* scheme at low total cost range. Keeping this in mind, we made the assumption of making the movement threshold toggle between two values for low and high mobility ranges. Thus, $m = 4$ if $0 \text{ km/hr} < V_i < 70 \text{ km/hr}$, and $m = 10$ for $V_i > 70 \text{ km/hr}$. Then we evaluated our *MVP_VTU* system with the new adaptive movement threshold and the results are plotted again in Figs. 2 and 3 and named *MVP_m* system to differentiate between the *MVP_VTU* that assumes a fixed value of $m = 3$ and the new *MVP_m* that uses both adaptive *VTU* and m values, for high and low mobility systems. The results illustrate the effect

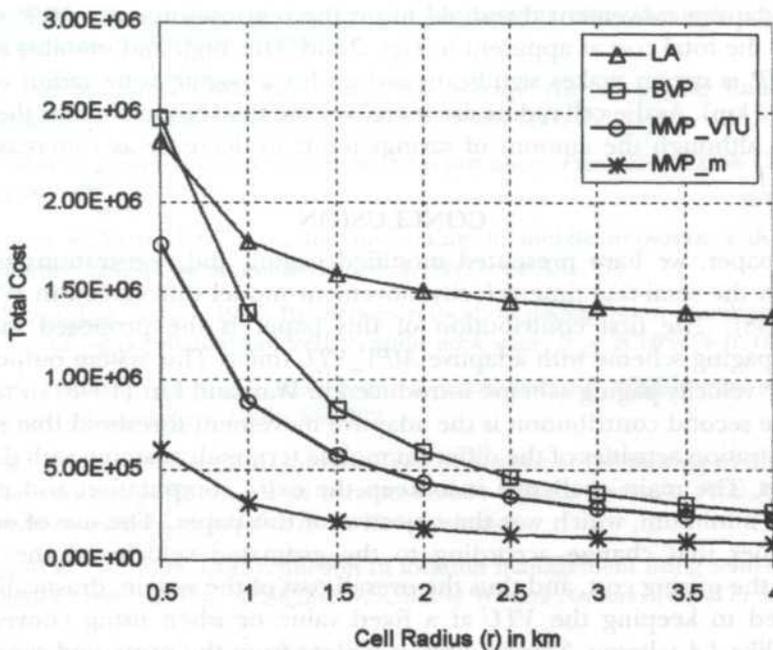


Fig. 3: Total system cost for low mobility class

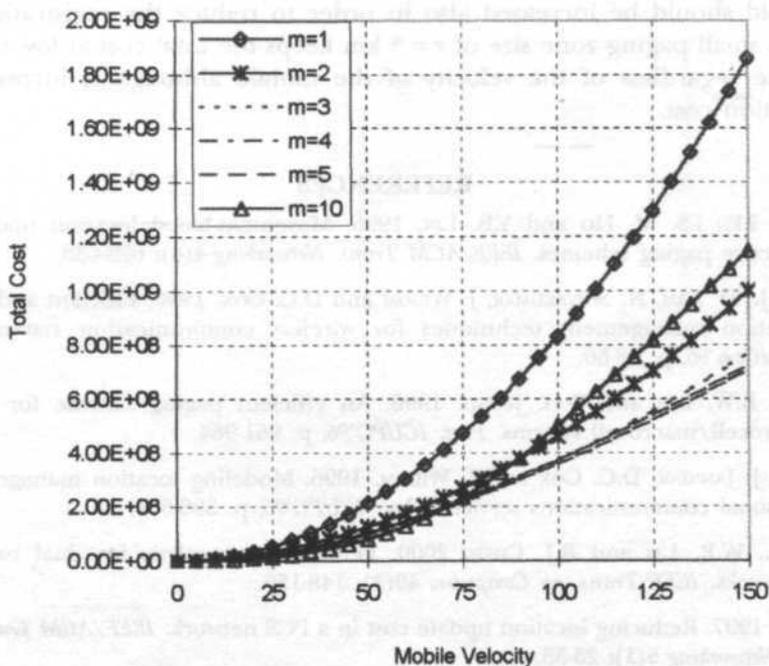


Fig. 4: Effect of movement threshold on total cost

of the adaptive movement threshold m on the registration cost (MVP_m) and, thus, on the total cost as apparent in Figs. 2 and 3 for high and mobility systems. The MVP_m system makes significant savings for a paging zone radius of small size ($r < 2$ km). As the cell radius size increases, the total cost maintains the lowest position although the amount of savings tends to decrease as r increases.

CONCLUSION

In this paper, we have presented modified paging and registrations schemes based on the semi-real time velocity-movement model introduced in Wan and Lin (1998). The first contribution of this paper is the proposed modified velocity-paging scheme with adaptive MPV_VTU timer. This system outperforms the basic velocity-paging scheme introduced in Wan and Lin (1998) in terms of cost. The second contribution is the adaptive movement threshold that reduces the registration activities of the different mobile terminals roaming with different velocities. The main challenge is to keep the extra computation and network cost at a minimum, which was the objective of this paper. The use of adaptive VTU values that change according to the estimated velocity of the mobile reduces the paging cost, and thus the overall cost of the system, drastically when compared to keeping the VTU at a fixed value or when using conventional systems like LA scheme. Nevertheless, it is clear from the presented results that a small movement threshold tends to reduce the total cost for low mobile velocity regardless of the cell size. As the velocity increases, the movement threshold should be increased also in order to reduce the registration cost. Using a small paging zone size of $r = 3$ km keeps the total cost at low levels all the time regardless of the velocity of the mobile although it increases the registration cost.

REFERENCES

- AKYILDIZ, I.F., J.S. M. HO and Y.B. LIN. 1996. Movement-based location update and selective paging schemes. *IEEE/ACM Trans. Networking* 4(4): 629-638.
- JANNINK, J., D. LAM, N. SHIVAKUMAR, J. WIDOM and D.C. COX. 1996. Efficient and flexible location management techniques for wireless communication systems, *Proc. MobiCom'96*, p. 38-50.
- KIM, D., B.W. LIM and D.G. JEONG. 1996. An efficient paging scheme for overlaid microcell/macrocell systems. *Proc. ICUPC'96*, p. 961-964.
- LAM, D., J. JANNINK, D.C. COX and J. WIDOM. 1996. Modeling location management in personal communications services. *Proc. ICUPC'96*, p. 596-601.
- LIN, Y.B., W.R. LAI and R.J. CHEN. 2000. Performance analysis for dual band PCS networks. *IEEE Trans. on Computers* 49(2): 148-159.
- LIN, Y.B. 1997. Reducing location update cost in a PCS network. *IEEE/ACM Transactions on Networking* 5(1): 25-33.
- NARAYANAN, S. and J. WIDOM. 1995. User Profile Replication for Faster Location Lookup in Mobile Environments. Technical Report, Stanford University.

- ROSE, C. 1995. Minimization of paging and registration costs through registration deadlines. *Proc. of ICC'95* p. 735-739.
- ROSE, C. and R. YATES. 1995. Minimizing the average cost of paging under delay constraints. *ACM-Baltzer Journal of Wireless Networks* **1(2)**: 211-219.
- ROSE, C. 1996. A greedy method of state based registration. *Proc. IEEE ICC'96 Conference*, p. 1158-1162.
- ROSE, C. and R. YATES. 1997. Location uncertainty in mobile networks: a theoretical framework. *IEEE Communications Magazine* **35(2)**: 94-101.
- SILVA, N.S. and J.M. BRAZIO. 1996. Reduction in location management signaling costs by the use of subpopulation tailored location area sizes. *Proc. ICUPC'96* p. 602-606.
- TABBANE, S. 1995. An alternative strategy for location tracking. *IEEE Journal on Selected Areas in Communications* **13(5)**: 880-892.
- THOMAS, R., H. GILBERT and G. MAZZIOTTO. 1998. Influence of the moving of the mobile stations on the performance of a radio mobile cellular network. *Proc. 3rd Nordic Seminar on Digital Land Mobile Radio Communications*.
- WAN, G. and E. LIN. 1998. Cost reduction in location management using semi-real time movement information. *ACM/Blatzer Journal of Wireless Networks (WINET)* **5(4)**: 245-256.
- VUDALI, M. 1996. The location area design in cellular and personal communications systems. *Proc. ICUPC'96*, p. 591-595.

INTRODUCTION

The goal of location management is to track the location of mobile stations in order to provide services such as mobile-to-mobile calls, paging, and other services throughout the network. Many research studies have been conducted in the area of location management in cellular networks. Keller et al. (1993) discussed the use of location management in IS-41. The authors discuss the drawbacks of changing the size of location areas and the use of adaptive location management. Keller and Yates

A New Congestion Control Algorithm for Active Networks

¹M.I. Buhari, ²M.H. Habaebi & ²B.M. Ali

¹King Fahd University of Petroleum and Minerals,
Dhahran, Saudi Arabia

²Department of Computer and Communications Engineering
Universiti Putra Malaysia
Serdang 43400 Selangor, Malaysia

Received: 9 April 2003

ABSTRAK

Makalah ini membincangkan dan menganalisis skim kawalan kesesakan yang berbeza untuk rangkaian-rangkaian aktif dan mencadangkan rangkaian baru. Skim yang dicadangkan mengenai satu keseimbangan antara bilangan kesilapan yang dipertimbangkan dan masa menunggu sebelum mengambil tindakan. Skim yang dilaksanakan bertindak ke atas konsep "Choke packet", iaitu sumber tersebut diarahkan untuk mengurangkan kadar transmisi. Pengurangan kadar transmisi yang munasabah dikekalkan pada sumber tersebut sambil nod perantaraan dalam pelantar jaringan aktif bertindak sebagai ganti biasanya nod matlamat bertindak. Skim yang dicadangkan disimulasi menggunakan OPNET dan keputusan simulasi dilaporkan. Keputusan menunjukkan bahawa teknik pelindapan sumber dilaksanakan dalam nod perantaraan jaringan aktif mengurangkan kesesakan secara drastik dan memperbaiki keseluruhan prestasi jaringan.

ABSTRACT

This paper discusses and analyses different congestion control schemes for active networks and proposes a new one. The proposed scheme strikes a balance between the number of errors considered and the waiting time before taking action. The scheme implemented works on "Choke packet" concept, where the source is instructed to reduce the transmission rate. Reasonable reduction of transmission rate is maintained at the source as the intermediate nodes in active networks platform take actions instead of the usually acting destination nodes. The proposed scheme was simulated using OPNET and simulation results are reported. Results indicate that the source-quenching technique implemented in the intermediate nodes of the active network reduces congestion drastically and improves overall network performance.

Keywords: Congestion control, source quench, choke packet, active network, OPNET

INTRODUCTION

The goal of congestion control is to control input traffic flows, so as to provide a reasonable tradeoff between packet losses and overall network throughput. Recently, many research findings have been reported in the area of congestion control in active networks. Rejaie *et al.* (1999) discussed the use of Rate Adaption Protocol (RAP) that suffers from the drawback of changing the rate of transmission at the source to half in case of congestion. Sisalem and Wolisz

(2000) discussed the Loss-Delay adaptation algorithm that uses Real-time Control Protocol (RTCP) for feedback messages. This algorithm too suffers from the slow response of RTCP that makes the congestion control scheme not fast enough to respond to the dynamics of active networks. TCP-Friendly Rate Control Protocol (TFRC) described by Floyd *et al.* (2000) starts slowly but initially takes actions on stabilizing the network when a single loss occurs. A combination of Window-based and Rate-based congestion control algorithms was proposed by (Rhee *et al.* 2000) where action is taken after some fixed number of errors has occurred instead of at every occurrence of error.

As the amount of processing increases due to parallel packet filtering concept being implemented in active networks, the probability of having congestion also increases. The fact is that the parallelisation of the access control rules increases the number of rules needed to represent the access list. To handle congestion, various congestion control schemes were analysed and discussed and a new scheme was proposed in this paper. The proposed scheme strikes a balance between number of errors considered and the time required to handle errors or congestion. The scheme implemented works on the "Choke packet" concept to handle congestion, where the source is instructed to reduce the transmission rate whenever there is congestion due to huge traffic. Reasonable reduction of transmission rate is maintained at the source as the intermediate nodes in an active networks platform take action instead of the usually acting destination nodes.

Related Work

Congestion control schemes can be of two types, namely, Window-Based & Rate-Based. Window-based category uses a congestion window either at the sender or the receiver end. Similar to TCP, each packet transmitted consumes one slot in the congestion window, while each packet received or the acknowledgment of a packet received frees one slot. The sender is allowed to transmit packets only when a free slot is available.

Rate-based category adjusts the transmission rate dynamically according to some network feedback mechanism. It can be subdivided into simple AIMD (Additive Increase, Multiplicative Decrease) schemes and model-based congestion control. Simple AIMD schemes mimic the behaviour of TCP congestion control. This results in a rate that displays the typical short-term saw-tooth like behaviour of TCP. This makes simple AIMD schemes unsuitable for continuous media streams. Model-based congestion control uses a TCP model by adapting the sending rate to the average long-term throughput of TCP. Model-based congestion control can produce much smoother rate changes that are better suited for continuous traffic (Widmer *et al.* 2001).

As in the case of Packet filtering, the usage of different access lists can vary the processing and in turn the impact of congestion on the network varies. Thus, we use the rate-based congestion control scheme. This helps us to change the rate dynamically when there is some change of access lists and so on. As the traffic is not continuous, we choose to use the simple AIMD form of the Rate-based category.

The following parameters are considered to compare the performance of congestion control scheme:

1. Decision Function: This indicates when the change in transmission rate is made. What is the impact of the protocol with and without congestion existence?
2. Decision Frequency: How frequently the changes are made. Do these changes impact the total performance of the network?
3. Stability: After a network undergoes some perturbation because of flows joining and leaving, if the network reaches steady state, no matter what the state of the protocol at the end of the perturbation is, the protocol eventually reaches the fair and TCP-friendly rate.
4. Scalability: The performance of a protocol for instance does not depend on the number of its receivers.

Various Congestion Control Schemes:

We have considered many rate based congestion control algorithms. The concepts behind those congestion control schemes and the impact of decision function and decision frequency on these congestion control schemes are discussed below.

1. The Rate Adaptation Protocol (RAP) (Rejaaie *et al.* 1999) is a simple AIMD scheme for unicast flows. The receiver acknowledges each data packet. The ACKs are used to detect packet loss and infer the round-trip time. When the protocol experiences congestion it halves the sending rate. In periods without congestion, the sending rate increases by one packet per round-trip time, thus mimicking the AIMD behaviour of TCP. The decisions on rate increase or decrease are made once per round-trip time. While TCP congestion control is appropriate for applications such as bulk transfer, some real-time applications (that is, where the data is being played out in real-time) find halving the sending rate in response to a single congestion indication to be unnecessarily severe, as it can noticeably reduce the user-perceived quality (Tanenbaum 1996). Thus, the impact of reducing the data transfer abruptly disturbs the network.

With regard to the decision frequency, changes are made every round-trip time (RTT). Due to the random nature of RTT, using the recent sample RTT as the step length is likely to result in a poor behaviour. RAP connections with shorter RTT are more aggressive and achieve a larger share of the bottleneck bandwidth. However, there are two issues to notice. First, in general, other measures of fairness can only be achieved by implementing the required machinery in the network. Second, as long as the unfairness problem is not resolved along TCP flows, being TCP-friendly implies accepting this unfairness.

2. Loss-Delay Adaptation Algorithm (Sisalem and Wolisz 2000) does not devise its own feedback mechanism to control the sending rate but relies

solely on the Real-Time Transport Control Protocol (RTCP) feedback messages provided by the Real-Time Transport Protocol (RTP). The increase and decrease factors for AIMD are dynamically adjusted to the network conditions. The amount of additive increase is then determined as the minimum of three independent increase factors to ensure that (i) flows with a low bandwidth can increase their rate faster than flows with a higher bandwidth; (ii) flows do not exceed the estimated bottleneck bandwidth; and (iii) flows do not increase their bandwidth faster than a TCP connection. The authors claim that the slow rate increase of LDA+ compensates for the fact that the rate is at most decreased to the throughput estimate of the complex TCP model. However, RTCP reports are generated infrequently (usually within several seconds). This makes LDA+ slow to react to changes in the network conditions. While looking at the decision function, we infer the following:

- LDA+ uses the real-time transport protocol (RTP) for collecting loss and delay statistics that are then used for adjusting the transmission rates of the senders in a manner similar to TCP connections suffering from equal losses and delays.
- With no observed losses, the sender can increase its transmission rate additively, otherwise it needs to reduce it multiplicatively. For the case of no losses, the flow's share can be increased by a value that does not exceed the increase of the bandwidth share of a TCP connection with the same round trip delay and packet size.

The in-frequency of the RTCP feedback messages dictates that an RTP sender cannot benefit fast enough from rapid changes in the network conditions. Thus, the goal of RTCP-based adaptation is to adjust the sender's transmission rate to the average available bandwidth and not react to rapid changes in buffer lengths of the routers.

3. TCP-Friendly Rate Control Protocol (TFRC) (Floyd *et al.* 2000) starts with a slow start phase but stops the phase at first loss event. The sender then computes a new fair rate and changes the sending rate. Several requirements for a loss rate estimator are formulated and the authors settle on the Average Loss Interval method which best fulfills these requirements. Once per round-trip time the TFRC receiver updates its parameters and sends a state report to the sender. The sender then computes a new fair rate from these parameters and adjusts the sending rate accordingly. A major advantage of TFRC is that it has a relatively stable sending rate while still providing sufficient responsiveness to competing traffic. As in Packet filtering, the network setup is made with proper care on the access lists and the access lists do not change often, it is not necessary to change the sending rate often. While looking at the decision function, we infer the following:
 - The sender explicitly adjusts its sending rate as a function of the measured rate of loss events, where a *loss event* consists of one or more packets dropped within a single round-trip time. Loss event rate is

different from loss packet rate because the loss impact on successive packets is counted as an event.

- In response to a decrease in the loss event rate, the sending rate is increased slowly. Similarly, do not halve the sending rate in response to a single loss event. However, in case of successive loss events, halve the sending rate.

With regard to the decision frequency, the receiver should report feedback to the sender at least once per round-trip time if it has received packets in that interval. If the sender has not received feedback after several round-trip times, then the sender should reduce its sending rate, and ultimately stop sending altogether.

4. TCP Emulation At Receivers (TEAR) (Rhee *et al.* 2000) is a hybrid protocol that combines aspects of window-based and rate-based congestion control. TEAR receivers calculate a fair receive rate which is sent back to the sender, who then adjusts the sending rate. To this end, the receivers maintain a congestion window that is modified similarly to TCP's congestion window. In contrast to TCP, the TEAR protocol does not directly use the congestion window to determine the amount of data to send but calculates the corresponding TCP sending rate. While looking at the decision function, we infer the following:

- If there is no congestion at slow-start or congestion-avoidance state, then the number of packets on transit from source to destination (congestion window size or *cwnd*) is increased by $1/\text{Last } cwnd$.
- Upon loss of packets, the *cwnd* is not modified. The reception of three packets after the losses will trigger three duplicate acknowledgments in TCP (assuming no delayed acknowledgments).

With regard to the decision frequency, changes are made when there is a loss of packet. The receiver takes the actions here.

5. Generic router assist (GRA) (Widmer *et al.* 2001), for instance, is a recent initiative that proposes general mechanisms located at routers to assist transport control protocols, which would greatly ease the design and implementation of congestion control protocols.

From the above discussion, it is clear that the change in the transmission rate must not be as often as in RAP or as slow as in LDA+ protocol. The stability of the active network should not be affected by having the processing at the intermediate nodes. To implement congestion control in active networks platform, we incorporated the following steps:

- i. In the initial stage of the network, the network starts with a slow-start Algorithm. The slow-start stage is similar to that of the TEAR algorithm.
- ii. The intermediate processing node (router) realises late arrival of the packets or out-of-sequence order of packets and then enters a stage of no more increase in the CWND. In case the situation of loss continues then

the intermediate node sends a Source Quench (SQ) message to the source indicating to it to reduce the rate of transmission.

EXPERIMENTAL SETUP

Slow-start Algorithm

When a TCP connection starts up, the TCP specification requires the connection to be conservative and assume that the available bandwidth to the receiver is small. TCP is supposed to use an algorithm called slow start to probe the path to learn how much bandwidth is available (Partridge and Shepard 1977; Lakshman *et al.* 1977). A simple description of the slow-start algorithm is as follows:

- (i) Send a TCP packet with one data segment and wait for an acknowledgement.
- (ii) Upon receiving the acknowledgement, send 50% more data every round trip.
- (iii) Do step 2, until a packet is lost. If so, stop the increase in data size.

Algorithm Design

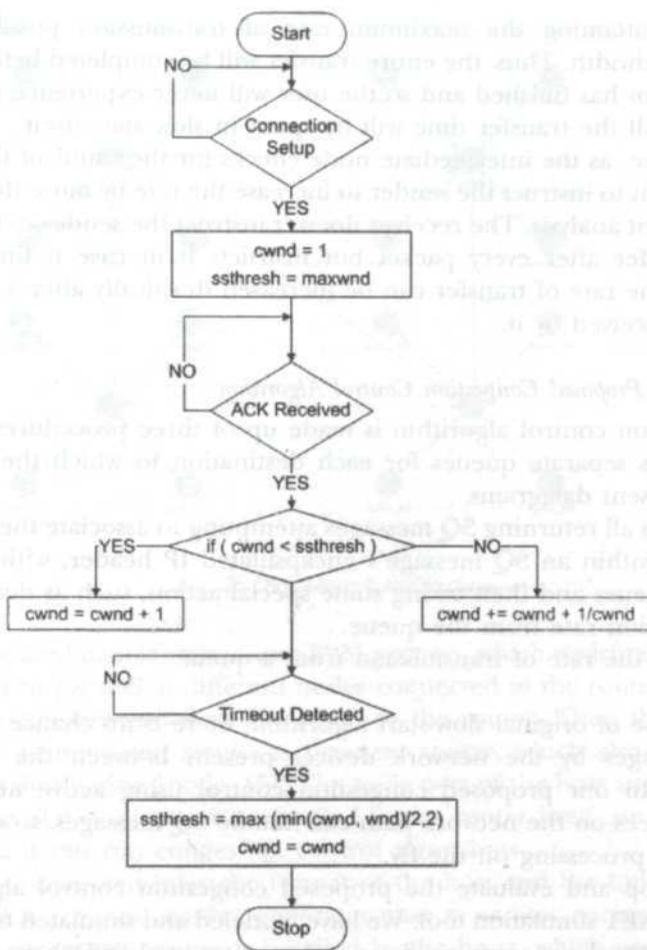
In the case of Slow-start, a back-check packet of full-window size is sent by the sender before it decides about the time-out. The congestion window (*cwnd*) is initially set to 1 and the incrementing is done exponentially. According to Wang and Crowcoft (1992), the congestion avoidance algorithm in fact is a time-out based multiplicative decrease/additive increase algorithm. When a timeout is detected, the threshold *ssthresh* is reduced to half of the current window size and then *cwnd* is stopped from increasing further.

During the slow-start phase, each acknowledgement increases *cwnd* by 1. After *cwnd* reaches *ssthresh*, *cwnd* is increased by $1/cwnd$ for each acknowledgement received (Wang and Crowcoft 1992). The following flowchart summarises the algorithm.

The actual sending window size (*wnd*) would not exceed the receiving size advertised by the receiver (*maxwnd*), therefore the sender always sets

$$wnd = \min(cwnd, maxwnd);$$

Networks should implement a feedback mechanism that allows traffic sources to control their transmission rate. Within the TCP/IP protocol suite, the ICMP Source Quench (SQ) message is created to provide the feedback from any overloaded gateway or host, back to the source. An ICMP source quench message is intended as a congestion control mechanism in IP. Source quench messages are used when a network gateway cannot forward a message because its message buffers are full. The gateway transmits a source quench message back to the source host machine to request that the source reduces its transmission rate until it no longer receives source quench messages from the gateway. Thus, this effectively throttles back the source's transmission rate. The gateway can transmit multiple source quench messages, one for each packet it receives from a source. The source machine is not required to respond to these source quench messages. When a source receives an SQ message, in principle it should decrease its rate of transmission. In practice, this does not happen (Finn 1995).



When the SQ message arrives, the rate of transmission is decreased. After some time elapses, the source could return to a higher rate of transmission. However, it is safe to keep it lower than the rate that it was using when it first received the SQ message and then slowly allow it to rise. Two phases are present in the case of the Congestion Control algorithm. They are the slow-start phase and the additive phase. In the slow-start phase, the rate of transmission (r) is reduced to a lower value. Then the rate r is allowed to grow. This phase concludes when the rate attains a threshold or another SQ message causes rate r to be reset. The additive phase commences after the rate r reaches the threshold. In this phase, the value of r grows but much more slowly.

Remedy for the Slow-start State Disadvantage

With slow-start algorithm, the rate of transmission is initialised to a slower value and then the rate of transmission is allowed to increase based on response from the destination. There is always a possibility that data transfer would complete

well before attaining the maximum rate of transmission possible on the available bandwidth. Thus, the entire transfer will be completed before the slow start algorithm has finished and so the user will never experience the full link bandwidth. All the transfer time will be spent in slow start itself.

In our case, as the intermediate node checks for the status of the network, it has the right to instruct the sender to increase the rate by more than one step based on event analysis. The receiver doesn't instruct the sender to increase the rate of transfer after every packet but instructs it in case it finds that the increase in the rate of transfer can be increased drastically after a few packets have been received by it.

Procedures for Proposed Congestion Control Algorithm

The congestion control algorithm is made up of three procedures as follows:

- i. Maintains separate queues for each destination to which the source has recently sent datagrams.
- ii. Examines all returning SQ messages attempting to associate the destination address within an SQ message's encapsulated IP header, with one of the above queues and then taking some special action, such as decreasing the transmission rate from the queue.
- iii. Controls the rate of transmission from a queue.

In the case of original slow-start algorithm, there is no chance of handling of SQ messages by the network devices present between the source and destination. In our proposed congestion control using active networks, the network devices on the network path can handle SQ messages, so we handle it using queue processing on the fly.

To develop and evaluate the proposed congestion control algorithm, we used the OPNET simulation tool. We have modeled and simulated two scenarios where the network processing is done at the router level and where the network processing is done at the host level.

Simulation Setup

The set-up consists of a 16-node network designed and simulated using OPNET. Our network model shown in *Fig. 1* consists of 4 subnets each comprising 4 nodes. We made use of OPNET's random packet generation algorithm at each of the nodes. We implemented a static routing algorithm that fixes the next router through which packets can be forwarded from the current one based upon the destination subnet and the host address present in the packet. Secondary routes are assigned for various routes.

As an example, let us suppose that the result of some computation is to be transmitted from node_0 to node_15. The data can pass either through hubs [0, 2 and 3] or [0,1 and 3] (*Fig. 1*). Parallel virtual machine (PVM) resides in active routers (hubs) and also in all the nodes. When data is passing through the first router, the router will partition the computation to be performed on the data into many modules based on availability of nodes. Each module along

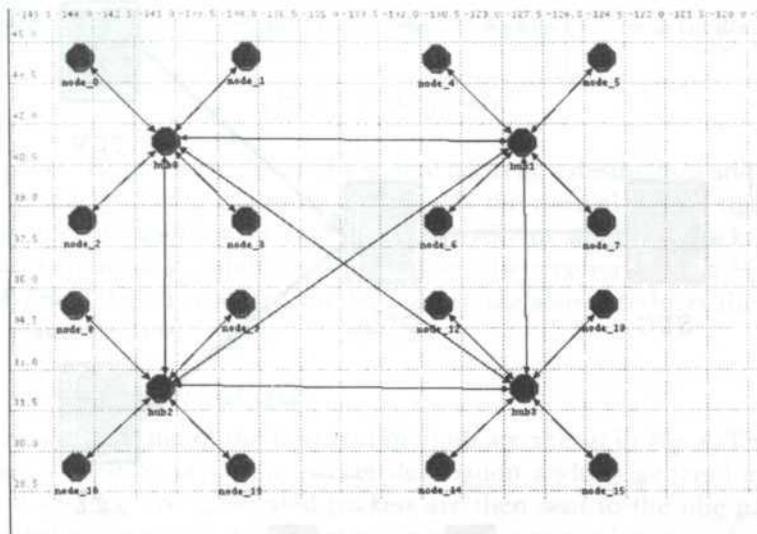


Fig. 1: Our network architecture

with its associated data is sent to one PVM process, which is virtually connected and running in parallel at different nodes connected to the router. The nodes will process the data and send the result to the router. Then the router will forward the accumulated results to the next router, which also forwards the result to the destination [node_15]. The node part of the host and the hub are shown in Fig. 2. Congestion is controlled in the router itself, since the router is active and it can run congestion control algorithms.

From Fig. 2, we can infer the format of the host and the hub. The hub is made up of a constant packet generator that generates packets at uniform speed. The packet generation is handled by the hosts, which possess the ideal packet generators (*src* in Fig. 2) in them. The processing is done on the hub and the packet is sent via the transmitter (*xmt*). Packets enter the hub through the receiver (*rcv*) node present. In the case of the node model of the hub, there exists a receiver and a transmitter for each connection from/to the hub. These connections are called streams. As each hub is connected to three other hubs and four nodes, the node diagram shows seven transmitters and seven receivers. A pair (transmitter and receiver) is connected together in logical form that is shown by the dotted lines.

The packet format for our capsule oriented active networks system is shown in Fig. 3 and is made up of the following:

The source, destination addresses and the protocol are used for the access control permit or denial decision on the packet. The packet type is used to identify whether the packet is a data packet or a control packet. The timestamp is used to calculate the time taken by the packet to travel from the source to destination. The file name or result and data fields respectively contain the name of the file or the result of access control processing and the data carried by the capsule.

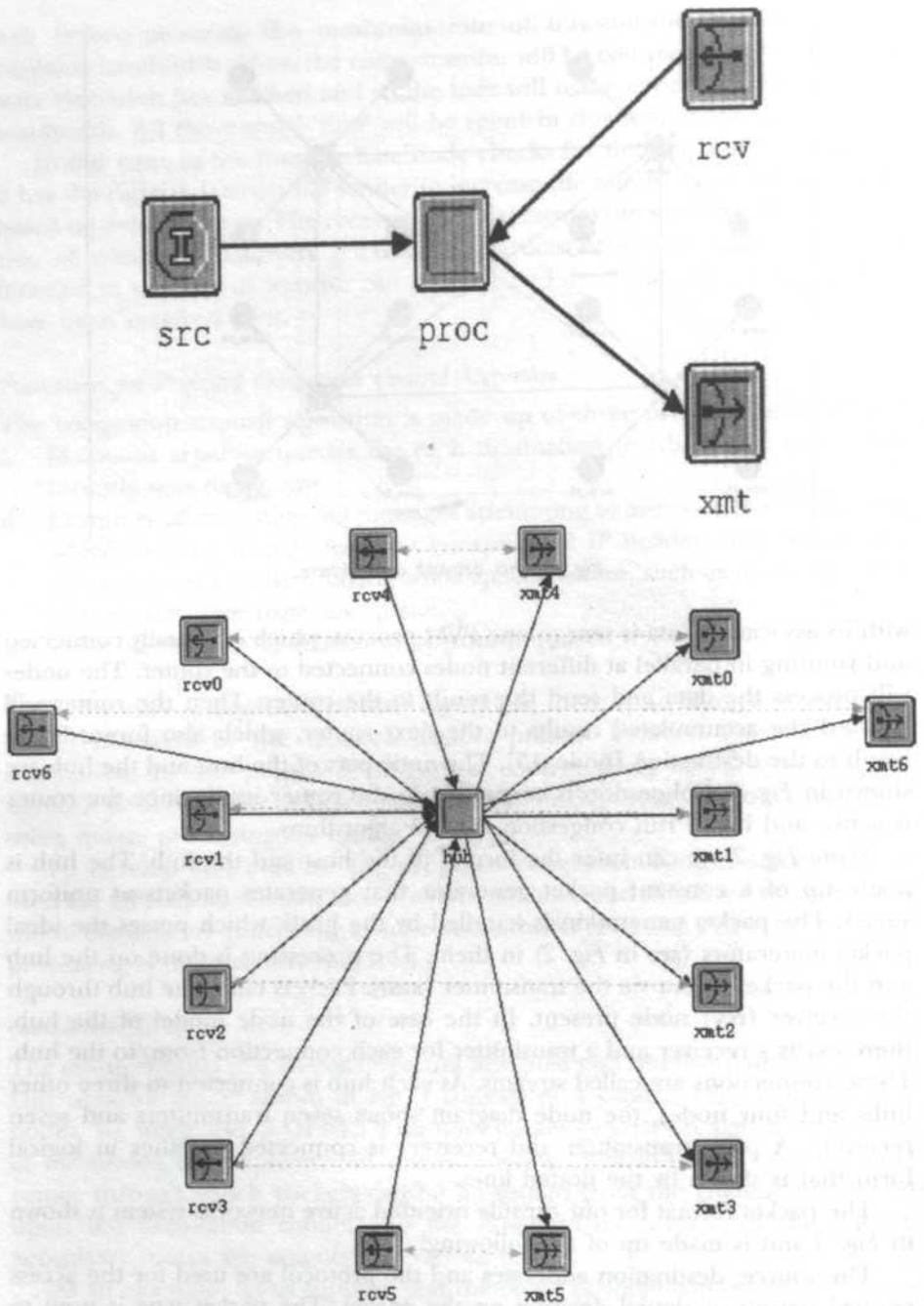


Fig. 2: Node diagram of the host and the hub

Source Address	Destination Address	Protocol	Packet Type	Timestamp	File Name OR Result	Data
----------------	---------------------	----------	-------------	-----------	---------------------	------

Figure 3. Packet Format

The packet format is made up of a source address, a destination address and the type of packet being transmitted. *Type 0* is the normal packet that travels from one node to another. In case of heavy traffic or when the packet arrival rate is more than the capability of the receiver, the receiver sends a SQ packet that is of *Type 1*. Upon receiving the SQ packet, the source reduces the pace at which it sends packets.

Processing Scheme of Host and Hub

The processing scheme of the host and the hub are shown in Fig. 4. The host's "init" part is used to make the packet destination address generation to fall between 0 and 15. The generated packets are then sent to the idle part. The transmitter then transmits the packet along any one transmit stream depending upon the destination address specified. In Fig. 4, the SRC_ARRIVAL indicates the activation of packet transfer from the idle state to the transmitter, upon arrival of a packet that needs to be transmitted. Moments after transmission of the packet, the state comes back to the idle state and waits for transmission or receiving of packets. The 'xmt' state is the state where the processing that ought to be done before and during the transmission of a packet is specified. On receiving a packet, the RCV_ARRIVAL is activated and the packet is processed by the routines specified in the 'rcv' state.

At the hub, the checking for the destination address is done by the routines specified in the 'route_pk' state and the packet is forwarded in the respective transmitter stream. The activation of the 'route_pk' state is done by the PK_ARRIVAL activation link. The processing at the hub plays a role in the calculation of RTT due to the fact that there is some active process at the hub that consumes some time. Once the packet reaches the destination, the packet is now processed to identify the time it has taken to reach the destination from the source. Depending upon the time taken an SQ message is generated. If the delay in transmission is for the first time, then the transmission rate is ignored. But if the delay continues then the SQ message is sent to the source.

To identify network congestion, we normally calculate the time taken for the packet to move from the source to the destination. Instead, using active networks, we are concerned with the next intermediate router between the source and the destination. If it takes a long time (which can be determined based on the distance and bandwidth between the node and next intermediate router or hop), it indicates that the network is congested and there is more likelihood of collision and loss of packets. The time calculation between the source and destination not only includes the time taken to travel from source to destination but also the processing that occurs at the hub. To handle this case, the calculations of time at the hubs before the receiving and sending of

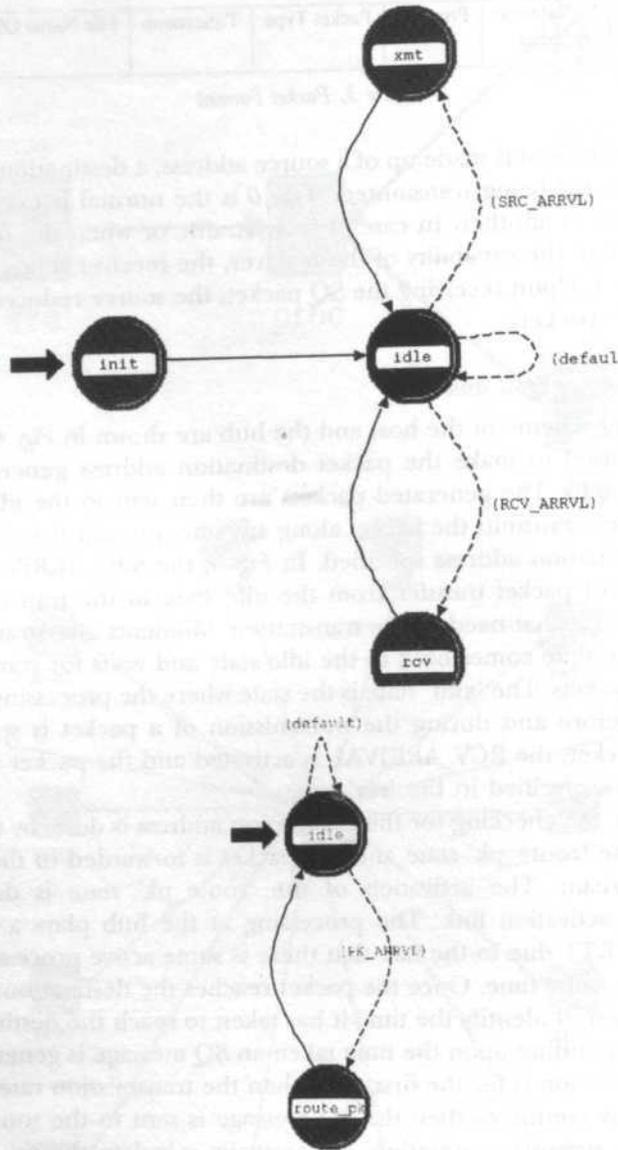


Fig. 4: Process module of the host (top diagram) and the hub (bottom diagram)

the packet will help to calculate the processing time (IP packet filtering time) and in turn can calculate the travel time alone. Based on RTT timing, the source needs to be informed if the destination identifies that the network is congested. After receiving the packet at the destination, the first task done is to calculate the time taken for the packet to reach the destination from the source. The source time is present in the packet header. This time is subtracted

from the current time at arrival to obtain the time taken. The SQ message is generated if the packets are generated at a faster rate by the source continuously. In this case, if the packets are received faster than the receiver can cope with, the receiver will send an SQ message to the source and so the source has to reduce the speed of transmitting packets.

Initial analysis was done for changing the time at which SQ messages are generated along with varying bits per second. Later, the processing was done at two levels, at the end stations and at the intermediate routers, and compared in terms of the recorded performances. In order to see that the situation of applying the active networks concepts of processing the data on the fly helps the congestion control algorithm, we have done the processing at the host level and also at the hub level. The processing time was analysed and the results are reported.

Procedure of Operation for Proposed Scheme

The pseudo code of the different operation procedures of the proposed congestion control algorithm is explained now.

INIT STATE:

Steps:

1. Set the packet distribution to be arranged for a uniform packet distribution.
2. Initialise the source quench, packet count and other variables.
3. Set the packet count value in the packet header.

SENDING NODE:

Steps:

1. Receive the packet from the source stream.
2. Set the destination address in the packet using a random function and then set the source address.
3. The packets are passed along the respective output streams in correlation with the destination.

RECEIVING NODE:

Steps:

1. Receive the packet at the receiver stream.
2. Get the current simulation time.
3. Check for packets at the other receiver streams also.
4. Increment the number of packets depending upon the number of packets that arrive.
5. Get the initial time at which the packet was sent out.
6. If initial time is -1, then the packet generated is a source quench message.
7. Get source and destination addresses and find the difference between the initialisation time and arrival time.
8. If the time difference is low for the first time, record a note about it. If this situation continues then the source quench message has to be generated.

9. Transmit the source quench message.
10. If the time difference is not high, the message could be positively received.

RESULTS

Analysis Scheme of the Model

The SQ message to the source controls the rate of data transmission onto the network and so directly impacts congestion control. The time taken for the packet to reach the destination from the source is calculated. If the source generates packets faster than the receiving capacity of the destination, the SQ message is sent to the source by the receiver. The source on receiving this SQ message reduces the rate of transfer of packets.

The model is simulated with various settings in the process model. The effect of data flow is analysed from one node to another with different round trip time. It has been found that the effect of changing the round trip time affects the whole model performance.

The following base models have been used in the simulation:

1. The error correction model: Point-to-Point error correction model (*dpt-ecc*).
2. The error allocation model: Point-to-Point error allocation model (*dpt-err*).
3. The propagation delay model: Whenever a point-to-point transmitter initiates a transmission, the computation of propagation delay is performed in order to determine the time required for the packet to reach the receiver in the destination node. A Point-to-Point propagation delay model is applied here.
4. The transmission delay model: This model is used to represent the simulated time interval that the transmitter requires to completely process and transmit the packet. A Point-to-Point transmission delay model is applied here.

The network was analysed for different data rates available in OPNET such as 4800bps, 9600bps and 19200bps and statistics obtained. These statistics were updated at intervals of simulated time determined by the *Update Intvl* (Update Interval) attribute set by the user in the simulation dialog box. The following statistics are presented in each update:

SQ Time: If the packet takes more than this time, then the source is informed to reduce its speed of packet transfer. The source will keep on increasing the speed of transmission at a gradual speed until it receives an SQ message.

SIM_T: The current simulation time in seconds.

REAL_T: The real running time of the simulation.

EV: The total number of events processed. An event usually signifies the arrival of a packet or an interrupt at a module, triggering some kind of response within the module.

AV_EV/S: The average number of simulation events per second of real time (measured since the beginning of the simulation).

IN_EV/S: The instantaneous number of simulation events per second

(measured using real time over the last update interval).

AV_EFF: The average efficiency of the simulation run. This is the simulation time divided by the real time.

The network under different data rates has been analysed with various values of *time_difference*, where *time_difference* = packet arrival time - packet generation time.

From the data presented in Table 1, 2 and 3, the followings were inferred:

1. The higher the data transmission rate, the higher the number of packets involved in congestion.
2. An increase in the data transmission rate results in a decrease in the time taken for the packet to reach the destination from the source.
3. The number of operations that ought to be performed on a packet keeps on increasing with the increase in the time duration.

TABLE 1
Packet information for 4800 bps

SQ Time	SIM_T	Real Time	EV	AV_EV/S	AV_EFF	Packets regenerated
0.7	1000	1.8	15160	8256	544.657	0
1.0	1000	1.8	15160	8198	540.83	0
1.5	1000	1.8	15160	8317	548.656	0
2.0	1000	1.8	15160	8288	546.707	0
2.00001	1000	1.9	15277	8051	527.057	9
2.001	1000	2.3	16402	7278	443.089	121
2.1	1000	2.2	16431	7411	451.089	124
2.25	1000	2.2	16487	7414	449.689	129
2.5	1000	2.3	16581	7315	441.217	141

TABLE 2
Packet information for 9600 bps

SQ Time	SIM_T	Real Time	EV	AV_EV/S	AV_EFF	Packets regenerated
0.0	1000	1.9	15175	8186	539.454	0
0.25	1000	1.8	15175	8516	561.24	0
0.5	1000	1.8	15175	8311	547.684	0
0.75	1000	1.9	15175	8111	534.528	0
1.0	1000	1.9	15175	8161	537.806	0
1.0005	1000	2.5	17225	6945	403.228	352
1.025	1000	2.4	17231	7046	408.916	355
1.05	1000	2.5	17238	6966	404.16	356
1.1	1000	2.5	17264	7031	407.313	360
1.2	1000	2.6	17322	6791	392.093	371
1.25	1000	2.5	17342	6921	399.131	373

TABLE 3
Packet information for 19200 bps

SQ Time	SIM_T	Real Time	EV	AV_EV/S	AV_EFF	Packets regenerated
0.25	1000	1.7	15178	8700	573.243	0
0.5	1000	1.8	15178	8381	552.188	0
0.7	1000	2.7	17844	6699	375.442	200
0.75	1000	2.6	17886	6782	379.195	204
1.0	1000	2.7	18153	6643	365.947	231
1.0005	1000	2.8	18153	6584	362.709	231
1.1	1000	2.7	18161	6648	366.096	233
1.2	1000	2.7	18178	6773	372.646	236

Therefore, to produce an SQ message, the time difference between the source and destination of packets was upset at different numbers of seconds for different data rate systems. The implication of the data from these tables is that when the SQ is issued and when the time taken for the data to reach the destination is less, there is a lesser likelihood of a collision. It is a tradeoff between placing the SQ time and the total time. The time impact on the network is due to two things: travel time and the processing time. The value of the SQ time is upset with regard to the set of operations or rules needed to make packet-filtering decisions. The number of rules varies according to the need of the network. If the number of computations that needs to be performed to take an action on the packet is high, then the processing time increases. This causes the time taken for the packet to reach the destination to increase. As the SQ time is based on the time required for the packet to move from source to destination, this increase in processing affects the SQ time also. In turn, causing the total time, for the packet to reach destination from source, to change based on the number of ACL rules as processing is done on-the-fly in the case of active networks. Thus with different access lists, the SQ time will also vary to cater for the changes in total time. Minimising the SQ time will safeguard the system from congestion but at the same time will affect the efficiency of network usage. On the other hand, fixing the SQ timer at a high value makes the network prone to congestion while resulting in high network utilisation. The use of source quench in an appropriate manner (in accordance with the network capability) reduces the probability of collisions and congestion.

The simulated system was tested by varying inter-arrival times (the time gap between the transmission of two consecutive packets) between packets under passive network criteria in order to identify the impact of varying packet inter-arrival time on the network performance. There was an increase in the End-to-End (ETE) delay when packets were transmitted with short inter-arrival time but network utilisation has increased because of faster packet transmission rate, as shown in *Fig. 5*.

The increase in the ETE delay indicates a possibility of congestion in the network. If the packet inter-arrival time is small, then there will be more packets

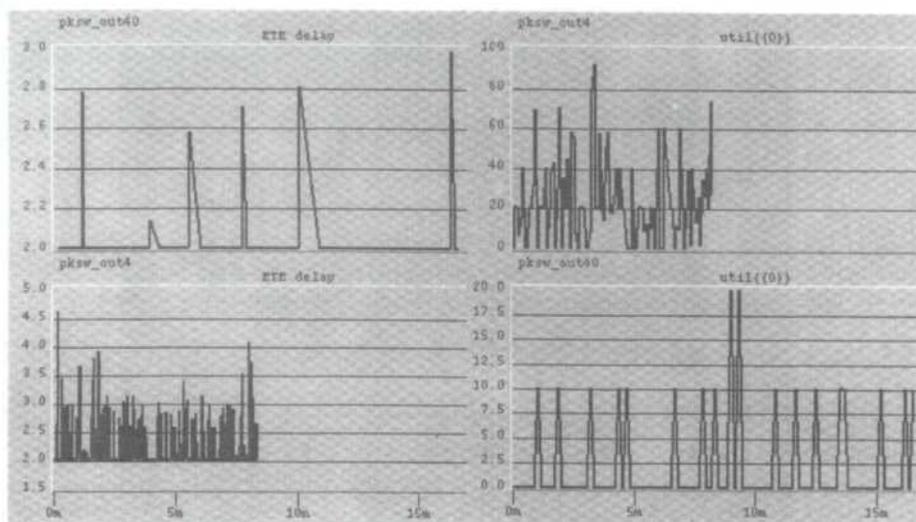


Fig. 5: ETE Delay & Utilisation with inter-arrival times as 4 and 40 sec. (top-right-graph: Inter-arrival time of 4 sec; bottom-right-graph: Inter-arrival time of 40 sec) (x-axis-left-graph: Simulation time; y-axis-left-graph: ETE delay time) (x-axis-right-graph: Simulation time; y-axis-right-graph: Utilisation)

in the network causing higher network utilisation and ETE delay density. The above graph helps us to infer that if the packets are sent with small inter-arrival time, there is a high possibility for congestion as well as high network utilisation.

Congestion is not due only to the transfer time from source to destination but also due to the IP packet filtering process at the hub. The process of identifying the situation of congestion in the network was done at the destination end as well as at the hub and the results were compared. The impact of the network congestion found at the end hosts has an equivalent impact on the hub too, as shown in Fig. 6. This makes it clear that if we take an action at the hub on the congestion control, it will reduce the impact of congestion throughout the network. The analysis was done when packets were generated at various packet inter-arrival times.

After identifying that the presence of some actions at the hub makes an impact on the network, we added the code to send the SQ message from the hub to source in case there occurs a very low ETE delay. Because of the presence of SQ action at the hub itself, we can see from Fig. 7 that the ETE delay towards the node end is reduced.

A highly congested network may end up in collision and loss of packets. To avoid this situation, we handled the congestion scenario at the router level itself. Comparing the top two graphs in Fig. 7, we can identify that the ETE delay is reduced and the density of ETE delay is also reduced. This is due to the fact that the intermediate router takes action and sometimes destroys the packet. Therefore, it is clear that the performance of the system improves with the inclusion of Source Quench at the router level.

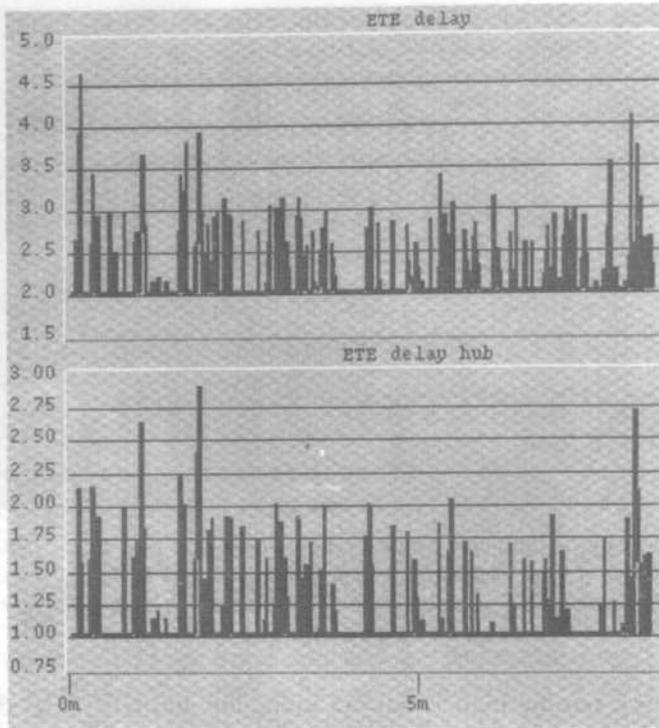


Fig. 6: ETE delay at Hub and Node (top-graph: ETE delay between source and destination; bottom-graph: ETE delay between source and hub) (x-axis: Simulation time; y-axis: ETE delay time)

DISCUSSION

Analysis of Various Congestion Control Schemes:

1. Rate Adaptation Protocol: Round trip delay (r) is measured once every round-trip-time. An increase in ' r ' is considered to be the situation of congestion and the window W is reduced to half. At the same time, if there is no increase in r , then the window size is increased by 1.

$$W \leftarrow W + \delta W$$

$$W \leftarrow cW; c = 0.5 \tag{1}$$

If c is small or W is large at its peak, then the scheme would loose the available bandwidth except when W is at its peak for few cycles. So, the utilisation of the bandwidth is not appropriate. Here, the net idle time is based on the bottleneck server and also on the idle due to slow increase of the window size. At i^{th} period, the idle time due to slow increase of window

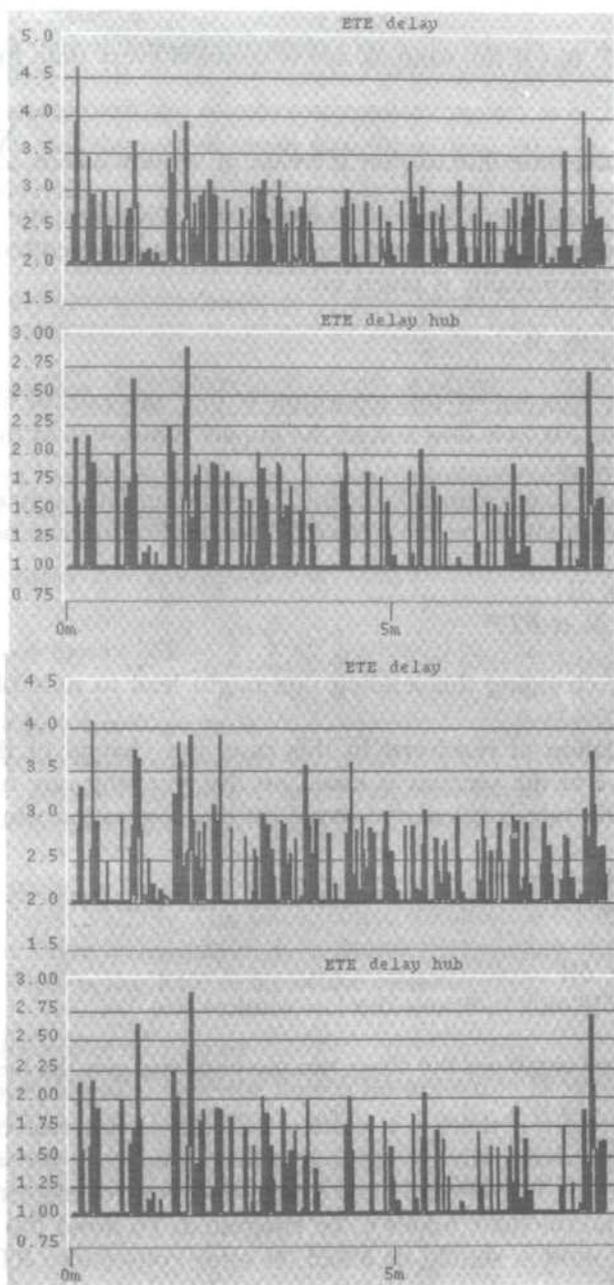


Fig. 7: ETE delay at Node and Hub (top-graph: without SQ at Hub; bottom-graph: with SQ at Hub)
 (x-axis: simulation time; y-axis: ETE delay time)
 (ETE delay: ETE delay between Source and Destination;
 ETE delay Hub: ETE delay between Source and Hub)

size is $\sum_{i=0}^{V-1} i$. But if the window size is doubled every time for an increase,

then the idle time due to slow increase of window size is $\sum_{i=0}^{V-1} (V-2^i)$.

2. Loss-Delay Adaptation Algorithm: As per this algorithm, the change of the window size (at time t) is made only after three independent analysis and so, the proportionality is based on:

$$W_t \propto [W_{t-1}, W_{t-2}, W_{t-3}] \quad (2)$$

The major concern of this algorithm is that the rate of change on the window size, $\partial W/\partial t$ is slow and so the impact is not seen at the proper time of the network.

3. TCP-Friendly Rate Control Protocol: The sending rate is changed based on the current situation on the network. This implication can be expressed as

$$\begin{aligned} \lambda &\leftarrow \lambda + \partial\lambda \\ \text{Where, } \partial\lambda &\propto RTT \end{aligned} \quad (3)$$

Frequently changing the sending rate might lead to unstable state of the network.

4. TCP Emulation at receivers: In this case, the change of the congestion window size at the receiver is based on the receiving rate of the receiver. The sending rate is calculated based on the congestion window size.

$$\begin{aligned} \partial W &\propto \partial\lambda_r \\ \partial\lambda_r &\propto \partial W \\ \text{So,} \\ \lambda &\leftarrow \lambda + \delta W \\ \text{Where } \delta W &\text{ only indicates the last window size} \end{aligned} \quad (4)$$

5. Our proposal combines the above two methods and gets the best out of them.

Special Scenario: If there is a sudden increase or decrease of the sending rate especially when the window size is at its maximum, then the rate adaptation algorithm and the TCP-Friendly rate congestion control algorithm cause the buffer at the intermediate nodes to be congested. To avoid this situation, the change on the window should be based on some consecutive RTT and not on individual RTT.

Along with the special scenario indicated, our proposed algorithm is not impacted by idle time and slow response like the Rapid adaptation and Loss-Delay algorithms. The possibility of unstable condition (as in TCP-Friendly rate algorithm) is avoided in our proposed system. Taking into consideration the receiving rate and the state of the network helps the proposed algorithm to be more stable.

Impact on ETE Delay

Using Fig. 5, we plan to analyse the impact on ETE delay by varying inter-arrival times, as 4 & 40 seconds. Here, the number of levels (*a*) is 2 (inter-arrival time as 4 & 40 seconds), and the number of replications (*r*) is 8. The differences between means and the overall mean give the impact on ETE delay.

TABLE 4
Impact on ETE delay

	Out4	Out40	
Mean	2.92375	2.215	2.569375
Impact on ETE delay	0.354375	-0.354375	

We can find out the experimental errors by finding the difference between the estimated response and the measured response using the equation:

$$\text{Experimental Error } e_{ij} = y_{ij} - \bar{y}_{ij} - \mu - \alpha_j \tag{5}$$

$$\text{Sum of Squared Errors SSE} = \sum_{i=1}^b \sum_{j=1}^a e_{ij}^2$$

- where y_{ij} = measured response
- \bar{y}_{ij} = estimated response
- μ = mean of all the measured response values
- α_j = mean of the each Column *j*

The percentage of variation explained by varying inter-arrival times (as 4 and 40 seconds) is 39.03%. The unexplained variation is 60.96%.

Looking at these percentages, it is found that varying inter-arrival times considerably impacts the ETE delay of the system.

Impact on Utilisation of the Network

Using Fig. 5, we plan to analyse the impact on utilisation by varying inter-arrival times as 4 & 40 seconds. Here, the number of levels (*a*) is 2 (inter-arrival time as 4 & 40 seconds), and the number of replications (*r*) is 6. The differences between means and the overall mean give the impact on utilisation.

TABLE 5
Impact on utilisation

	Out40	Out4	
Mean	6.8543333	55	30.9271667
Impact on utilisation	-24.072834	24.072833	

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained by varying inter-arrival times is 76.55%. The unexplained variation is 23.44%.

Looking at these percentages, it is found that varying inter-arrival times considerably impacts the utilisation of the network. Thus, we can visualise that the impact of the inter-arrival time on the utilisation is more than that towards ETE delay of the system.

Impact on ETE delay at Hub and Node

Using Fig. 6, we plan to analyse the impact on ETE delay at the hub and the node. Here, the number of levels (a) is 2 (ETE delay at hub, ETE delay at node), and the number of replications (r) is 16. The differences between means and the overall mean give the impact on ETE delay both at hub and node.

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained by the ETE delay at the node/hub is 62.025%. The unexplained variation is 37.97%.

Looking at these percentages, it is found that there is considerable impact on the ETE delay at the hub with respect to the ETE delay at the node.

TABLE 6
Impact on ETE Delay at hub and node

	Node	Hub	
Mean	2.97625	1.7425	2.359375
Impact on ETE delay	0.616875	-0.616875	

Impact on ETE Delay at Hub and Node without SQ

Using Fig. 7, we plan to analyse the impact of ETE delay at the node and hub without SQ enabled. Here, the number of levels (a) is 2 (ETE delay at node, ETE delay at hub), and the number of replications (r) is 16. The differences between means and the overall mean give the impact on ETE delay at hub and node without SQ.

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained by the ETE delay at the node and hub is 56.47%. The unexplained variation is 43.53%.

Looking at these percentages, it is found that the ETE delay at the node has significant impact on the ETE delay at the hub without SQ being enabled.

TABLE 7
Impact on ETE Delay at hub and node without SQ

	ETE Delay	HUB	
Mean	2.97625	1.846875	2.4115625
Impact on ETE delay	0.5646875	-0.564688	

Impact on ETE Delay at Hub and Node with SQ

Using Fig. 7, we plan to analyse the impact on ETE delay at the node and hub with SQ enabled. Here, the number of levels (a) is 2 (ETE delay at node, ETE delay at hub), and the number of replications (r) is 13. The differences between means and the overall mean give the impact on ETE delay at hub and node with SQ.

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained by the ETE Delay at node/hub is 54.28%. The unexplained variation is 45.72%.

Looking at these percentages, it is found that the ETE delay at the node has significant impact on the ETE delay at the hub with SQ being enabled. Thus, we can visualise that there is significant impact on the ETE delay at the hub by the ETE delay at the node irrespective of the SQ being enabled or not.

TABLE 8
Impact on ETE Delay at hub and node with SQ

	HUB	ETE Delay	
Mean	1.9284615	3.0684615	2.49846154
Impact on ETE delay	-0.57	0.57	

Impact on ETE Delay at Hub with and without SQ

Using Fig. 7, we plan to analyse the impact on ETE delay at the hub with and without SQ enabled. Here, the number of levels (a) is 2 (with SQ, without SQ), and the number of replications (r) is 17. The differences between means and the overall mean give the impact on ETE delay at hub with and without SQ.

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained with or without using SQ is 0.031%. The unexplained variation is 99.97%.

Looking at these percentages, it is found that the variation in having or not having SQ does not impact on the ETE delay at the hub.

TABLE 9
Impact on ETE Delay at hub with and without SQ

	With SQ	Without SQ	
Mean	1.7811765	1.7652941	1.77323529
impact on ETE delay	0.0079412	-0.007941	

Impact on ETE Delay at Node with and without SQ

Using Fig. 7, we plan to analyse the impact of ETE delay at the node with and without SQ enabled. Here, the number of levels (a) is 2 (with SQ, without SQ), and the number of replications (r) is 13. The differences between means and the overall mean give the impact on ETE delay at node with and without SQ.

Using Equation 5, we can find out the experimental errors by finding the difference between the estimated response and the measured response.

The percentage of variation explained by using or non-using of SQ algorithm is 0.3076%. The unexplained variation is 99.69%.

Looking at these percentages, it is found that the variation in having or not having SQ does not impact on the ETE delay at the node.

TABLE 10
Impact on ETE Delay at node with and without SQ

	With SQ	Without SQ	
Mean	3.0353846	2.9669231	3.00115385
Impact on ETE delay	0.0342308	-0.034231	

CONCLUSION

Efficiency of a network can be improved by providing improved quality of service. Thus, enhancement to the quality of service of Transmission Control Protocol (TCP) is required to improve network performance. Enhancement of QoS using Source Quench messages to control the speed at which the source sends data to the receiver was implemented. The retransmission time for the network was set by first analysing the time taken by a packet to reach the destination from the source, along with the processing time.

From the discussion, we have indicated that the processing of the packets to check for congestion was done on intermediate nodes, which is using the processing on-the-fly concept of active networks. With the active networks oriented congestion control, our discussion states that there is significant impact on the ETE delay at the hub by the ETE delay at the node irrespective of the SQ being enabled or not. Being a simulated system, the communication time is not as appropriate as the case of a real system and also the processing is done on one processor. The possibility of congestions is identified earlier at the routers, and the necessary action is taken.

REFERENCES

- GREGORY, G. FINN. 1995. A connectionless congestion control algorithm. *Proceedings of ACM SIGCOMM 19(5)*: 12-30.
- FLOYD, S., J. PADHYE and J. WIDMER. 2000. Equation-based congestion control for unicast applications. *Proc. ACM SIGCOMM*, p. 43-56. Stockholm, Sweden.
- LAKSHMAN, T.V., and U. MADHOW. 1997. The performance of TCP/IP for networks with high Bandwidth-delay products and random loss. *IEEE/ACM Transactions on Networking 5(3)*: 336-350.
- PARTRIDGE, C. and T. J. SHEPARD. 1997. TCP/IP performance over satellite links. *IEEE Network 11(5)*: 44-49.
- REJAE, R., M. HANDLEY and D. ESTRIN. 1999. Rap: An end-to-end rate-based congestion control mechanism for real-time streams in the internet. *Proceedings of IEEE INFOCOM* p. 1337-1345.
- RHEE, I., V. OZDEMIR and Y. Yi. 2000. TEAR: TCP emulation at receivers – flow control for multimedia streaming. *Tech. Report*, Department of Computer Science, NCSU.
- SISALEM, D. and A. WOLISZ. 2000. LDA+ TCP-friendly adaptation: A measurement and comparison study. *IEEE International Conference on Multimedia and Expo (III)*, p. 1619-1622.
- TANENBAUM, A. S. 1996. *Computer Networks*. New Jersey: Prentice-Hall International.
- WANG, Z. and J. CROWCOFT. 1992. Eliminating periodic packet losses in the 4.3-Tahoe BSD TCP congestion control algorithm. *Computer Communication Review 22(2)*: 9-13.
- WIDMER, J., R. DENDA and M. MNAUVE. 2001. A survey on TCP-friendly congestion control. *IEEE Network*, May/June: 28-37.

Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer Overlaid by Dense Sand Layer

Abdulhafiz O. Al-Shenawy & Awad A. Al-Karni

King Saud University, Civil Eng. Dept., P. O. Box 800

Riyadh, 11421, Saudi Arabia

Received: 8 May 2003

ABSTRAK

Pengiraan keupayaan galas muktamad tapak cetek ke atas sistem tanah dua lapisan bergantung pada corak permukaan kegagalan yang dibangunkan di bawah tapak. Bagi lapisan tanah liat lemah yang dilapisi oleh lapisan pasir tumpat atas, kajian terdahulu menganggap bahawa permukaan kegagalan adalah kegagalan ricih menebuk melalui lapisan pasir atas dan mod kegagalan Prandtl dalam lapisan tanah liat lemah bawah. Dengan mengadaptasi anggapan ini dalam kajian ini, persamaan keupayaan galas muktamad diperoleh sebagai satu fungsi sifat-sifat tanah, kelebaran tapak, dan ketebalan tanah atas. Kertas ini membentangkan kajian berparameter yang terperinci bagi parameter reka bentuk termasuk kesan sudut geseran, nisbah ketebalan lapisan pasir dengan kelebaran tapak, nisbah kedalaman pembenaman dengan kelebaran tapak, dan nisbah kejeleketan tanah liat dengan keluaran berat unit tanah liat oleh kelebaran tapak. Carta reka bentuk dibangunkan dalam bentuk tiada dimensi untuk julat parameter reka bentuk yang sangat luas. Carta yang baru memberi pilihan lain kepada sesiapa yang percaya bahawa carta reka bentuk dibangunkan berdasarkan analisis had atas lebih anggar keupayaan galas disebabkan sifat penyelesaian had atas. Carta reka bentuk baru adalah terhad kepada tapak cetek.

ABSTRACT

Calculation of the ultimate bearing capacity of shallow footing on a two layered system of soil depends on the pattern of the failure surface that develops below the footing. For a weak clay layer overlaid by a top dense sand layer, previous studies assumed that the failure surface is a punching shear failure through the upper sand layer and Prandtl's failure mode in the bottom weak clay layer. By adapting this assumption in this study, the ultimate bearing capacity equation was derived as a function of the properties of soils, the footing width, and the topsoil thickness. The paper presents a detailed parametric study of the design parameters including the effect of angle of friction, the ratio of the thickness of sand layer to the footing width, the ratio of the depth of embedment to the footing width, and the ratio of the clay soil cohesion to the product of the clay unit weight by the footing width. Design charts were developed in dimensionless form for very wide ranges of design parameters. The available method based on the limit equilibrium analysis was developed in dimensionless form and for a limited range of design parameters. The new charts give another option for those who believe that the design charts developed based on the upper limit analysis overestimate the bearing capacity due to the very nature of the upper bound solution. The new design charts are limited to shallow footings.

Keywords: Shallow footing, bearing capacity, two layered system, weak clay layer, dense sand layer, design chart

INTRODUCTION

The function of a foundation is to transfer the load of the superstructure to the underlying soil formation without overstressing the soil. The soil must be capable of carrying the load for structure(s) placed upon it without shear failure and with the resulting settlement being tolerable for that structure.

Many investigations on the subject of ultimate bearing capacity have been carried out during the past century. Subsequently, numerous proposals have been advanced regarding considerations, criteria, and procedures for evaluation of the ultimate bearing capacity of soils. Among the very early contributors were Prandtl (1921) who developed a solution for a surface strip footing over a perfectly plastic cohesive-frictional weightless half-space. Reissner (1924) extended the solution of Prandtl to include the effect of a uniform surcharge load on the resistance of penetration of ultimate applied load. Since real soils possess weight, Terzaghi (1943) was the first to introduce the concept of ultimate bearing capacity and presented a comprehensive theory for the evaluation of such capacity of shallow foundations. Subsequently, the bearing capacity theory went through many modifications to account for different features such as foundation shape, load inclination, ground slope, nonsymmetrical loads, and water table. The general bearing capacity theories proposed by Meyerhof (1963), Hansen (1970), Vesic (1973) and others are now routinely used in foundation design.

The bearing capacity theories mentioned above involve cases in which the soil supporting the foundation is homogeneous and extends to a considerable depth. However, in practice, layered soil profiles are often encountered. For layered clayey soil, Button (1953) was the first to analyse footings on layered soils of different cohesion. Many other studies were conducted for clayey layers including those of Sivareddy and Srinivasan (1967), Brown and Meyerhof (1969), Desai and Reese (1970a, b) and Merifield *et al.* (1999). In another case, many authors studied the bearing capacity of a sand layer overlaying a clay layer. These studies were conducted by Meyerhof (1974), Meyerhof and Hanna (1978), Hanna and Meyerhof (1980), Hardy and Townsend (1982), Okamura *et al.* (1997), Kenny and Andrawes (1996), Burd and Frydman (1997), and Michalowski and Shi (1995). For footings resting over a two-layer $c-\phi$ soil, the ultimate bearing capacity was studied by Purushothamaraj *et al.* (1974), Satyanarayana and Garg (1980), Florkiewicz (1989), and Azam and Wang (1991).

In this study, design charts were developed using the punching shear model in a dimensionless form since those of Hanna and Meyerhof (1980) were not presented in a nondimensionalised form, which limits their application. The new design charts were developed for very wide ranges of design parameters. The presented charts here may be useful in overcoming the problem of the design charts that were developed by Michalowski and Shi (1995) which may overestimate the bearing capacity by a significant amount because of the very nature of the upper bound solution on which the derivation is based.

FORMULATION

According to Fig. 1, a shallow footing with width B is resting on a two-layered soil. The top layer is a dry granular soil with thickness H , unit weight γ , and angle of friction ϕ . The bottom layer is assumed to be a normally consolidated clay with undrained cohesion c_u . The punching shear failure mechanism is assumed to be developed here which is characterised by the formation of two vertical shear bands inside the granular layer and a Prandtl-type failure in the cohesive layer as shown by Fig. 1. By assuming a constant vertical stress (σ_z) acting along the width (B) of horizontal slice with thickness dz at depth z from the base of the footing (see Fig. 2), and considering equilibrium of the vertical forces then

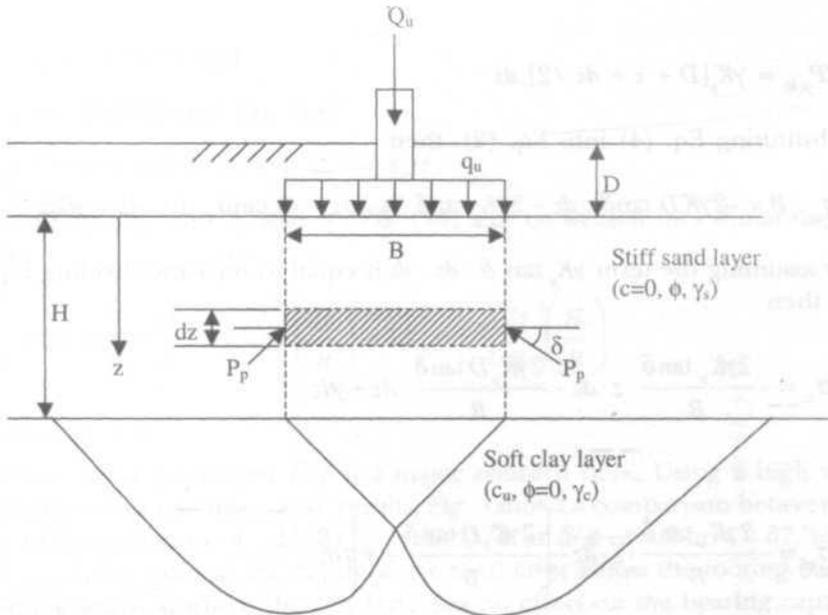


Fig. 1: Punching shear model of strip footing over a two-layered soil

$$\sigma_z \cdot B - (\sigma_z + d\sigma_z) \cdot B - 2dP \cdot \sin\delta + \gamma B dz = 0 \quad (1)$$

which gives

$$-d\sigma_z \cdot B - 2dP_p \sin \delta + \gamma B dz = 0 \quad (2)$$

where dP_p is the passive force which inclines upwards at an angle δ to the horizontal and equal to $dP_{p(h)}/\cos\delta$ where

$$dP_{p(h)} = p \cdot dz = \left[\gamma DK_p + \gamma \left(z + \frac{dz}{2} \right) K_p \right] \cdot dz \quad (3)$$

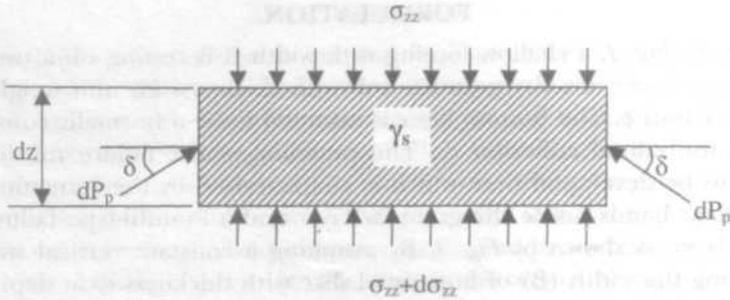


Fig. 2: Applied forces on a strip dz of the failure zone at depth z from the base of footing

or

$$dP_{p(h)} = \gamma K_p [D + z + dz/2] \cdot dz \tag{4}$$

By substituting Eq. (4) into Eq. (2), then

$$d\sigma_z \cdot B = -2\gamma K_p D \tan \delta \cdot dz - 2\gamma K_p \tan \delta \cdot z \cdot dz - \gamma K_p \tan \delta \cdot dz \cdot dz + \gamma B dz \tag{5}$$

By assuming the term $\gamma K_p \tan \delta \cdot dz \cdot dz$ is equal to zero and dividing Eq. (5) by B , then

$$d\sigma_z = -\frac{2\gamma K_p \tan \delta}{B} \cdot z \cdot dz - \frac{2\gamma K_p D \tan \delta}{B} \cdot dz + \gamma dz \tag{6}$$

or

$$d\sigma_z = -\frac{2\gamma K_p \tan \delta}{B} \cdot z \cdot dz - \left[\frac{2\gamma K_p D \tan \delta}{B} - \gamma \right] \cdot dz \tag{7}$$

By integrating Eq.(7)

$$\sigma_z = -\frac{\gamma K_p \tan \delta}{B} \cdot z^2 - \left[\frac{2\gamma K_p D \tan \delta}{B} - \gamma \right] z + C \tag{8}$$

where C is the integration constant, and Eq. (8) can be rewritten as

$$\sigma_z = -A_1 \cdot z^2 - A_2 z + C \tag{9}$$

where

$$A_1 = \frac{\gamma K_p \tan \delta}{B} \tag{10}$$

and

$$A_2 = \left[\frac{2\gamma K_p D \tan \delta}{B} - \gamma \right] \quad (11)$$

By applying the boundary conditions, at $z=0$; $\sigma_{zz}=q_u$ and from Eq. (9), then

$$C = q_u \quad (12)$$

and at $z=H$, $\sigma_{zz} = q_{up}$ where $q_{up} = 5.14c_u + \gamma(D + H)$ and from Eq. (9), then

$$q_{up} = -A_1 H^2 - A_2 H + C \quad (13)$$

or

$$C = q_{up} + A_1 H^2 + A_2 H \quad (14)$$

Then from Eq.(12) and Eq. (14)

$$q_u = 5.14c_u + \gamma(D + H) + A_1 H^2 + A_2 H \quad (15)$$

By substituting with A_1 and A_2 , Eq. (15) can be written in a dimensionless form as:

$$\frac{q_u}{\gamma B} = 5.14 \frac{c_u}{\gamma B} + \frac{D}{B} + K_p \tan \delta \left(\frac{H}{B} \right)^2 + 2K_p \tan \delta \left(\frac{D}{B} \right) \left(\frac{H}{B} \right) \quad (16)$$

Evaluation of δ/ϕ

The effect of the parameter δ/ϕ is a major concern here. Using a high value of δ/ϕ will lead to unconservative results. Fig. 3 shows a comparison between the results of the variation of $(H/B)_{critical}$ with D/B at δ/ϕ of 1.0 and 0.67, where $(H/B)_{critical}$ is the ratio of the depth of the sand layer below the footing base to the footing width at which the clay layer has no effect on the bearing capacity. A significant effect is shown. For example, the $(H/B)_{critical}$ at $D/B=0$ is increased from 3.5 to 4.77 when δ/ϕ is decreased from 1.0 to 0.67, respectively. The difference between the two values increases as D/B increases. The figure also shows a comparison with the results of Michalowski and Shi (1995). The results of Michalowski and Shi (1995) show that the value of δ/ϕ is close to one when H/B is small and reduces as H/B increases. This conclusion is in agreement with suggestion of Meyerhof (1974). However, the experimental results of Hanna and Meyerhof (1980) suggested lower values of δ/ϕ . By using the values of δ/ϕ suggested by Hanna and Meyerhof (1980) the following relationships were developed

$$\delta/\phi = a + b \frac{q_2}{q_1} + c \left(\frac{q_2}{q_1} \right)^2 \quad (17)$$

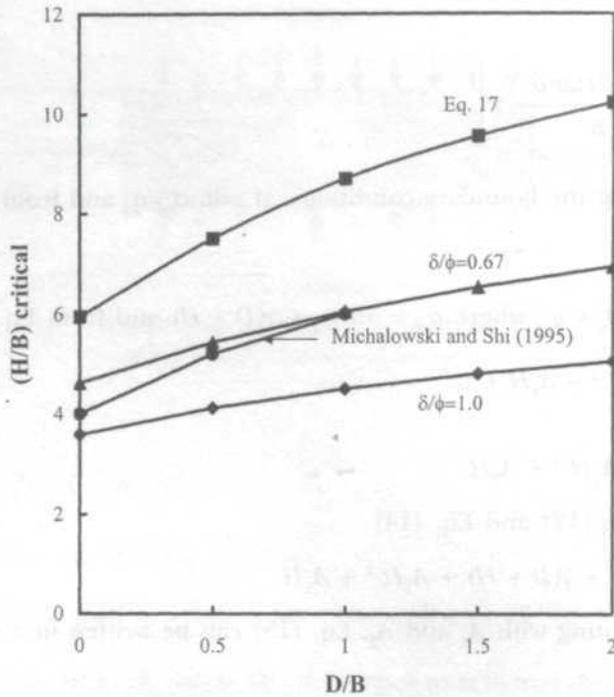


Fig. 3: Effect of δ/ϕ on the variation of $(H/B)_{critical}$ with D/B

where

$$a = 0.00829 \left(\frac{q_2}{q_1} \right) - 0.00872 \tag{18}$$

$$b = 0.000744 \left(\frac{q_2}{q_1} \right) + 1.0621 \tag{19}$$

$$c = -0.009 \left(\frac{q_2}{q_1} \right) - 0.0515 \tag{20}$$

and q_1 is the upper layer (sand layer) bearing capacity and q_2 is the lower layer (clay layer) bearing capacity. By using the relationship in Eq. (17), the variation of $(H/B)_{critical}$ with D/B is shown in Fig. 3. Since the value of $(H/B)_{critical}$ for Michalowski and Shi (1995) is based on the upper bound analysis, it overestimates the bearing capacity as discussed before in the introduction. However, the value of $(H/B)_{critical}$ based on Eq. 17 gives even more conservative results for the bearing capacity.

The results in Figs. 4 a and b show the variation of $(H/B)_{critical}$ with $c_u/\gamma B$. Figs. 4 a and b compare the results of Michalowski and Shi (1995) with the

results of Eq. (16) at δ/ϕ equal to one and δ/ϕ from Eq. (17), respectively. In Fig. 4a the results are close in values but differ in trend since those of Michalowski and Shi (1995) are concave up, while the curves of δ/ϕ equal to one are concave down. However, in Fig. 4b the results are different in values but agree in trend as the results of Michalowski and Shi (1995) are concave up, and the ones for δ/ϕ from Eq. (17) are concave up also. It may be concluded that using Eq. (17) to calculate δ/ϕ and using it in Eq. (16) is more reliable since the difference in values shown in Fig. 4b is due to the overestimation associated with the upper bound solution. Thus, Eq. 17 is used in solving Eq. 16 and in developing the design charts.

Parametric Study

The bearing capacity of the shallow foundation resting on layered soils, with an upper dense sand layer and bottom soft clay layer, depends mainly on five parameters. These parameters include ϕ , H/B , D/B , and δ/ϕ , which are related to the sand layer and the parameter $c_u/\gamma B$ which is related to the clay layer. These parameters are presented in dimensionless form to generalise their effect

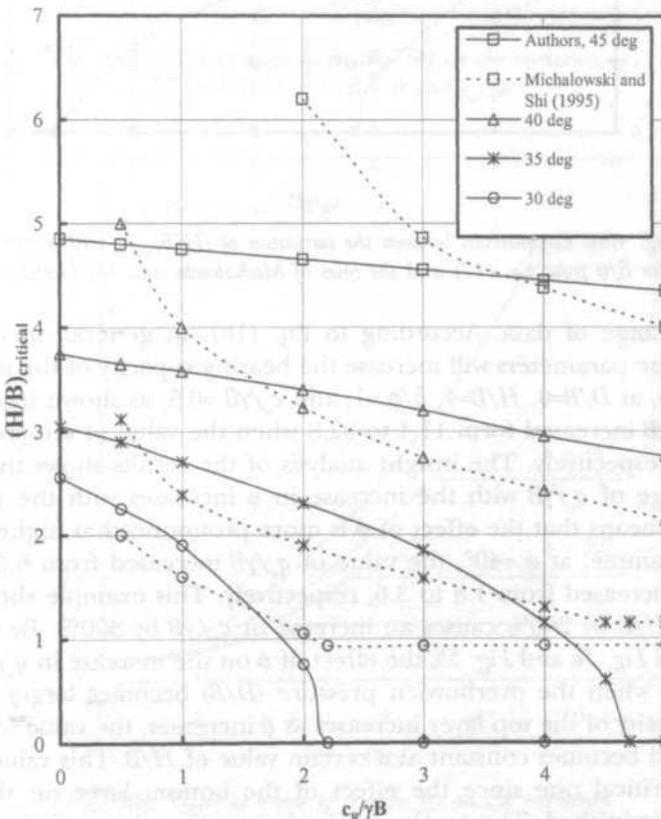


Fig. 4(a): Comparison between the variation of $(H/B)_{critical}$ with $c_u/\gamma B$ for $\delta/\phi = 1.0$ and the ones of Michalowski and Shi (1995)

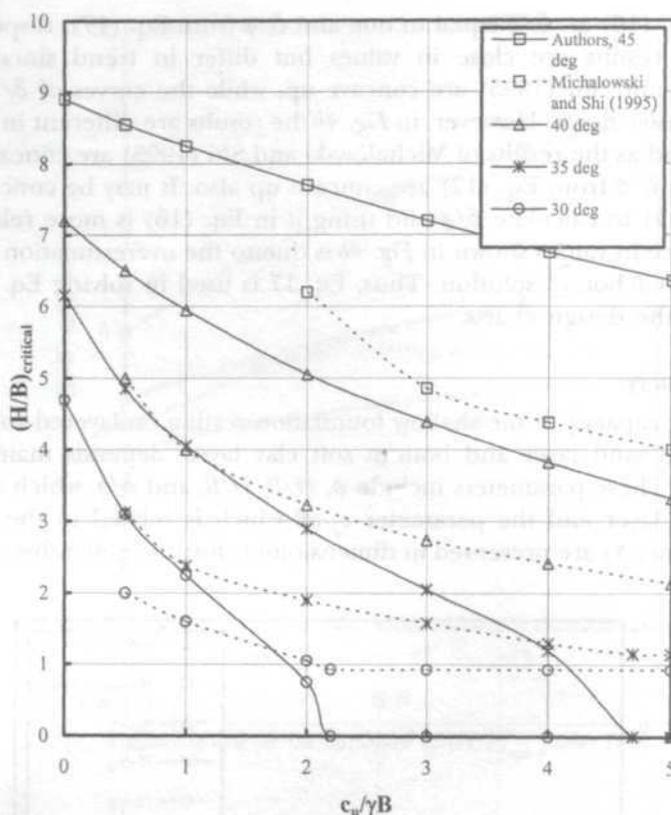


Fig. 4(b): Comparison between the variation of $(H/B)_{critical}$ with $c_u/\gamma B$ for δ/ϕ from Eq. (17) and the ones of Michalowski and Shi (1995)

for a wide range of data. According to Eq. (16), in general, increasing the values of these parameters will increase the bearing capacity of the layered soil. For example, at $D/B=0$, $H/B=4$, $\delta/\phi=1$, and $c_u/\gamma B=0.5$, as shown in Fig. 5a, the value of $q_u/\gamma B$ increased from 11.4 to 22.8 when the value of ϕ increased from 30° to 40° , respectively. The insight analysis of the results shows that the rate of the change of $q_u/\gamma B$ with the increase in ϕ increases with the increase in H/B , which means that the effect of ϕ is more pronounced at higher values of H/B . For example, at $\phi=40^\circ$, the value of $q_u/\gamma B$ increased from 6.54 to 18.46 when H/B increased from 1.8 to 3.6, respectively. This example shows that an increase in H/B by 200% causes an increase in $q_u/\gamma B$ by 300%. By comparing the results in Fig. 5a and Fig. 5b, the effect of ϕ on the increase in $q_u/\gamma B$ is more pronounced when the overburden pressure (D/B) becomes larger. Since the bearing capacity of the top layer increases as ϕ increases, the value of $q_u/\gamma B$ also increases and becomes constant at a certain value of H/B . This value of H/B is called the critical one since the effect of the bottom layer on the bearing capacity is diminished. The results in Fig. 5 show that the critical value of H/B increases as ϕ increases. From Fig. 6, the effect of the angle of friction (ϕ) on

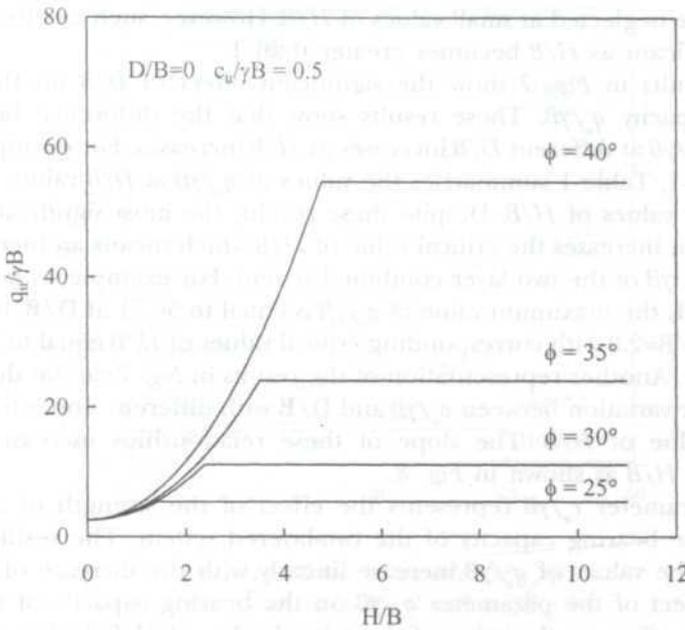


Fig. 5(a): Effect of angle of friction (ϕ) on the variation of $q_u/\gamma B$ with H/B at $D/B = 0$ and $c_u/\gamma B = 1$

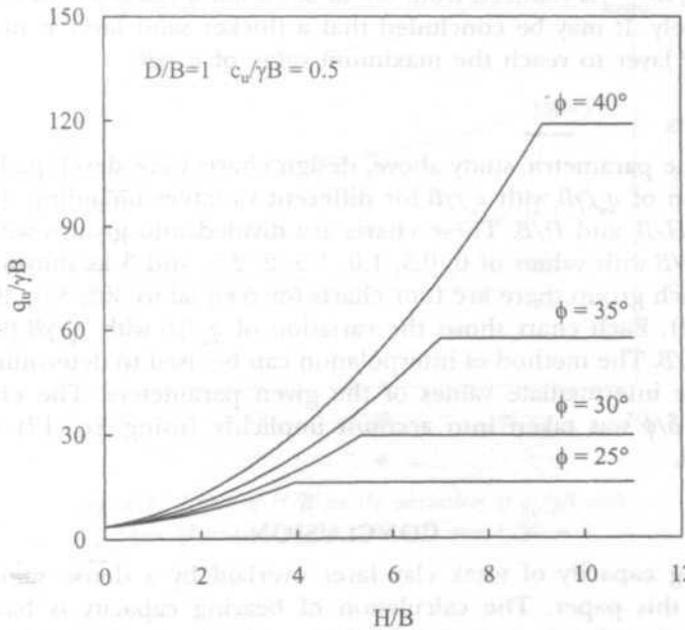


Fig. 5(b): Effect of angle of friction (ϕ) on the variation of $q_u/\gamma B$ with H/B at $D/B = 1$ and $c_u/\gamma B = 1$

$q_u/\gamma B$ can be neglected at small values of H/B . However, such an effect becomes more significant as H/B becomes greater than 1.

The results in Fig. 7 show the significant effect of D/B on the ultimate bearing capacity $q_u/\gamma B$. These results show that the difference between the values of $q_u/\gamma B$ at different D/B increases as H/B increases. For example, at $\phi=40$ and $q_u/\gamma B=1$, Table 1 summarises the values of $q_u/\gamma B$ at D/B values of 0 and 2 at different values of H/B . Despite these results, the most significant effect of D/B is that it increases the critical value of H/B which means an increase in the values of $q_u/\gamma B$ of the two layer combined system. For example, at $\phi=40$ and $q_u/\gamma B=1$, $\delta/\phi=1$, the maximum value of $q_u/\gamma B$ is equal to 54.71 at $D/B=0$ and equal to 182 at $D/B=2.0$ with corresponding critical values of H/B equal to 6 and 10.6, respectively. Another representation of the results in Fig. 7 can be shown in Fig. 8 as a linear variation between $q_u/\gamma B$ and D/B with different slopes for each line of each value of H/B . The slope of these relationships increases with the increase of H/B as shown in Fig. 8.

The parameter $c_u/\gamma B$ represents the effect of the strength of the bottom layer on the bearing capacity of the two-layered system. The results in Fig. 9 show that the values of $q_u/\gamma B$ increase linearly with the increase of $c_u/\gamma B$. The greatest effect of the parameter $q_u/\gamma B$ on the bearing capacity of the layered system is its effect on the value of the critical value of H/B . Unlike the effect of the angle of friction, the critical value of H/B is reduced as the value of $q_u/\gamma B$ increases as shown in Fig. 10. For example at $\phi=40^\circ$, $D/B=0$, and $\delta/\phi=1$, the value of $(H/B)_{\text{critical}}$ is reduced from 6.0 to 3.8 when $q_u/\gamma B$ is increased from 1 to 4, respectively. It may be concluded that a thicker sand layer is needed for a weaker clay layer to reach the maximum value of $q_u/\gamma B$.

Design Charts

Based on the parametric study above, design charts were developed to present the variation of $q_u/\gamma B$ with $c_u/\gamma B$ for different variables including the angle of friction ϕ , H/B , and D/B . These charts are divided into groups with different values of D/B with values of 0, 0.5, 1.0, 1.5, 2, 2.5, and 3 as shown in Figs. 11 to 17. In each group there are four charts for ϕ equal to 30° , 35° , 40° , and 45° (see Fig. 11). Each chart shows the variation of $q_u/\gamma B$ with $c_u/\gamma B$ for different values of H/B . The method of interpolation can be used to determine the value of $q_u/\gamma B$ for intermediate values of the given parameters. The effect of the parameter δ/ϕ was taken into account implicitly (using Eq. 17) during the calculations.

CONCLUSION

The bearing capacity of weak clay layer overlaid by a dense sand layer was studied in this paper. The calculation of bearing capacity is based on the assumption that the pattern of the failure surface is a punching shear failure through the sand layer and Prandtl's failure mode in the weak clay layer. By adapting this assumption in this study, a bearing capacity equation was derived

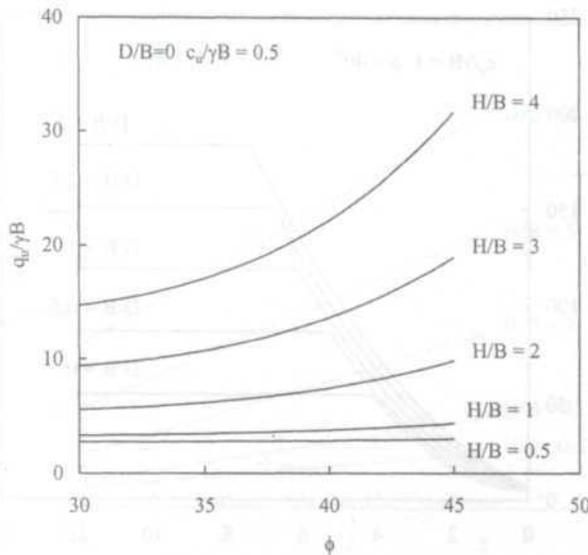


Fig. 6(a): Effect of H/B on the variation of $q_u/\gamma B$ with angle of friction (ϕ) at $D/B = 0$ and $c_u/\gamma B = 0.5$

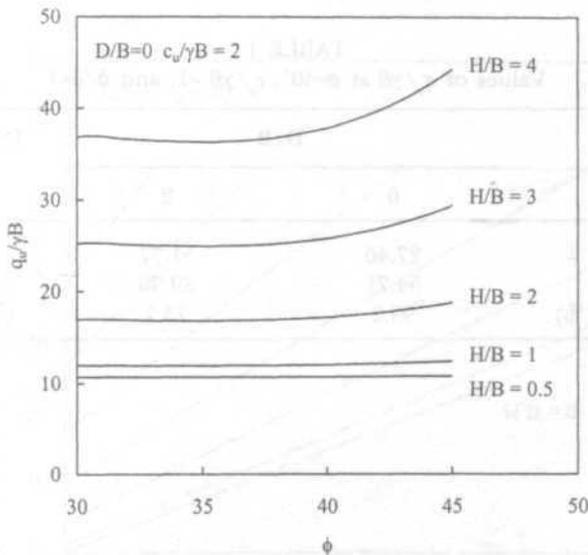


Fig. 6(b): Effect of H/B on the variation of $q_u/\gamma B$ with angle of friction (ϕ) at $D/B = 0$ and $c_u/\gamma B = 2$

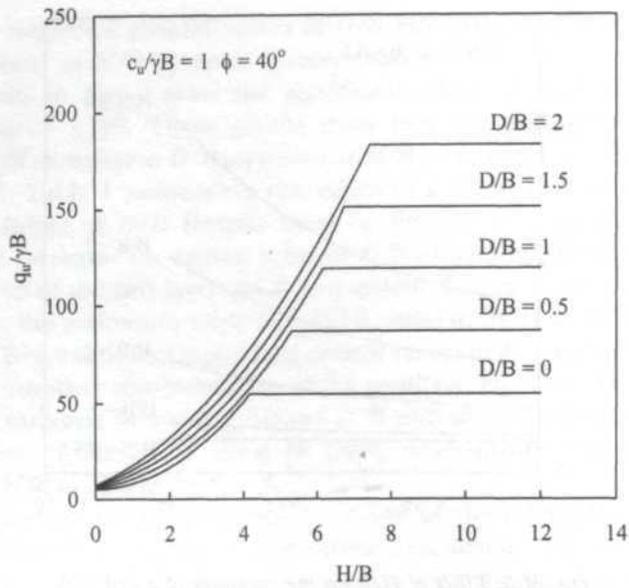


Fig. 7: Effect of D/B on the variation of $q_u/\gamma B$ with H/B at $\phi = 40^\circ$ and $c_u/\gamma B = 1$

TABLE 1
Values of $q_u/\gamma B$ at $\phi=40^\circ$, $c_u/\gamma B=1$, and $\delta/\phi=1$

H/B	D/B		Difference (%)
	0	2	
4	27.46	51.77	88.5
6	54.71	89.96	64.4
Difference (%)	99.2	73.7	

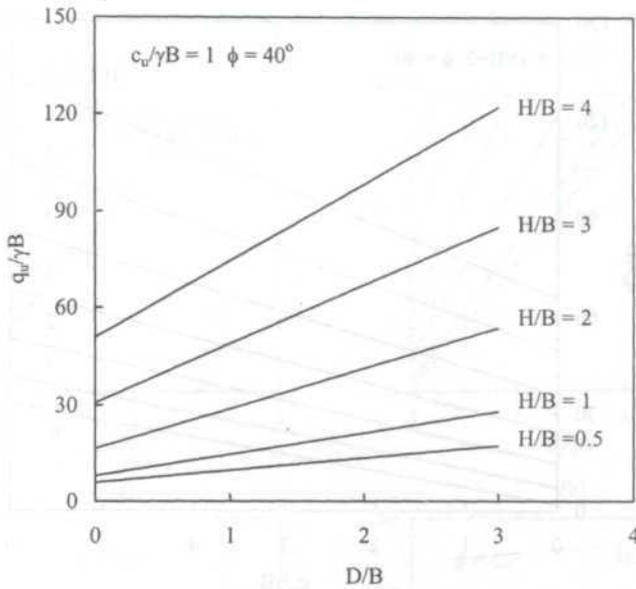


Fig. 8: Variation of q_u/gB with D/B at different values of H/B at $\phi = 40^\circ$ and $c_w/\gamma B = 1$

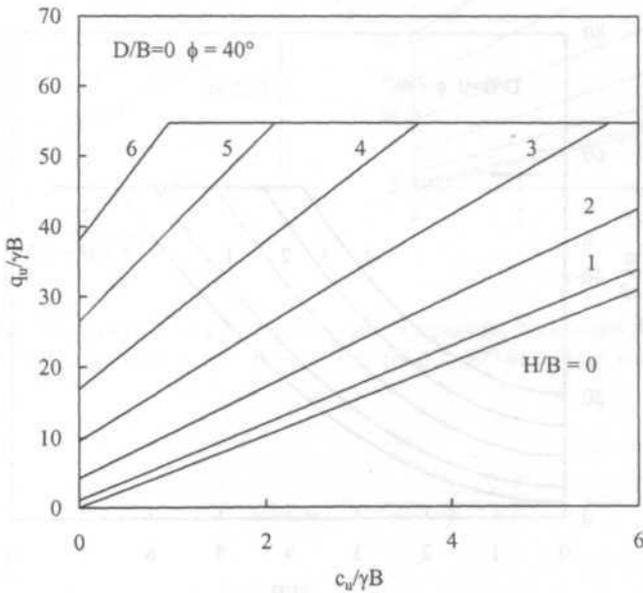


Fig. 9(a): Variation of $q_u/\gamma B$ with $c_w/\gamma B$ at different values of H/B at $\phi = 40^\circ$ and $D/B = 0$

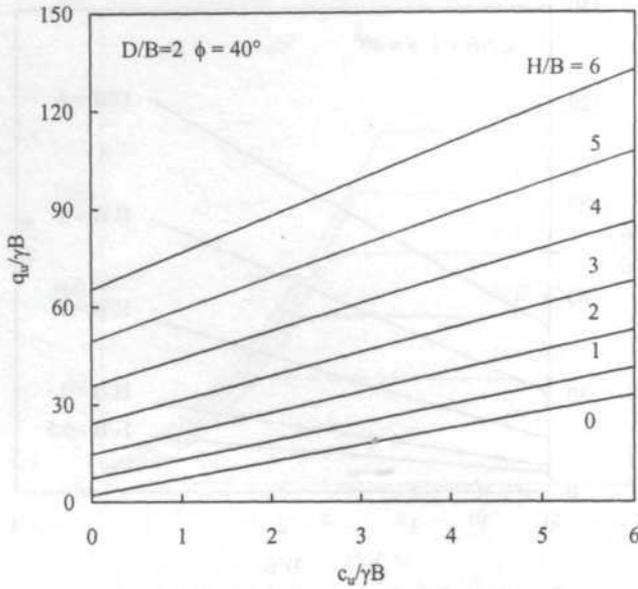


Fig. 9(b): Variation of $q_u/\gamma B$ with $c_u/\gamma B$ at different values of H/B at $\phi = 40^\circ$ and $D/B = 2$

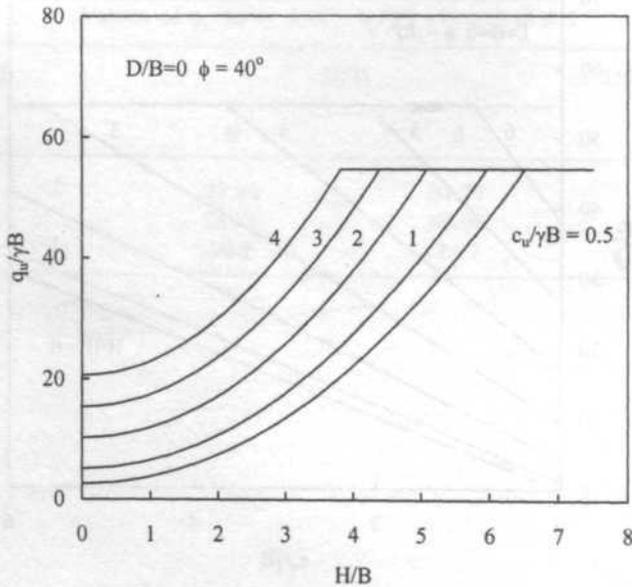


Fig. 10: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ at different values of H/B at $\phi = 40^\circ$ and $D/B = 2$

Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer

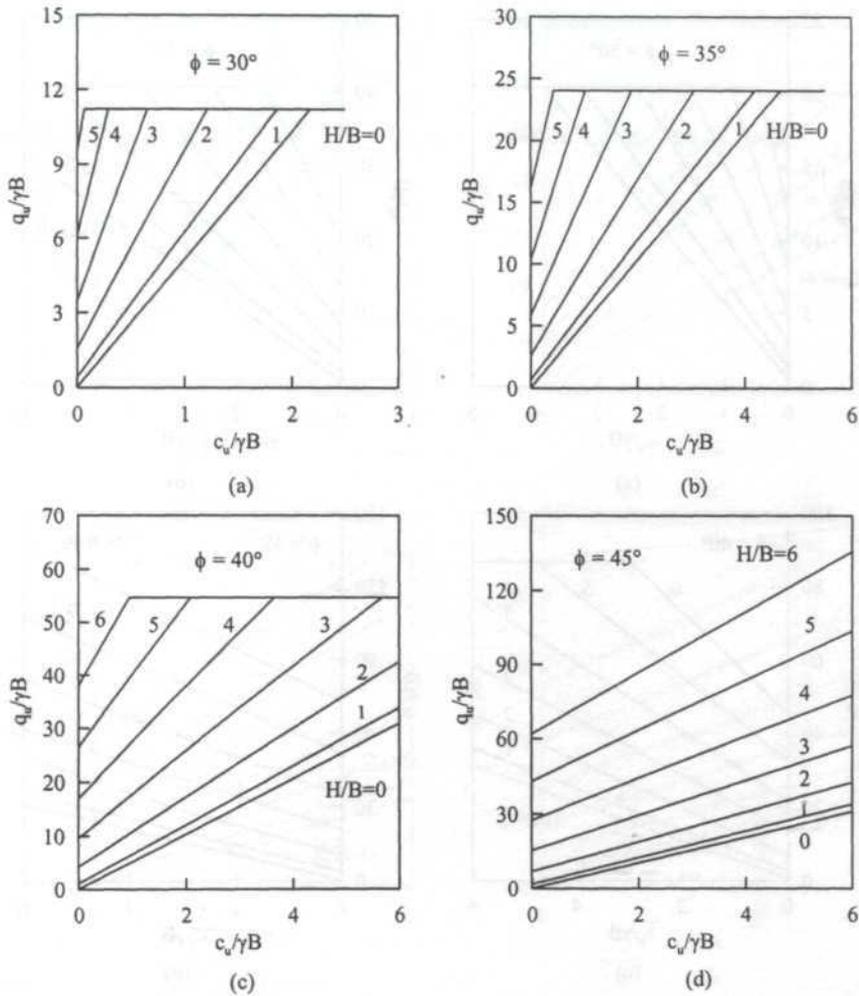


Fig. 11: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 0$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

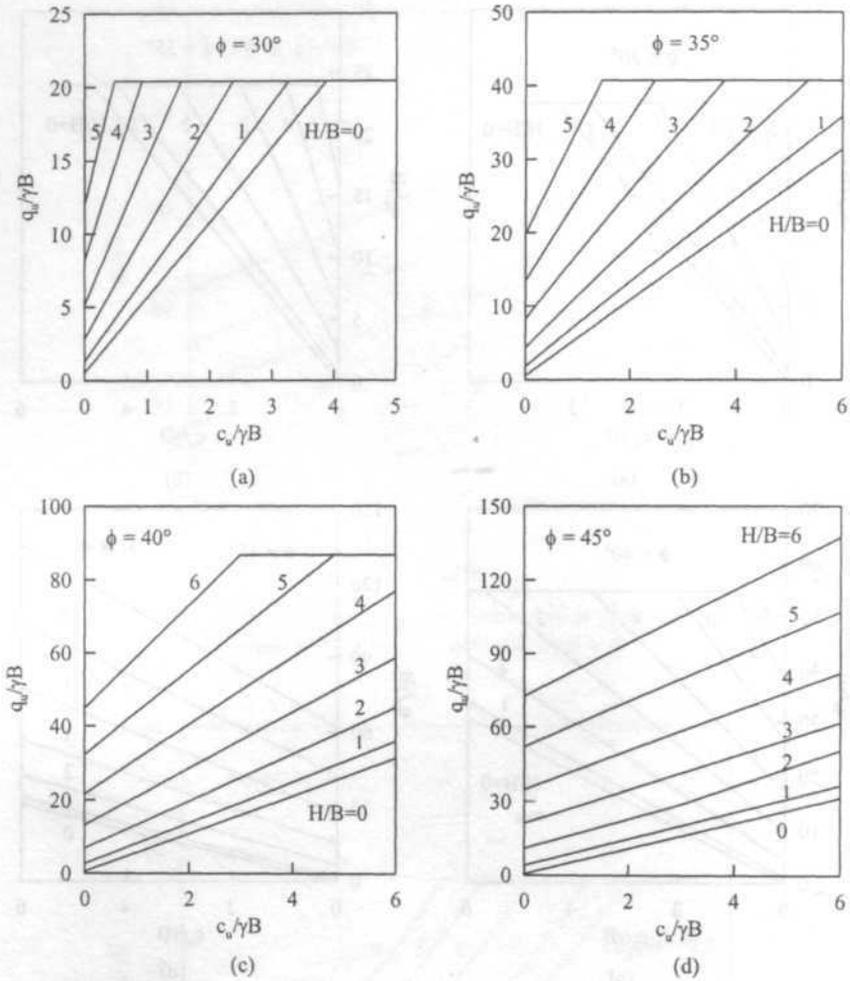


Fig. 12: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 0.5$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer

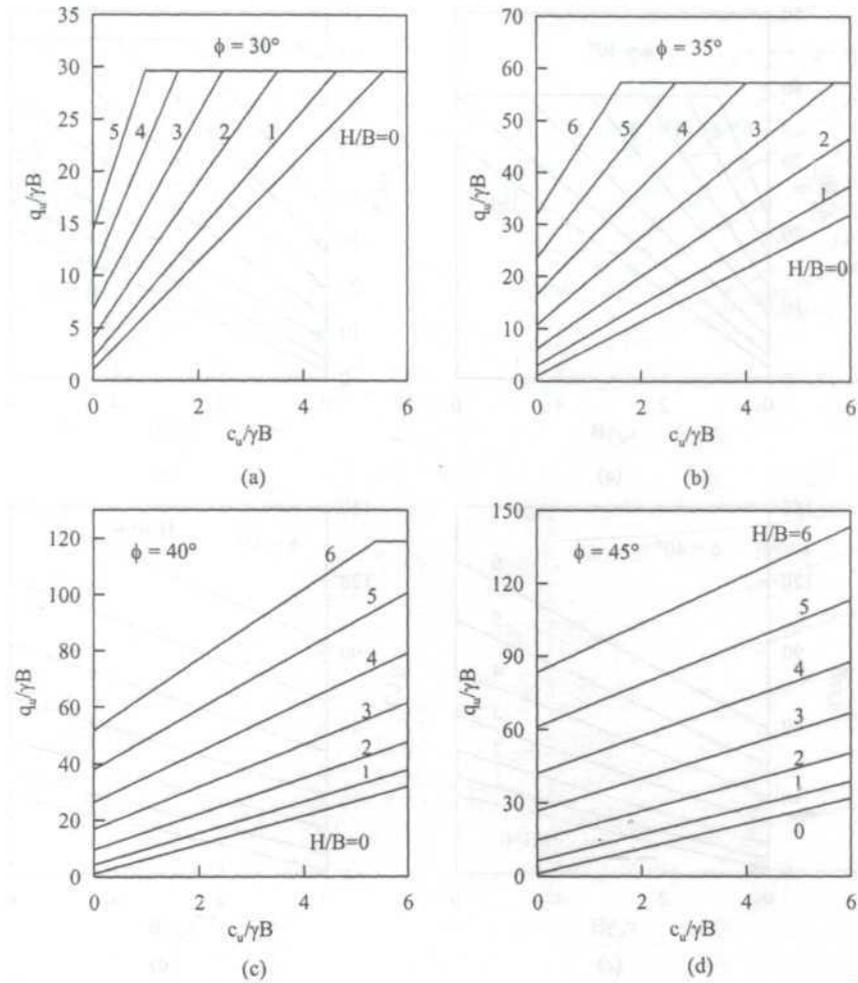


Fig. 13: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 1.0$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

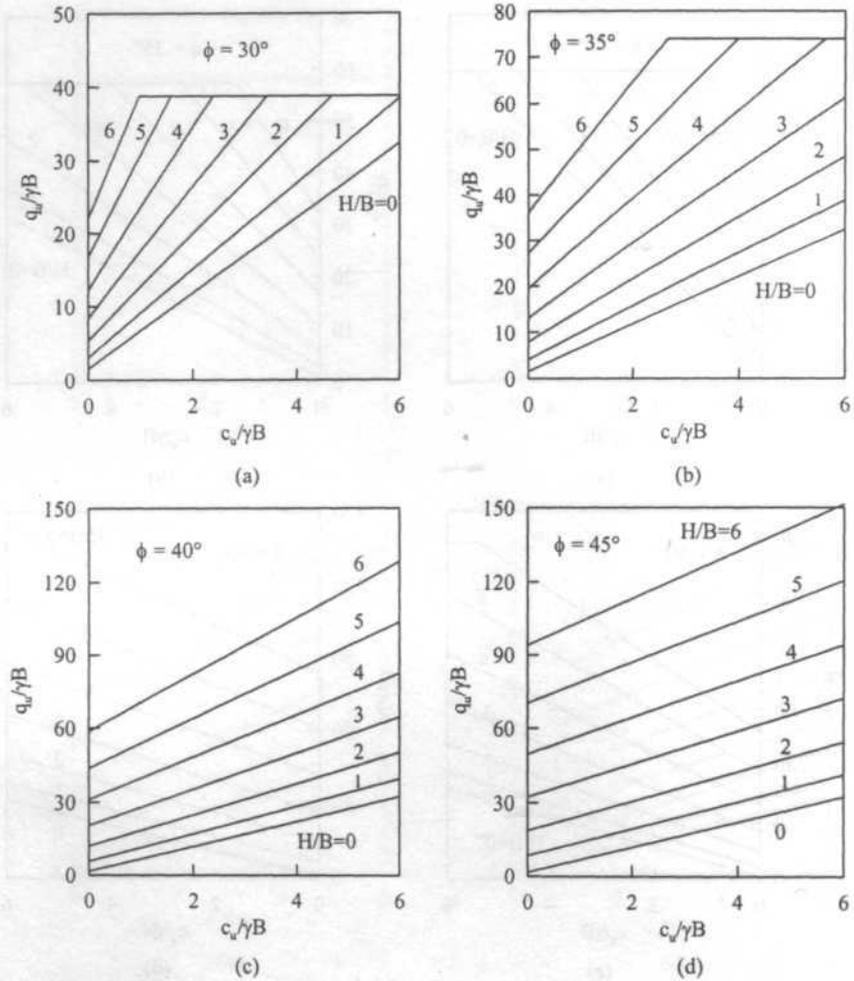


Fig. 14: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 1.5$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer

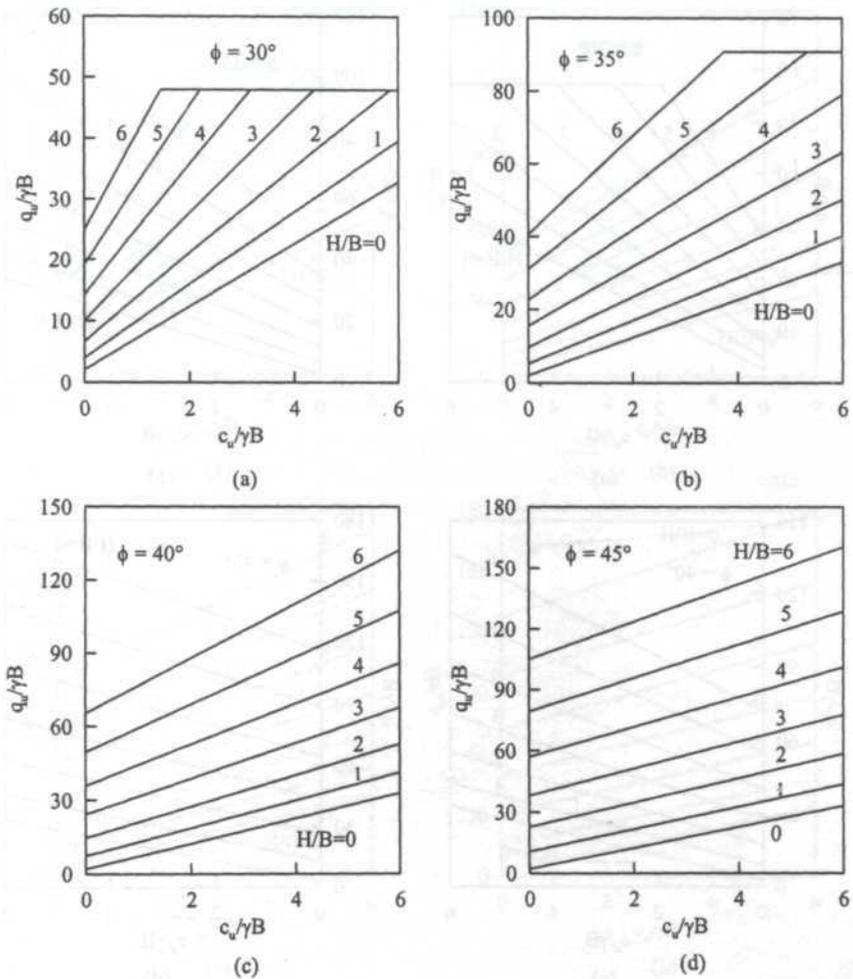


Fig. 15: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 2.0$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

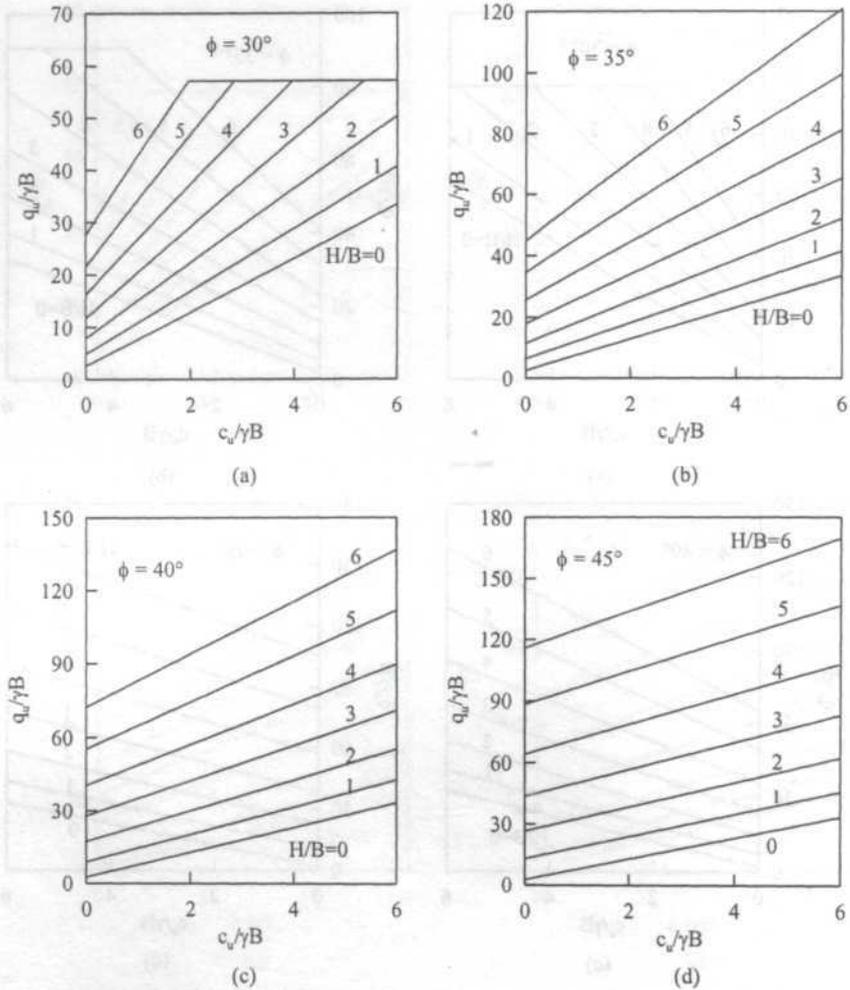


Fig. 16: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 2.5$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

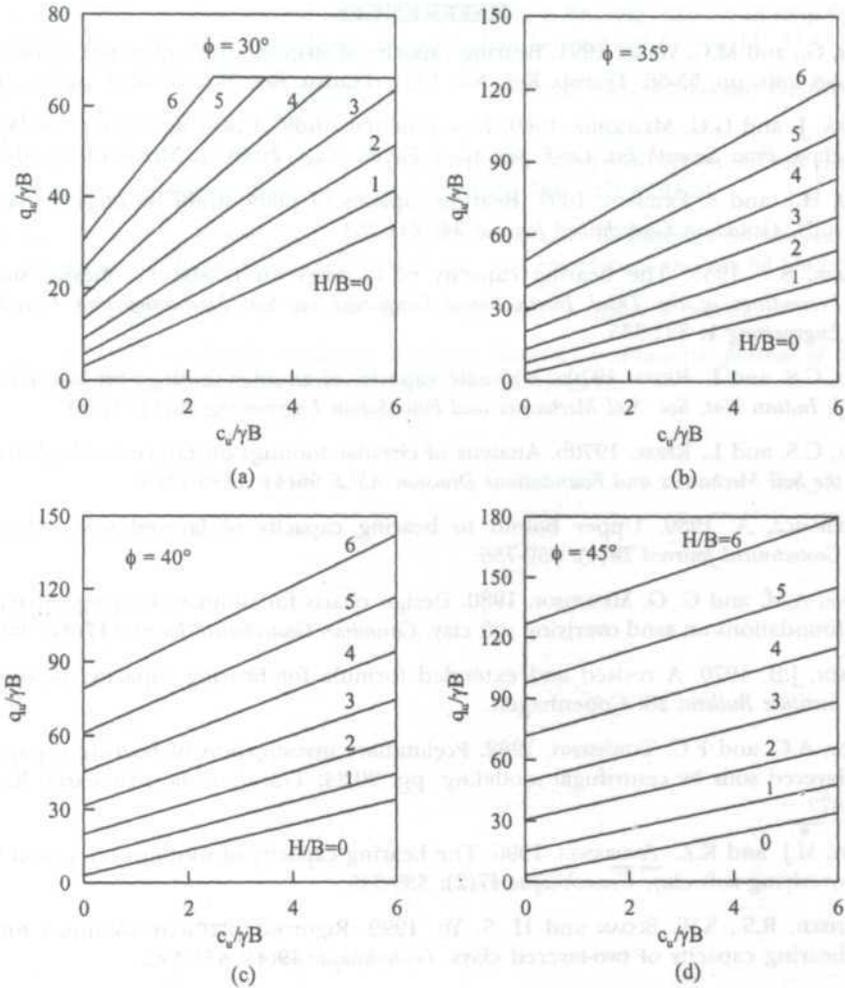


Fig. 17: Variation of $q_u/\gamma B$ with $c_u/\gamma B$ for sand-clay foundation soil at $D/B = 3.0$: (a) $\phi = 30^\circ$, (b) $\phi = 35^\circ$, (c) $\phi = 40^\circ$ and (d) $\phi = 45^\circ$

as a function of the properties of soils, the footing width, and the topsoil thickness. The effect of the punching shear parameter (δ) was considered and evaluated from empirical relationships that were developed based on the experimental results of Hanna and Meyerhof (1980). Based on this analysis, design charts were developed using the punching shear model in a dimensionless form since those of Hanna and Meyerhof (1980) were not presented in a nondimensionalised form, which limits their application. The presented charts here may be useful in overcoming the problem of design charts that were developed by Michalowski and Shi (1995) which may overestimate the bearing capacity by a significant amount depending on the very nature of the upper bound solution on which the derivation is based.

REFERENCES

- AZAM, G., and M.C. WANG. 1991. Bearing capacity of strip footing supported by two-layer c- ϕ soils. pp. 55-66. Transp. Res. No. 1331, Transp. Res. Board, Washington, D. C.
- BROWN, J. and G.G. MEYERHOF. 1969. Experimental study of bearing capacity in layered clays. *Proc. Seventh Int. Conf. Soil Mech. Found. Eng.*, 2: 45-51. Mexico City, Mexico.
- BURD, H.J. and S. FRYDMAN. 1997. Bearing capacity of plane-strain footings on layered soils. *Canadian Geotechnical Journal* 34: 241-253.
- BUTTON, S.S. 1953. The bearing capacity of footings on two-layer cohesive subsoil. *Proceedings of the Third International Conference on Soil Mechanics and Foundation Engineering* 1: 332-335.
- DESAI, C.S. and L. REESE. 1970a. Ultimate capacity of circular footings on layered soils. *J. Indian Nat. Soc. Soil Mechanics and Foundation Engineering* 96(1): 41-50.
- DESAI, C.S. and L. REESE. 1970b. Analysis of circular footings on layered soils. *Journal of the Soil Mechanics and Foundations Division ASCE* 96(4): 1289-1310.
- FLORKIEWICZ, A. 1989. Upper bound to bearing capacity of layered soils. *Canadian Geotechnical Journal* 26(4): 730-736.
- HANNA, A.M. and G. G. MEYERHOF. 1980. Design charts for ultimate bearing capacity of foundations on sand overlying soft clay. *Canadian Geotechnical Journal* 17(2): 300-303.
- HANSEN, J.B. 1970. A revised and extended formula for bearing capacity. *Geotechnical Institute Bulletin* 28: Copenhagen.
- HARDY, A.C. and F.C. TOWNSEND. 1982. Preliminary investigation of bearing capacity of layered soils by centrifugal modeling, pp. 20-24. Transportation Research Record 872.
- KENNY, M.J. and K.Z. ANDRAWES. 1996. The bearing capacity of footings on a sand layer overlying soft clay. *Geotechnique* 47(2): 339-345.
- MERIFIELD, R.S., S.W. SLOAN and H. S. YU. 1999. Rigorous plasticity solutions for the bearing capacity of two-layered clays. *Geotechnique* 49(4): 471-490.
- MEYERHOF, G.G. 1963. Some recent research on the bearing capacity of foundations. *Canadian Geotechnical Journal* 1(1): 16-26.
- MEYERHOF, G.G. 1974. Ultimate bearing capacity of sand overlying clay. *Canadian Geotechnical Journal* 11(2): 224-229.
- MEYERHOF, G.G. and A.M. HANNA. 1978. Ultimate bearing capacity of foundations on layered soil under inclined load. *Canadian Geotechnical Journal* 15(4): 565-572.
- MICHALOWSKI, R.L. and L. SHI. 1995. Bearing capacity of footings over two-layer foundation soils. *Journal of Geotechnical Engineering ASCE* 121(5): 421-428.
- OKAMURA, M., J. TAKEMURA and T. KIMURA. 1997. Centrifuge model tests bearing capacity and deformation of sand layer overlying clay. *Soils and Foundations* 37(1): 73-88.
- PRANDTL, L. 1921. Über die Eindringungsfestigkeit plastischer Baustoffe und die Festigkeit von Schneiden. *Zeitschrift für Angewandte Mathematik und Mechanik* 1(1): 15-20.

Derivation of Bearing Capacity Equation for a Two Layered System of Weak Clay Layer

- PURUSHOTHAMARAJ, P., B.K. RAMIAH and K.N. V.RAO. 1974. Bearing capacity of strip footing in two layered cohesive-friction soils. *Canadian Geotechnical Journal* **11**: 32-45.
- REISSNER, H. 1924. Zum Erddruckproblem. *Proceedings, First International Conference on Applied Mechanics*, pp. 295-311. Delft.
- SATYANARAYANA, A.M. and R.K. GARG. 1980. Bearing capacity of footings on layered C- ϕ soils. *Journal of Geotechnical Engineering Division ASCE* **106** Gt7: 819-824.
- SIVAREDDY, A. and R. J. SRINIVASAN. 1967. Bearing capacity of footings on layered clays. *Proceedings ASCE, Journal of the Soil Mechanics and Foundations Division* **93** SM2: 83-99.
- TERZAGHI, K. 1943. *Theoretical Soil Mechanics*. New York: J. Wiley & Sons.
- VESIC, A. S. 1973. Analysis of ultimate loads of Shallow foundations. *Journal of the Soil Mechanics and Foundations Division, ASCE* **99** SM1: 45-73.

Modelling of Changes in Evapotranspiration for an Area in Peninsular Malaysia

Lee Teang Shui, M.M.M. Najim, M. Aminul Haque & *Huang Yuk Feng

*Faculty of Engineering, Universiti Putra Malaysia
Serdang, 43400 Selangor, Malaysia*

**Kuala Lumpur Infrastructure University College
Kajang, 43000 Selangor*

Received: 10 June 2003

ABSTRAK

Satu kajian telah dijalankan untuk menyemakkan kepekaan taksiran penyejatpeluhan disebabkan perubahan iklim. Kajian ini menggunakan data 30 tahun daripada stesyen meteorologi terletak di estet padi FELCRA, Seberang Perak, Semenanjung Malaysia. Kesan disebabkan perubahan pembolehubah iklim, suhu, sinaran suria, kelembapan nisbi dan laju angin terhadap penyejatpeluhan dianalisis. Keputusan menunjukkan bahawa suhu min, kelembapan nisbi min, laju angin min dan sinaran sejagat berubah $+0.182\text{ }^{\circ}\text{C}$, -0.73% , -0.0365 m/s dan $+0.146\text{ MJ/m}^2$ sedekad masing-masing. Analisis statistik taksiran penyejatpeluhan hasil kaedah-kaedah terpilih menunjukkan bahawa kaedah-kaedah Penman-Monteith, Blaney-Criddle dan kaedah penyejatan Pan adalah taksiran serupa ($P=0.05$) dan adalah sesuai untuk kawasan kajian. Bila perubahan pada iklim sekarang dikuatkuasakan pada dekad-dekad akan datang, kadar taksiran melebihi berbentuk linear bagi kaedah-kaedah Blaney-Criddle dan Penman, sebaliknya berupa eksponen bagi kaedah Penman-Monteith. Tokohan keseluruhan yang dijangka selepas 5 dekad ialah 5.3% dan 6.9% bagi kaedah-kaedah Penman dan Blaney-Criddle masing-masing. Kaedah Penman-Monteith meramalkan tokohan penyejatpeluhan sebanyak 74.4% pada lima dekad akan datang.

ABSTRACT

A study was carried out to check the sensitivity of evapotranspiration estimation due to changes in climate. The study used 30 years of data from the meteorological station in the FELCRA paddy estate, Seberang Perak, Peninsular Malaysia. The effect of changes in the climatic variables, temperature, solar radiation, relative humidity and wind speed on evapotranspiration were analyzed. Results showed that the mean temperature, mean relative humidity, mean wind speed, and net global radiation have changed by $+0.182\text{ }^{\circ}\text{C}$, -0.73% , -0.0365 m/s , and $+0.146\text{ MJ/m}^2$ respectively per decade, while the short wave radiation received has decreased by 0.0037 MJ/m^2 per decade. The statistical analysis of the evapotranspiration estimations using selected methods showed that the Penman-Monteith, Blaney-Criddle and Pan evaporation methods give similar estimations ($P = 0.05$) and are suitable for the study area. When changes to the present climate are imposed for future decades, the over-estimation rate is linear with the Blaney-Criddle and Penman methods whereas for the Penman-Monteith method it is exponential. The overall increment expected after 5 decades is 5.3% and 6.9% with Penman and Blaney-Criddle methods respectively. The Penman-Monteith shows a 74.4% increment in the evapotranspiration over the next five decades.

Keywords: Evapotranspiration, estimation methods, climate change, global warming

INTRODUCTION

The emission of green house gases such as carbon dioxide; radioactively active gases and aerosol precursors modifies the climate and a major concern is temperature increase. The future global warming is predicted to increase by 0.1°C to 0.3°C over a period of a decade (Raper *et al.* 1996). The other climatic parameters that could be affected because of the global warming are rainfall, cloudiness, humidity, windiness and evaporation or evapotranspiration.

As any change in the climate will change the evaporation, evapotranspiration, rainfall and so on, the water budget in any region will be altered and this will have major consequences on agriculture. The climate change will affect water resources and consequently will have an impact on all the agricultural activities.

Chattopadhyay and Hulme (1997) reported that the temperature in southern and central parts of India increased for all the seasons from 1940 – 1990. They analysed evaporation time series data for different parts of India and showed that both pan evaporation and potential evapotranspiration have decreased in recent years. The main reasons behind this were the increased relative humidity and decreased radiation. They also suggested that the increase or change in potential evapotranspiration would be unequal between regions and seasons. Peterson *et al.* (1995) also reported a decrease in potential evapotranspiration and suggested that the increase in cloudiness and decrease in solar radiation are mostly responsible for this.

Thomas (2000) reported that an average decrease in potential evapotranspiration by 0.21% per year could be attributed to widespread reduction in short wave radiation. Brutsaert and Parlange (1998) suggested that the decrease in relative humidity due to the global climate change has increased the actual evaporation rates. This has led to a decrease in the potential evaporation.

However, Chattopadhyay and Hulme (1997) also found a bigger influence of the energy term in the Penman method (Penman 1948) on potential evapotranspiration change over tropics. As such, the evapotranspiration will increase if the relative humidity decreases and solar radiation increases due to a decrease in cloudiness because of global warming.

Increasing concentrations of greenhouse gases started to increase troposphere temperature. This phenomenon is affecting the cloudiness, humidity and windiness. Received solar radiation could change if cloudiness is affected by greenhouse warming. Humidity and wind speed might also be altered by greenhouse warming, inducing changes in climate. The objective of the study is to check the sensitivity of the evapotranspiration to the climate changes (temperature, solar radiation, relative humidity and wind speed) in the study area to select a proper method for estimating evapotranspiration.

METHODOLOGY

Study Area and Data Collection

The study area, Seberang Perak paddy estate, is located at 40° 7' N and 101° 4' E, and lies 10 km from the west coast of Peninsular Malaysia. Seberang Perak has a tropical climate characterised by a high rainfall of about 2100 mm with monthly peaks in April and October. Two peak-wet seasons are in March-April (rainfall between 175 - 200 mm per month) and October-November (rainfall between 200 - 300 mm per month). The distinct dry seasons are from December to February (150- 175 mm per month) and June to September (less than 150 mm per month).

Sunshine duration is about 7 hours or more from January to May while it decreases gradually to 5.5 hours from June to December. Net radiation is 17.0 MJm⁻² or more from February to September, while the lowest radiation is in November and December. Average air temperature in the project area is just above 26°C. The maximum temperature of the project area is about 32°C and the minimum is about 23°C, which are more or less uniform throughout the year. Total evaporation starts to increase from December to March/April reaching a maximum (>110 mm per month) and the minimum is recorded in November, which is less than 100 mm.

The climate data for this study were collected from the Sitiawan meteorological station of the Malaysian Meteorological Services. Daily values of data for a period of 30 years (1972 - 2001) were used for this study. The data collected were temperature (maximum, minimum), relative humidity (maximum, minimum), wind speed, solar radiation, sunshine duration, atmospheric pressure, and pan evaporation.

Evapotranspiration Methods and Simulation

The 30 years of daily data of mean temperature, mean relative humidity, wind speed and net global radiation were used to analyse the trends of change in a decade. Eight evapotranspiration estimation methods, the most common and widely used by other researchers for the purpose of evapotranspiration estimation, were tested for their validity for the study area. The eight methods tested in this study were Penman (Penman, 1948), Penman-Monteith (Monteith 1965 and 1981), Pan Evaporation, Kimberly-Penman (Jensen *et al.* 1990), Priestley-Taylor (1972), Hargreaves (Salazar *et al.* 1984), Samani-Hargreaves (Samani and Hargreaves 1985) and Blaney-Criddle (Allen and Pruitt 1986). The details of these methods are given in Appendix I.

The inputs (temperature, solar radiation, wind speed and relative humidity) were changed individually in all the methods between $\pm 30\%$ with a step variation of $\pm 10\%$. All the inputs were varied as a percentage of daily historical values. When monthly average values were needed for Blaney-Criddle, Hargreaves and Samani Hargreaves the daily-varied values were used in the calculation of the monthly average values of the required inputs. The climatic trends found in this study were next applied to forecast possible future trends in evapotranspiration.

RESULTS AND DISCUSSION

The analysis shows that the mean temperature has increased by 0.182°C per decade during the last 30 years. The mean relative humidity has dropped by 0.73% per decade. The mean wind speed has dropped by 0.0365 m/s per decade. The net global radiation has increased by 0.146 MJ/m² but the net incoming short wave radiation has decreased by 0.0037 MJ/m². Nicholls *et al.* (1996) reported that the earth surface temperature has increased since the 1850s by about 0.5°C or about 0.04°C per decade. Raper *et al.* (1996) estimated that the future global warming would modify the atmospheric temperature between 0.1 °C and 0.3 °C per decade. This study in the west coast of Peninsular Malaysia on the temperature increase trend shows agreement with those estimated by Raper *et al.* (1996). This study shows an increasing trend in pan evaporation that is attributed to the changes in climatic parameters, especially with the increase in temperature and decrease in relative humidity, similar to what Cohen *et al.* (2002) showed at Bet Dagan.

The monthly averages of the evapotranspiration estimates by all the eight methods were tested with a Randomised Complete Block Design where a mean separation procedure was done to verify the differences between different methods of estimations. The results of a two-way ANOVA are given in Table 1.

TABLE 1
Comparison of evapotranspiration estimation methods

Evapotranspiration estimation method	Mean*
Blaney-Criddle	3.276 a
Hargreaves	4.486 e
Kimberly-Penman	3.989 b
Pan	3.229 a
Penman	3.550 c
Penman Monteith	3.152 a
Priestley-Taylor	4.329 d
Samani-Hargreaves	4.454 e

*Values followed by the same letter are not significantly different at $P = 0.05$

The methods of Blaney-Criddle, Pan and Penman-Monteith give the lowest values and there are no significant differences among them ($P = 0.05$). All other five methods are significantly different from the Blaney-Criddle, Pan and Penman-Monteith methods. The estimates of Penman, Kimberly Penman and Priestley-Taylor methods significantly differ from each other. The methods of Hargreaves and Samani-Hargreaves give the highest values. These two methods do not have any significant differences between them ($P=0.05$). Thus, the methods of Blaney-Criddle, Pan and Penman-Monteith are suitable for the study area in the west coast of Peninsular Malaysia. However, the Penman-Monteith has universal acceptance (Allen *et al.* 1998).

Fig. 1 shows the percentage change in reference evapotranspiration to a percentage change in temperature. All the eight methods show a corresponding linear trend of increase/decrease of evapotranspiration with a temperature increase/decrease. The three temperature-based methods of Hargreaves (H in Fig. 1), Samani-Hargreaves (SH) and Blaney-Criddle (BC) are most sensitive to temperature change, with Hargreaves showing the highest trend while Blaney-Criddle the lowest. The Pan method (Pan) does not have any variable in terms of temperature. The Kimberly-Penman (KP) method shows the lowest variation with temperature change. All the combination methods used in this study, Kimberly-Penman, Penman (P), Penman-Monteith (PM) and Priestley-Taylor (PT) methods, respond in a similar way and the variation of evapotranspiration is more or less the same. Out of the four methods suitable for the project area, the Penman-Monteith gives 8.8% and 28.7% change in evapotranspiration for a change of +10% and +30% in temperature respectively, whereas the Blaney-Criddle shows a 9.0% and 27.0% while Penman shows 8.0% and 26.1% for similar changes.

Six methods have relative humidity as an input either directly or indirectly. Blaney-Criddle, Kimberly-Penman and Penman methods show a linear trend (Fig. 2). Evapotranspiration decreases with an increase in relative humidity when Blaney-Criddle, Kimberly-Penman and Penman methods are applied. Blaney-Criddle shows the largest change in evapotranspiration (-28.1% change in evapotranspiration for a +30% increment in relative humidity). Kimberly-Penman and Penman methods show -15.6% and -9.4% changes in evapotranspiration respectively for a +30% increase in relative humidity. Pan, Penman-Monteith and Priestley-Taylor show linear relationships where an increase in relative humidity increases the evapotranspiration.

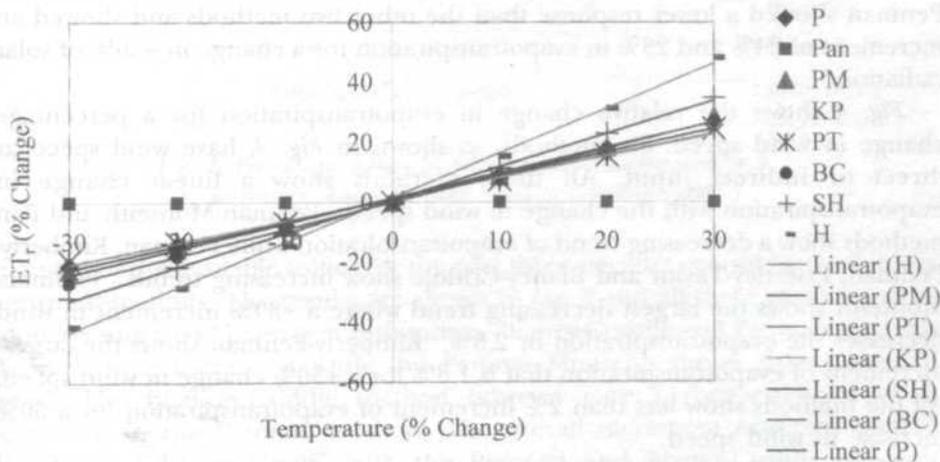


Fig. 1: Percentage change in reference evapotranspiration to a percentage change in temperature at Seberang Perak

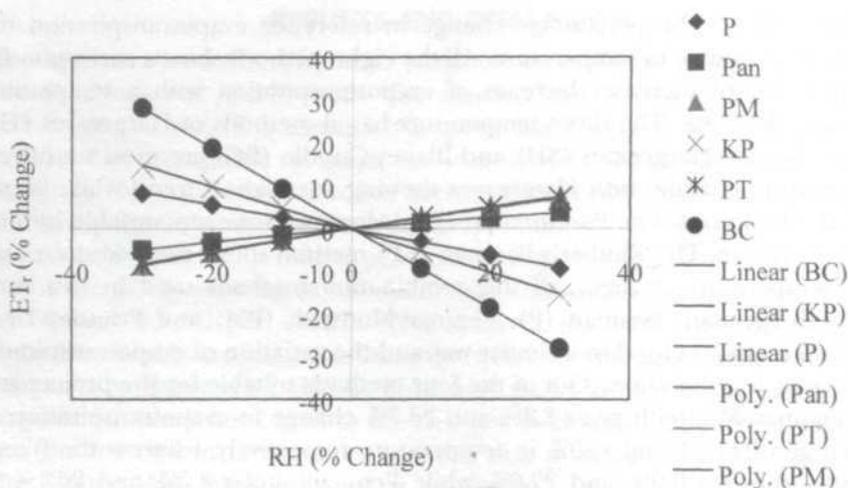


Fig. 2: Percentage change in reference evapotranspiration to a percentage change in relative humidity at Seberang Perak

Solar radiation received at a location varies with the change in cloudiness due to global warming. Out of eight methods tested in this study, all combination methods (Penman, Penman-Monteith, Kimberly-Penman, and Priestley-Taylor) use solar radiation as an input. All four methods show a linear trend in evapotranspiration estimates with a percentage change in solar radiation (Fig. 3). With the Penman-Monteith and Priestley-Taylor, a +30% increment in solar radiation results in estimating 26.5% more evapotranspiration. Penman-Monteith did not response more than 20% when solar radiation is incremented by +30% in the study by McKenney and Rosenberg (1993). Penman and Kimberley-Penman showed a lower response than the other two methods and showed an increment of 24% and 23% in evapotranspiration for a change in +30% of solar radiation.

Fig. 4 shows the relative change in evapotranspiration for a percentage change in wind speed. Six methods, as shown in Fig. 4, have wind speed as direct or indirect input. All these methods show a linear change in evapotranspiration with the change in wind speed. Penman-Monteith and Pan methods show a decreasing trend of evapotranspiration while Penman, Kimberly-Penman, Priestley-Taylor and Blaney-Criddle show increasing trends. Penman-Monteith shows the largest decreasing trend where a +30% increment in wind decreases the evapotranspiration by 2.5%. Kimberly-Penman shows the largest increment of evapotranspiration that is 1.8% for a +30% change in wind speed. All the methods show less than 2% increment of evapotranspiration for a 30% increase in wind speed.

The average trend in climate change in the study area was used in a simulation to evaluate the future trends in evapotranspiration estimation by the three methods (Penman-Monteith, Penman and Blaney-Criddle) that gave non-significant differences. The pan evaporation will vary with the temperature

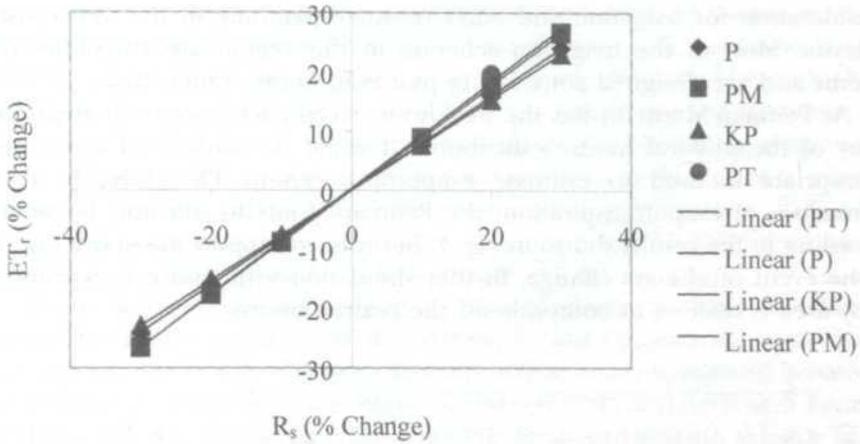


Fig. 3: Percentage change in reference evapotranspiration to a percentage change in net global radiation at Seberang Perak

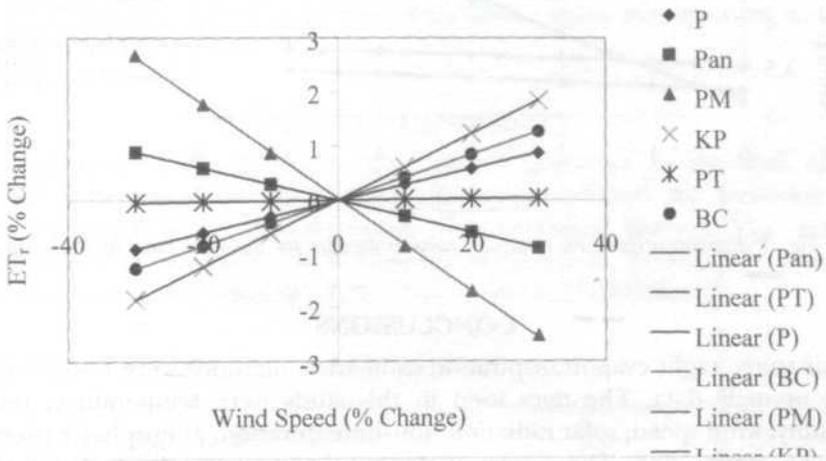


Fig. 4: Percentage change in reference evapotranspiration to a percentage change in wind speed at Seberang Perak

change but the available values for the past three decades cannot be used in the future estimations. The results are shown in Fig. 5 and all three methods show an increasing trend in evapotranspiration. Blaney-Criddle and Penman methods show a linear increment while the Penman-Monteith shows an exponential trend. The Blaney-Criddle method however gave higher estimates when compared to the Penman method. The overall increment expected after 5 decades is 5.3% and 6.9% with the Penman and Blaney-Criddle methods respectively. The Penman-Monteith shows a 74.4% increment in the evapotranspiration in the next five decades. Penman-Monteith shows the same exponential trend when average monthly evapotranspiration for each month is considered. The increasing trend of evapotranspiration has to be taken into

consideration for irrigation and water resource planning in the west coast of Malaysia. Most of the irrigation schemes in this region are run-of-the river systems and are designed considering past evapotranspiration data.

As Penman-Monteith has the worldwide acceptance because it encumbers many of the physical nature's attributes, it could be considered as the most appropriate method to estimate evapotranspiration. Therefore, in future estimations of evapotranspiration, the Penman-Monteith will only be suitable according to the results shown in Fig. 5. In order to support the above findings in the event of climate change, further simulation with future data from the study area is needed to comprehend the real variations.

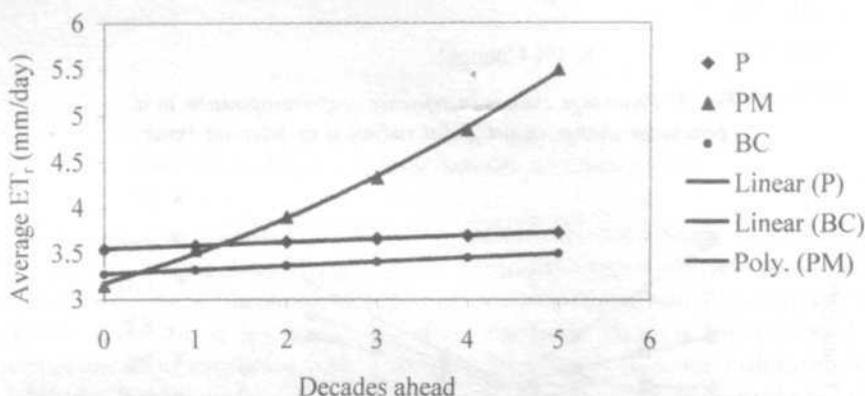


Fig. 5: Evapotranspiration trends in coming decades for the West coast of Malaysia

CONCLUSIONS

In this study, eight evapotranspiration estimation methods were tested with 30 years of daily data. The data used in this study were temperature, relative humidity, wind speed, solar radiation, sunshine duration, atmospheric pressure, and pan evaporation. The results of these estimates were checked statistically and it was found that the Penman-Monteith, Blaney-Criddle and Pan methods estimate lower values of evapotranspiration with no significant difference among them ($P = 0.05$). All the other methods were significantly different from these three methods. The analysis of the last 30 years data showed that the mean temperature has increased by 0.182°C per decade. The mean relative humidity has dropped by 0.73% per decade. The mean wind speed has dropped by 0.0365 m/s per decade. The net global radiation has increased by 0.146 MJ/m^2 but the net incoming short wave radiation has decreased by 0.0037 MJ/m^2 .

All the eight methods show a linear trend of increase or decrease of evapotranspiration with corresponding temperature increase or decrease. The three temperature-based methods of Hargreaves, Samani-Hargreaves and Blaney-Criddle are most sensitive to temperature change. Evapotranspiration decreases with an increase in relative humidity with the Blaney-Criddle, Kimberly-Penman

and Penman methods. The Penman-Monteith and Priestley-Taylor methods show the highest effect on evapotranspiration where a +30% increment in solar radiation results in 26.5% more evapotranspiration. Penman-Monteith shows the largest decreasing trend where a +30% increment in wind decreases the evapotranspiration by 2.5%. Logically, when wind speed increases, evaporation should follow suit. Thus there could be a limitation as to the values of physical parameters that can be input into the Penman-Monteith equation and in fact into all the other equations. In other words there will be limitations to the applications of these equations, which should not be alarming.

When the present trend in climate change (based on historical data) is imposed for future decades, the Blaney-Criddle and Penman methods show a linear increment while the Penman-Monteith indicates an exponential behaviour. The overall increment expected after 5 decades is 5.3% and 6.9% with Penman and Blaney-Criddle methods respectively. The Penman-Monteith shows a 74.4% increment in the evapotranspiration in the next five decades. Penman-Monteith shows the same exponential trend when average monthly evapotranspiration for each month is considered. Since it is the FAO recommended method, serious concerns will have to be placed on future water management to avoid being caught unprepared.

ACKNOWLEDGEMENTS

The authors wish to expressive their sincere gratitude to the staff of the FELCRA Seberang Perak Paddy Estate Irrigation Scheme, the Drainage and Irrigation Department, the Malaysian Meteorological Service. The authors would also like to thank The Ministry of Science, Technology and the Environment for the funding of the Project IRPA 01-02-04-0422.

REFERENCES

- ALLEN, R.G. and W.O. PRUITT. 1986. Rational use of the FAO Blaney-Criddle formula. *J. Irrig. Drain. Eng.* **112(IR2)**: 139-155.
- ALLEN, R.G., L.S. PEREIRA, D. RAES and M. SMITH. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper-56. Rome, Italy.
- BRUTSAERT, W. and M.B. PARLANGE. 1998. Hydrologic cycle explains the evaporation paradox. *Nature* **396**: 30.
- CHATTOPADHYAY, N. and M. HULME. 1997. Evaporation and potential evapotranspiration in India under conditions of recent and future climate change. *Agricultural and Forest Meteorology* **87**: 55 - 73.
- COHEN, S., A. IANETZ and G. STANHILL. 2002. Evaporative climate changes at Bet Dagan, Israel, 1964-1998. *Agricultural and Forest Meteorology* **111**: 83 - 91.
- JENSEN, M.E., R.D. BURMAN and R.G. ALLEN. 1990. Evapotranspiration and irrigation water requirements. ASCE Manual 70.