

# Effect of Production Time on Energy Use of a Tube-Ice Making Tower

**Chittin Tangthieng**

Department of Mechanical Engineering, Faculty of Engineering, Chulalongkorn University,  
Phayathai Road, Pathum Wan, Bangkok, 10330, Thailand  
E-mail: fmectt@eng.chula.ac.th

## Abstract

One of the important industries for a tropical country is the ice-making industry. In particular, tube ice is a form of ice for consuming, which is available in many grocery stores. This research is a numerical study of the effect of the production cycle time on the energy use of a tube-ice making process. The system is assumed one dimensional in the radial direction and unsteady. The governing equations are composed of the wall and the ice regions. The boundary condition at the wall surface is a convective type whereas the boundary condition at the interface is a isothermal solidification. The governing system is transformed into a dimensionless form, which is numerically solved by the finite difference method. The numerical solution in terms of ice thickness is compared with the data measured from the ice factory. There is a qualitative agreement between the numerical prediction and data obtained from the field measurement. From the numerical prediction, it can be seen that the average energy consumption is increased by approximately 5.8 kJ/m per minute with increase of the production cycle time. On the other hand, as the production cycle time increases, the average specific energy consumption reaches a minimum value of 0.341 kJ/g at a production cycle time of 14 minutes. Reduction of the production cycle time from 28 to 14 minutes leads to the higher value of the average specific energy consumption by 0.7 percent.

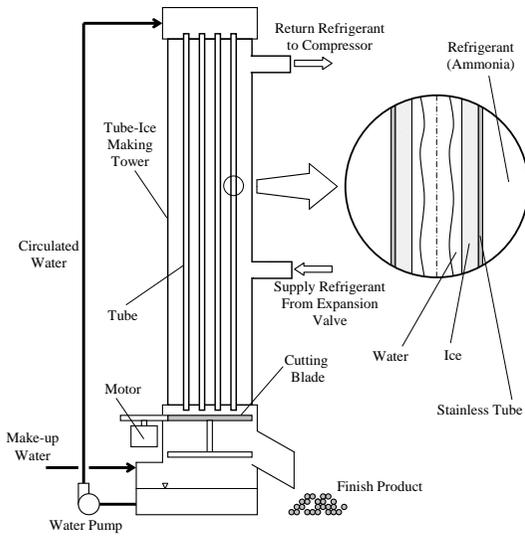
**Keywords:** tube ice, numerical method, production cycle time, energy consumption

## 1. Introduction

Tube ice is one of the important consuming products, especially for a tropical country. Because of its hygiene, tube ice is a favorable product for most consumers. Nevertheless, the limitation of the ice making industry is the transportation over a long distance. Therefore, tube-ice making industries are considered local ones, which are well-distributed in most of the districts in the country.

In a tube-ice making plant, the tube-ice making tower, which serves as an evaporator of the refrigeration system, is a major component to produce tube ice.

Figure 1 illustrates the components of the tube-ice making tower. In general, the refrigerant is ammonia, and the saturation temperature is set at  $-8^{\circ}\text{C}$ . Heat transfer occurs between ammonia and liquid water through the tube. The liquid water is circulated by being pumped to the top of the tower, and it falls down to the bottom sump by gravity. Make-up water is added to compensate for the formation of ice. The tube is usually made of stainless steel with a nominal diameter of  $1\frac{1}{4}$  inch.



**Figure 1** Components of tube-ice making tower

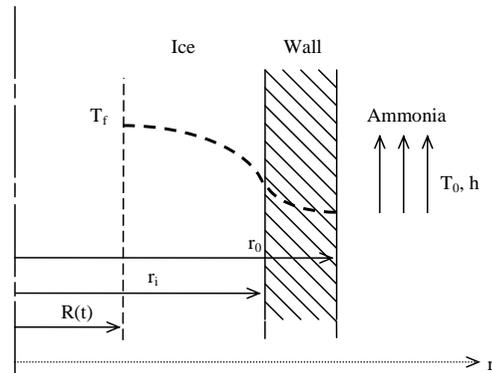
During tube-ice production, ice will form on the inner surface of the tube and grow until the diameter of the inner core is approximately 1 cm. Thereafter, the defrost process begins, and the tube ice is cut to a desired length by falling through a cutting blade at the bottom. The total production cycle is approximately 28 minutes, whereas the defrost period takes only 5 to 10 minutes.

In this study, a numerical method is employed to predict the average energy consumption, and the average specific energy consumption (SEC) of the tube-ice making process. A comparison between the numerical predictions and the field data is made. The effect of the different production cycle time on the average energy consumption and the average SEC is investigated. The results of this research may be used as a guideline to improve the energy efficiency of the tube-ice making process.

## 2. Problem Formulation

To formulate the mathematical model, the major assumptions for the tube-

ice making process are made as follows: (i) The problem is one dimensional and transient. (ii) The interface where solidification takes place is sharp and isothermal at  $T_f$ . (iii) Ammonia is at the saturation temperature,  $T_0$ , with a convective heat transfer coefficient of  $h$ . (iv) The initial temperature of the tube is also at  $T_0$ . (v) Convection of liquid water is neglected. The schematic diagram of the tube-ice making problem is shown in Figure 2.



**Figure 2** The schematic diagram of the problem under consideration

According to these assumptions, the mathematical models for the tube-ice making process are [1]:

(i) Ice Region

$$\frac{1}{\alpha_s} \frac{\partial T_s}{\partial t} = \frac{\partial^2 T_s}{\partial r^2} + \frac{1}{r} \frac{\partial T_s}{\partial r} \quad (1)$$

$$r = R(t); T_s = T_f, k_s \frac{\partial T_s}{\partial r} = \rho_s \Delta H \frac{dR}{dt} \quad (2a)$$

$$r = r_i; T_s = T_w, k_s \frac{\partial T_s}{\partial r} = k_w \frac{\partial T_w}{\partial r} \quad (2b)$$

$$t = 0; R(0) = r_i \quad (2c)$$

(ii) Wall Region

$$\frac{1}{\alpha_w} \frac{\partial T_w}{\partial t} = \frac{\partial^2 T_w}{\partial r^2} + \frac{1}{r} \frac{\partial^2 T_w}{\partial r^2} \quad (3)$$

$$r = r_i; T_s = T_w, k_s \frac{\partial T_s}{\partial r} = k_w \frac{\partial T_w}{\partial r} \quad (4a)$$

$$r = r_0; -k_w \frac{\partial T_w}{\partial r} = h(T_w - T_0) \quad (4b)$$

$$t = 0; T_w = T_0 \quad (4c)$$

### 3. Mathematical Analysis

To facilitate the computational procedure, a coordinate transformation is employed to the governing equations by introducing the following dimensionless variables:

$$\hat{t} = \frac{\alpha_s t}{(r_0 - r_i)^2} = \frac{\alpha_s t}{D^2} \quad (5)$$

$$\hat{r}_s = \frac{r - r_i}{r_i - R(t)} \quad (6)$$

$$\hat{r}_w = \frac{r - r_i}{r_0 - r_i} \quad (7)$$

$$\phi = \frac{r_i - R(t)}{r_0 - r_i} = \frac{r_i - R(t)}{D} \quad (8)$$

$$\theta_s = \frac{T_s - T_0}{T_f - T_0} \quad (9)$$

$$\theta_w = \frac{T_w - T_0}{T_f - T_0} \quad (10)$$

Substituting the dimensionless variables into equations (1-4) gives:

#### (i) Ice Region

$$\frac{\partial^2 \theta_s}{\partial \hat{r}_s^2} + \frac{\phi}{\hat{r}_s \phi + R_r} \frac{\partial \theta_s}{\partial \hat{r}_s} + \left( \hat{r}_s \phi \frac{d\phi}{d\hat{t}} \right) \frac{\partial \theta_s}{\partial \hat{r}_s} - \phi^2 \frac{\partial \theta_s}{\partial \hat{t}} = 0 \quad (11)$$

$$\hat{r}_s = 0; \theta_s = \theta_w, \frac{\partial \theta_s}{\partial \hat{r}_s} - R_1 \phi \frac{\partial \theta_w}{\partial \hat{r}_w} = 0 \quad (12a)$$

$$\hat{r}_s = -1; \theta_s = 1, \frac{\partial \theta_s}{\partial \hat{r}_s} + \frac{\phi}{Ste} \frac{d\phi}{d\hat{t}} = 0 \quad (12b)$$

$$\hat{t} = 0; \phi = 0 \quad (12c)$$

#### (ii) Wall Region

$$\frac{\partial^2 \theta_w}{\partial \hat{r}_w^2} + \frac{1}{\hat{r}_w + R_r} \frac{\partial \theta_w}{\partial \hat{r}_w} - \frac{R_2}{R_1} \frac{\partial \theta_w}{\partial \hat{t}} = 0 \quad (13)$$

$$\hat{r}_w = 0; \theta_s = \theta_w, \frac{\partial \theta_s}{\partial \hat{r}_s} - R_1 \phi \frac{\partial \theta_w}{\partial \hat{r}_w} = 0 \quad (14a)$$

$$\hat{r}_w = 1; \frac{\partial \theta_w}{\partial \hat{r}_w} + Bi \theta_w = 0 \quad (14b)$$

$$\hat{t} = 0; \theta_w = 0 \quad (14c)$$

It can be seen that the moving boundary conditions (2b) at  $r=R(t)$  are transformed to the fixed boundary conditions (12b) at  $\hat{r}_s = -1$ , which are favorable to the numerical procedure [2]. In addition, five dimensionless parameters appearing in equations (11-14) are:

$$R_1 = \frac{k_w}{k_s} \quad (15)$$

$$R_2 = \frac{\rho_w C_{pw}}{\rho_s C_{ps}} \quad (16)$$

$$Ste = \frac{C_{ps}(T_f - T_0)}{\Delta H} \quad (17)$$

$$Bi = \frac{h D}{k_w} \quad (18)$$

$$R_r = \frac{r_i}{r_0 - r_i} = \frac{r_i}{D} \quad (19)$$

In this study,  $R_1$ ,  $R_2$ ,  $R_r$  and  $Ste$  are treated as constants, whereas  $Bi$  is not due to the convective heat transfer coefficient of the nucleate boiling.

### 4. Finite Difference Analysis

Equations (11-14) can be solved by using the finite difference method. Because of the parabolic partial differential equations (11-14), the fully implicit scheme is employed to ensure numerical stability [3]. The partial differential terms appearing in equations (11) and (13) are replaced with the finite difference approximation as

follows:

$$\frac{\partial^2 \theta}{\partial \hat{r}^2} = \frac{\theta_{i+1}^{n+1} - 2\theta_i^{n+1} + \theta_{i-1}^{n+1}}{(\Delta \hat{r})^2} \quad (20)$$

$$\frac{\partial \theta}{\partial \hat{r}} = \frac{\theta_{i+1}^{n+1} - \theta_{i-1}^{n+1}}{2\Delta \hat{r}} \quad (21)$$

$$\frac{\partial \theta}{\partial \hat{t}} = \frac{\theta_i^{n+1} - \theta_i^n}{\Delta \hat{t}} \quad (22)$$

By substituting equations (20-22) into the governing equations (11-14) and rearranging, the finite difference equations for nodes at the different locations, i.e., the interior nodes, the interfacial node at  $\hat{r}_s = \hat{r}_w = 0$ , and the boundary node at  $\hat{r}_w = 1$ , become:

(i) Interior Nodes

$$\begin{aligned} & \left(1 - K \frac{\Delta \hat{r}}{2}\right) \theta_{i-1}^{n+1} - \left(2 + L \frac{\Delta \hat{r}^2}{\Delta \hat{t}}\right) \theta_i^{n+1} + \left(1 + K \frac{\Delta \hat{r}}{2}\right) \theta_{i+1}^{n+1} \\ & = - \left(L \frac{\Delta \hat{r}^2}{\Delta \hat{t}}\right) \theta_i^n \end{aligned} \quad (23)$$

For the ice region:

$$K = \frac{\varphi}{\hat{r}_s \varphi + R_r} + \hat{r}_s \varphi \frac{d\varphi}{d\hat{t}}, \quad L = \varphi^2, \quad \Delta \hat{r} = \Delta \hat{r}_s, \quad \theta = \theta_s \quad (24)$$

For the wall region:

$$K = \frac{1}{\hat{r}_w + R_r}, \quad L = \frac{R_2}{R_1}, \quad \Delta \hat{r} = \Delta \hat{r}_w, \quad \theta = \theta_w \quad (25)$$

(ii) Interfacial Node ( $\hat{r}_s = \hat{r}_w = 0$ ) [4]

$$\begin{aligned} & \left(1 - \frac{K_b \Delta \hat{r}_b}{2}\right) \theta_{i-1}^{n+1} - \left[\left(\frac{\Delta \hat{r}_a}{\Delta \hat{r}_b} + \frac{L_b \Delta \hat{r}_a \Delta \hat{r}_b}{2 \Delta \hat{t}}\right) \left(1 + \frac{K_a \Delta \hat{r}_a}{2}\right) \frac{J_b}{J_a}\right] \theta_i^{n+1} \\ & + \left(1 + \frac{L_a \Delta \hat{r}_a^2}{2 \Delta \hat{t}}\right) \left(1 - \frac{K_b \Delta \hat{r}_b}{2}\right) \theta_{i+1}^{n+1} + \left(1 + \frac{K_a \Delta \hat{r}_a}{2}\right) \left(\frac{J_b \Delta \hat{r}_a}{J_a \Delta \hat{r}_b}\right) \theta_{i+1}^{n+1} \\ & = - \left[\left(\frac{L_a \Delta \hat{r}_a^2}{2 \Delta \hat{t}}\right) \left(1 - \frac{K_b \Delta \hat{r}_b}{2}\right) + \left(\frac{L_b \Delta \hat{r}_a \Delta \hat{r}_b}{2 \Delta \hat{t}}\right) \left(1 + \frac{K_a \Delta \hat{r}_a}{2}\right) \frac{J_b}{J_a}\right] \theta_i^n \end{aligned} \quad (26)$$

where

$$\begin{aligned} K_a &= \frac{\varphi}{R_r}, \quad L_a = \varphi^2, \quad J_a = 1, \quad \Delta \hat{r}_a = \Delta \hat{r}_s, \quad \theta = \theta_s = \theta_w, \\ K_b &= \frac{1}{R_r}, \quad L_b = \frac{R_2}{R_1}, \quad J_b = R_1 \varphi, \quad \Delta \hat{r}_b = \Delta \hat{r}_w \end{aligned} \quad (27)$$

(iii) Boundary Node ( $\hat{r}_w = 1$ )

$$\begin{aligned} & 2\theta_{w,i-1}^{n+1} - \left(2 + L \frac{\Delta \hat{r}_w^2}{\Delta \hat{t}} + 2 \text{Bi} \Delta \hat{r}_w \left(1 + K \frac{\Delta \hat{r}_w}{2}\right)\right) \theta_{w,i}^{n+1} \\ & = - \left(L \frac{\Delta \hat{r}_w^2}{\Delta \hat{t}}\right) \theta_{w,i}^n \end{aligned} \quad (28)$$

where

$$K = \frac{1}{1 + R_r}, \quad L = \frac{R_2}{R_1} \quad (29)$$

Equations (23-29) are a set of nonlinear equations due to the appearance of  $\varphi$  and  $d\varphi/dt$  terms. Thus, to linearize the equations, those two terms must be replaced by an approximation as follows:

$$\varphi = \frac{\varphi_i^{n+1} + \varphi_i^n}{2} \quad (30)$$

$$\frac{d\varphi}{d\hat{t}} = \frac{\varphi_i^{n+1} - \varphi_i^n}{\Delta \hat{t}} \quad (31)$$

By substituting equations (30-31) into equations (23-29), a set of these equations will be linearized in the form of a tridiagonal system. They can be solved by using Thomas algorithm. Note that an initial estimation of  $\varphi_i^{n+1}$  is required to run the numerical scheme. To verify the value of  $\varphi_i^{n+1}$ , the boundary condition (12b) will be utilized and be rewritten in a finite difference form.

$$\varphi_i^{n+1} = \sqrt{\varphi_i^{n,2} + \left(\frac{\text{Ste} \Delta \hat{t}}{\Delta \hat{r}_s}\right) \left(\theta_{s,i+2}^{n+1} - 4\theta_{s,i+1}^{n+1} + 3\theta_{s,i}^{n+1}\right)} \quad (32)$$

The value of  $\phi_i^{n+1}$  will be iterated until convergence.

It is noticed that Bi can be written as a function of h, which is no longer a constant due to nucleated boiling. In this case, the Stefan and Abdelsalan correlation [5] is utilized. By substituting the properties of ammonia [6] into this correlation, h for the nucleated boiling of ammonia at the saturation temperatures of  $-8^\circ\text{C}$  can be written as:

$$h = 6.307 \times 10^6 \left( \frac{T_{w,0} - T_0}{T_0} \right)^{2.0303} \quad (33)$$

The subscript w,0 represents the outer surface temperature of the tube. Rewriting equation (33) into a dimensionless form yields:

$$Bi = 1.304 \theta_{w,0}^{2.0303} \quad (34)$$

During the computational scheme, Bi will be iterated until convergence. When both  $\phi_i^{n+1}$  and Bi have converged, the program will record the temperature profile and the ice thickness, and then proceed to the next time step until the final (or production cycle) time, where ice thickness is approximately 12.5 mm, is reached.

It should be noted here that the physical initial condition at  $\hat{t} = 0$  is a singularity due to the zero ice thickness, and it cannot be used as an initial condition for the numerical scheme. To avoid this difficulty, the numerical initial condition must be shifted to initial time at  $\hat{t} \ll 1$ , at which the radius of curvature of the tube is much greater than the ice thickness. The temperature distribution and the ice thickness at  $\hat{t} \ll 1$  can be determined by the similarity solution [7].

The values of  $R_1$ ,  $R_2$ , and Ste can be determined by substituting the properties of

stainless steel and properties of ice given in Tables 1 and 2 into equations (15-17). For the value of  $R_r$ , the inner diameter of a standard 1¼-inch tube is 35.04 mm with the wall thickness of 3.56 mm. As a result, the values  $R_1$ ,  $R_2$ , Ste, and  $R_r$  are:

$$\begin{aligned} R_1 &= 7.158, R_2 = 1.863, \\ Ste &= 0.0485, R_r = 4.921 \end{aligned} \quad (35)$$

**Table 1** Properties of ice [8]

Properties	Values
$\rho_s$ (kg/m <sup>3</sup> )	920
$k_s$ (W/m-K)	1.91
$C_{ps}$ (kJ/kg-K)	2.022
$\Delta H$ (kJ/kg)	333.7

**Table 2** Properties of stainless steel [8]

Properties	Values
$\rho_w$ (kg/m <sup>3</sup> )	7,900
$k_w$ (W/m-K)	14.1
$C_{pw}$ (kJ/kg-K)	0.451

The computational results are represented in terms of the average energy consumption ( $EC_{avg}$ ) and the average specific energy consumption ( $SEC_{avg}$ ) at different production cycle times. Mathematically,  $EC_{avg}$  is the integral of the cooling-transfer-rate curve over a given production cycle time ( $t_p$ ) whereas the  $SEC_{avg}$  is the average integral of the SEC curve over a given production cycle time:

$$EC_{avg} = \int_{t=0}^{t=t_p} (\text{Cooling Transfer Rate}) dt \quad (36)$$

$$SEC_{avg} = \frac{1}{t_p} \int_{t=0}^{t=t_p} (SEC) dt \quad (37)$$

The cooling transfer rate is given by:

$$\begin{aligned} \text{Cooling Transfer Rate} &= 2\pi r_0 h (T_{w,0} - T_0) \\ &= 2\pi k_w (T_f - T_0) (R_r + 1) \text{Bi} \theta_{w,0} \end{aligned} \quad (38)$$

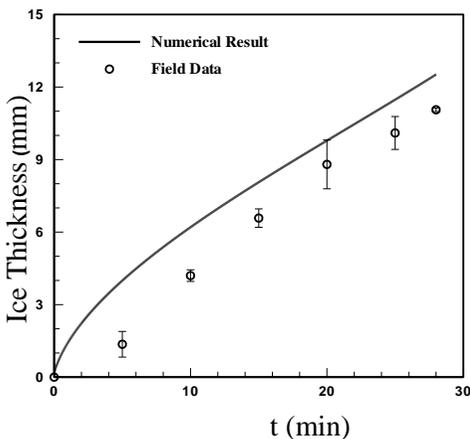
The SEC is defined as the ratio of the cooling transfer rate to the production rate, which is given by:

$$\begin{aligned} \text{Production Rate} &= \rho_s \frac{dA_{ice}}{dt} = -2\pi \rho_s R \frac{dR}{dt} \\ &= 2\pi \frac{k_s}{C_{ps}} (R_r - \psi) \frac{d\psi}{d\hat{t}} \end{aligned} \quad (39)$$

$$\begin{aligned} \text{SEC} &= \frac{\text{Cooling Transfer Rate}}{\text{Production Rate}} \\ &= \frac{C_{ps} (T_f - T_0) R_l (R_r + 1) \text{Bi} \theta_{w,0}}{(R_r - \psi) (d\psi / d\hat{t})} \end{aligned} \quad (40)$$

### 5. Results and Discussion

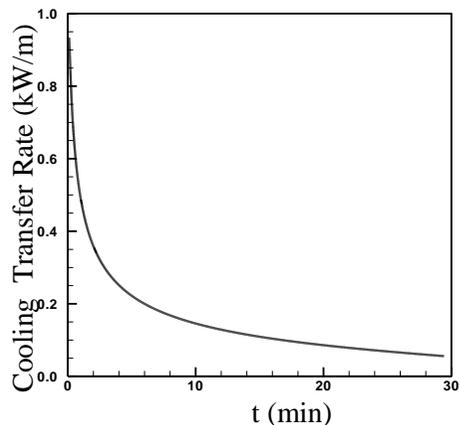
To verify the validity of numerical results, a comparison between the ice thickness obtained from the numerical prediction and that obtained from the field data measuring is made and illustrated in Figure 3.



**Figure 3** Comparison of the ice thickness obtained from two methods

In Figure 3,  $T_0$  is set at  $-8^\circ\text{C}$ . It is noticed that both the numerical results and field data are in qualitative agreement. The ice thickness obtained from the numerical prediction is greater than that obtained from the field data. It is explained that the beginning of ice formation on the inner surface of the tube is not at  $t = 0$  for the field data. This may stem from heat accumulation in the system after the defrost process. According to Figure 3, if the trend line is drawn based on the field data measurements, the trend line will intersect the horizontal axis at approximately 3 minutes. As a result, ice will be formed about 3 minutes after the beginning of the refrigeration process.

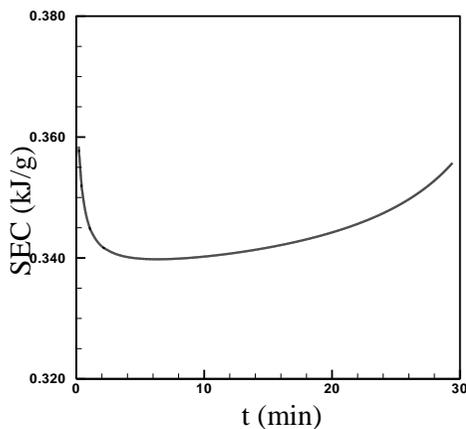
Figure 4 illustrates the relation of the cooling transfer rate obtained from the numerical prediction as a function of time. It can be seen that after the beginning of the process, the cooling transfer rate decreases rapidly. Thereafter, it will decrease at a slower rate at the end of the process. The reason is that the ice layer acts as if it is an insulator. As a result, the thicker the ice layer, the lower the cooling transfer rate from the ammonia to the liquid water.



**Figure 4** Variation of the cooling transfer rate with time

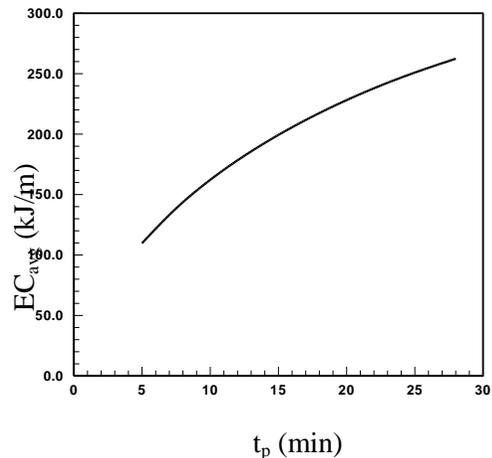
Figure 5 illustrates the relation of the SEC obtained from the numerical

prediction as a function of time. After the beginning of the process, the SEC first decreases, reaches a minimum point, and increases toward the end of process. It is explained that at the first part, the SEC decreases because the cooling transfer rate is still high leading to a high production rate. As time passes, the ice layer gets thicker and prevents the cooling effect from diffusing through the layer itself. Thus, the production rate gets lower, resulting in a higher value of SEC. The change of SEC over the entire production cycle is between 0.34 and 0.36 kJ/g. It should be noted that the lower the SEC, the more efficiently the energy is consumed. From Figure 5, if the ice-making process is terminated prior to the given production cycle time (i.e., 28 minutes), the value of  $SEC_{avg}$  should be lower. Thus, the relations between  $t_p$  and  $EC_{avg}$  and  $SEC_{avg}$  are investigated.



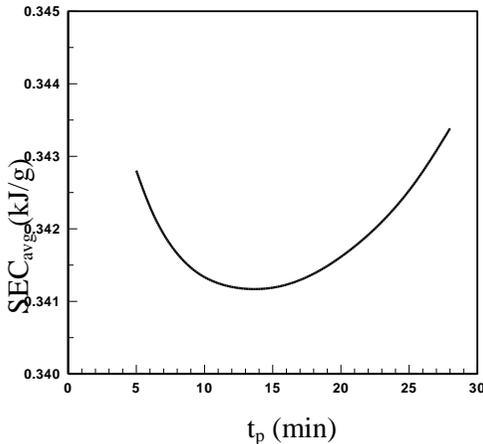
**Figure 5** Variation of the specific energy consumption with time

The variation of  $EC_{avg}$  with different  $t_p$  is depicted in Figure 6. As expected,  $EC_{avg}$  increases with increasing  $t_p$ . By setting the value of  $EC_{avg}$  at  $t_p = 5$  minutes as a reference case,  $EC_{avg}$  is increased by 48 and 108 percent when  $t_p$  increases from 5, to 10 and 20 minutes, respectively, with the average value of 5.8 kJ/m per minute.



**Figure 6** Effect of the production cycle time on the average energy consumption

The variation of  $SEC_{avg}$  with different  $t_p$  is depicted in Figure 7.  $SEC_{avg}$  decreases at the beginning when  $t_p$  is 5 minutes, reaches a minimum point, and increases until  $t_p$  reaches 28 minutes. Physically,  $SEC_{avg}$  represents the average energy consumption per mass of the produced ice over the entire cycle time. The lower the value of  $SEC_{avg}$ , the higher the energy efficiency of the process. According to Figure 7, the minimum point is located when  $t_p$  is approximately 14 minutes, with the minimum  $SEC_{avg}$  of 0.341 kJ/g. By comparing  $t_p$  at 14 minute with that at 28 minutes, the value of  $SEC_{avg}$  is decreased by 0.7 percent. In practice, the manufacturer tends to operate the ice-making system on the right side of the minimum point. Thus, it would be beneficial to the manufacturer to reduce the production cycle time from 28 minutes to 14 minutes, leading to a higher energy efficiency of the ice-making process. In the future, this numerical prediction can be adapted to different operating conditions, such as different tube diameters, refrigerant properties, or saturated refrigerant temperature to provide a proper production cycle time.



**Figure 7** Variation of the average specific energy consumption with the production cycle time

## 6. Conclusions

In this study, a mathematical model of the tube-ice making process has been formulated. The finite difference approach is employed to solve the governing system. The numerical solutions to predict the average energy consumption and the average specific energy consumption as a function of the production cycle time are obtained. A comparison between the numerical predictions and the field data is made, which results in qualitative agreement. As the production cycle time increases, the average energy consumption is increased by approximately 5.8 kJ/m per minute, whereas the average specific energy consumption reaches a minimum value of 0.341 kJ/g at the production cycle time of 14 minutes. Based on this study, it is recommended for ice-making manufacturers to reduce the production cycle time from 28 to 14 minutes. This results in a higher value of the average specific energy consumption by approximately 0.7 percent.

## 7. Nomenclature

$Bi$  Biot number  
 $C_h$  constant in equation (33) [W/m<sup>2</sup>-K]  
 $C_p$  specific heat at constant pressure [J/kg-K]

$C_{Bi}$  constant in equation (34)  
 $D$  tube thickness [m]  
 $EC$  energy consumption [kJ/m]  
 $h$  convective heat transfer coefficient [W/m<sup>2</sup>-K]  
 $\Delta H$  latent heat of fusion [J/kg]  
 $k$  thermal conductivity [W/m-K]  
 $J$  term for the finite difference equations  
 $K$  term for the finite difference equations  
 $L$  term for the finite difference equations  
 $r$  distance in the radial direction [m]  
 $r_i$  inner diameter of the tube [m]  
 $r_0$  outer diameter of the tube [m]  
 $\hat{r}$  dimensionless distance in the radial direction  
 $R$  radius of the ice layer [m]  
 $R_1$  wall-to-ice heat capacity ratio  
 $R_2$  wall-to-ice thermal conductivity ratio  
 $R_r$  ratio of the inner diameter to the tube thickness  
 $SEC$  specific energy consumption [kJ/g]  
 $Ste$  Stefan number  
 $t$  time [s, min]  
 $t_p$  production cycle time [min]  
 $\hat{t}$  dimensionless time  
 $T$  temperature [°C, K]

## Greek Symbols

$\alpha$  thermal diffusivity [m<sup>2</sup>/s]  
 $\phi$  dimensionless ice thickness  
 $\theta$  dimensionless temperature  
 $\rho$  density [kg/m<sup>3</sup>]

## Subscripts

0 saturation state  
 avg average  
 f freezing point  
 s ice region  
 w wall region

## 8. References

- [1] Necati Özisik, M., Heat Conduction, John-Wiley & Sons, New York, USA, pp. 416-430, 1993.

- [2] Cheung, F. B. and Cha, S. W., Finite-Difference Analysis of Growth and Decay of a Freeze Coat on a Continuous Moving Cylinder, *Numerical Heat Transfer*, Vol. 12, pp. 41-56, 1987.
- [3] Chapra, S. C. and Canale, R. P., *Numerical Methods for Engineers*, McGraw-Hill, New York, USA, pp. 738-741, 1990.
- [4] Tangthieng, C, Cheung, F. B. and Shih, Y. C., Solidified Layer Growth and Decay Characteristics During Freeze Coating of a Binary Substance, *Journal of Thermophysics and Heat Transfer*, Vol. 16, pp. 379-388, 2002.
- [5] Stephan, K. and Abdelsalam, M., Heat-Transfer Correlations for Natural Convection Boiling, *International Journal of Heat and Mass Transfer*, Vol. 23, pp. 73-87, 1980.
- [6] Carey, V. P., *Liquid-Vapor Phase-Change Phenomena: an Introduction to the Thermophysics of Vaporization and Condensation Processes in Heat Transfer Equipment*, Taylor & Francis, USA, pp. 630, 1992.
- [7] Piruchvet, N., and Tangthieng, C., A Numerical Study to Predict the Ice Thickness, the Ice Production Rate and the Cooling Load of the Block-Ice Making Process, the 18th Conference on Mechanical Engineering Network of Thailand, Article CST29, Khonkaen, Thailand, 18-20 October, 2005.
- [8] Incropera, F. P. and Dewitt, D. P., *Fundamentals of Heat and Mass Transfer*, John-Wiley & Sons, New York, USA, pp. 905-916, 2002.

# Study of Conversion Efficiency of Tube-type Solar Thermal Collectors

**Thananchai Leephakpreeda**

Sirindhorn International Institute of Technology, Thammasat University

P.O. BOX 22 Thammasat Rangsit Post Office,

Pathum Thani, Thailand, 12121

Email address: thanan@siit.tu.ac.th

## Abstract

There has been more and more attention to research on design and usage of solar thermal collectors so as to satisfy economic development. Nowadays, the extent of applicability of a tube-type solar thermal collector is an intriguing problem for new engineering design. In this paper, analytical and experimental studies of a multi-tube type solar thermal collector are presented as a tool. Thermodynamic analysis of the thermal collector is derived to describe and evaluate its performance in aspects of energy conversion efficiency. A small-scale rig of a multi-tube type solar thermal collector is made from materials available in market. It is experimentally tested during 9-hr of daytime in real use. It is found that experimental results give a good agreement with the mathematical model of conversion efficiency, which relates to material parameters of the solar thermal collector.

**Keywords:** solar thermal collector, conversion efficiency, tube-type collector

## 1. Introduction

Conceptually, a solar thermal collector is intended to absorb radiation from sunlight in order to provide usable heat. This class of solar heating system is widely used in many countries especially with high incidents of sunny climates like Thailand. There are various types of solar thermal collectors. Up to now, flat plate and box-type collectors have been used in household and industrial applications. In conventional designs, a black metal sheet with built in pipes is placed to face toward the sun and it is covered with a transparent plate such as glass, which allows solar energy to pass through, but reduces heat loss from the black metal sheet [1-2]. A heat-transport fluid such as water circulating through pipes is heated up and passed to a storage tank located above a

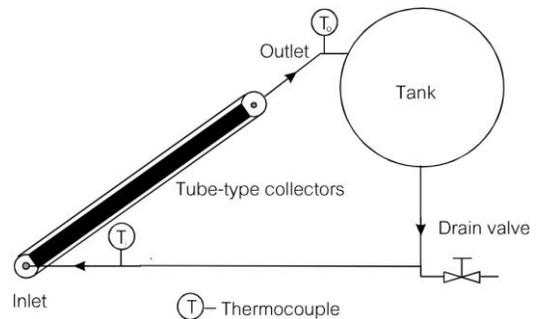
solar thermal collector. High efficiency of this type of collector can be obtained if the solar metal absorber directly faces toward the sun with right altitude angle and azimuth angle all the time [3]. Some improved solar thermal collectors are equipped with automatic sun tracking system [4]. The approaches have been not very attractive to practitioners owing to the high costs of instrumentation and maintenance.

To avoid those limitations, solar thermal collectors are currently built with multiple transparent tubes, which heat up solar absorbers that are made of black-coated metal pipes inside. An advantage of this design over flat-plate type collectors is that the round profiles of the tubes are always perpendicular to the sun's rays and therefore the solar radiation absorbed is approximately constant over the course of a

day. This allows them to reach considerably higher temperature [5]. In turn, the tube-type collectors have drawn more and more researchers' attention to obtain the most effective use [6]. In [7], efficiency characteristics of tube-type solar thermal collectors are proposed to investigate performance by using supercritical CO<sub>2</sub> as a working fluid in test studies. Experimental results report an interesting fact that there are variations of the collector efficiency against the ratio of temperature difference between inlet fluid temperature and ambient air temperature, to the solar radiation, under different seasons during the year. Unfortunately, analytical explanations have not explicitly been given in their works or even in other open literature [8]. Understanding this relation might be a significant contribution to the design of the solar thermal collectors. In this study, such a characteristic model of the collector efficiency is determined with the principle of energy balance in section 3. The details of the experimental setup of a multi-tube type solar thermal collector are discussed in section 2. Section 4 yields the experimental verification of the proposed characteristic model. The conclusion is given in section 5.

## 2. Experimental setup

The schematic diagram of a multi-tube type solar thermal collector is tested in closed circuit under outdoor conditions as shown in Fig. 1.



**Fig. 1** Schematic diagram of tube-type solar thermal collector.

The upper 50-litre insulated storage tank is used for hot water storage. The water is circulated by a thermal siphon effect through the solar thermal collector by heating up the cool water heading into the storage tank. With a conventional inclined angle of  $\sim 10^\circ$ , the solar thermal collector consists of ten double tubular pipes connected in sequence; the outer tubes are made of transparent UV-resistance acrylic with 20-mm outside diameter and 2-mm thickness while the inner tube is made of black rubber pipe with 15-mm outside diameter and 3-mm thickness. They are selected for low cost and easy assembly. Fig. 2 shows the testing panel of the 1-m long solar thermal collectors. The water temperature at inlet and outlet as well as the ambient temperature and the water temperature in the storage tank are measured by type-K thermocouples. The incident solar radiation is measured with a pyranometer. All measurements from thermocouples and pyranometer are taken by a PC-based data acquisition system with 5-minute sampling time during experiments.



**Fig. 2** Testing panel of solar thermal collector.

### 3. Analysis on conversion efficiency

From upstream to downstream of flow in the solar thermal collector, the useful heat collected along the pipe per unit time can be expressed by:

$$Q = \rho \dot{V} C_p (T_o - T_i) \quad (1)$$

where  $\rho$  is the density of the working fluid ( $\text{kg/m}^3$ ),  $\dot{V}$  is the volume flow rate ( $\text{m}^3/\text{s}$ ),  $C_p$  is the constant pressure specific heat ( $\text{J}/(\text{kgK})$ ),  $T_o$  is the temperature at outlet (K), and  $T_i$  is the temperature at inlet (K).

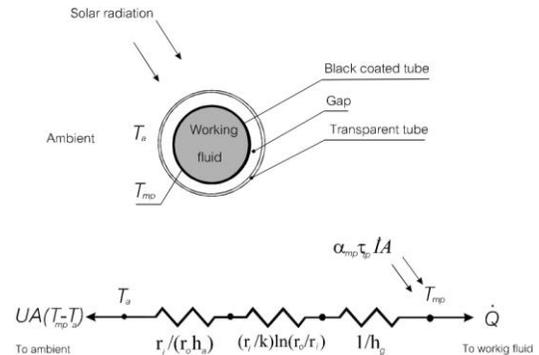
The conversion efficiency expresses the fraction of incident heat that is collected by the working fluid to the incident solar radiation [7].

$$\eta = \frac{Q}{IA} \quad (2)$$

where  $I$  is the solar radiation on collector tilted surface ( $\text{W}/\text{m}^2$ ), and  $A$  is the tilted surface area of the collector ( $\text{m}^2$ ).

It should be noticed from Eq. (2) that there is no explicit description on how the parameters of the solar thermal collector influence the conversion efficiency. By applying the principle of energy balance to the solar thermal collector, one portion of incident radiation from the outer transparent pipe that is transmitted to the black coated pipe, is absorbed to heat up the working fluid. Another portion is lost by the heat transfer back to ambient. The heat transfer

in radial direction is considered for simplicity. It should be noted that the heat conduction in the circumferential direction can be neglected owing to a thin-walled tube. The working fluid is lumped in this analysis, e.g., the mean temperature of the working fluid. The convective heat transfer between the thin-wall tube and the working fluid is neglected since the flow rate of the working fluid is significantly low when it is caused by the thermal siphon effect. Fig. 3 illustrates the thermal circuit of the heat transfer in the tube-type solar thermal collector.



**Fig. 3** Thermal circuit of tube-type solar thermal collector.

Hence, the useful heat collected along the pipe can be governed as well by:

$$Q = \alpha_{mp} (\tau_{tp} IA) - UA(T_{mp} - T_a) \quad (3)$$

where  $\alpha_{mp}$  is the absorption factor of the inner black-coated pipe,  $\tau_{tp}$  is the transmission factor of the outer transparent pipe,  $U$  is the overall heat transfer coefficient ( $\text{W}/(\text{m}^2\text{K})$ ),  $T_{mp}$  is the mean surface temperature of the inner black-coated pipe (K), and  $T_a$  is the ambient temperature (K). The overall heat transfer coefficient can be related to material properties as:

$$U = \frac{1}{r_i/r_o h_a + (r_i/k) \ln(r_o/r_i) + 1/h_g} \quad (4)$$

where  $h_a$  is the convection heat transfer coefficient of the outside air ( $\text{W/m}^2$ ),  $k$  is the thermal conductivity of the transparent pipe ( $\text{W/m}$ ),  $r_o$  and  $r_i$  are respectively, the outer and inner radii of the transparent pipe (m), and  $h_g$  is the convection heat transfer coefficient between gap ( $\text{W/m}^2$ ). Now, the conversion efficiency in Eq. (2) can be expressed in terms of parameters of the solar thermal collector as:

$$\eta = \alpha_{mp} \tau_{tp} - U \frac{(T_{mp} - T_a)}{I} \quad (5)$$

In this work, the derivation of Eq. (5) yields contributive enlightenment in design of the solar thermal collector. It can be seen that the optimal design parameters of the solar thermal collector can be obtained by selecting transparent pipe with high radiation transmission and black-coated pipe with high radiation absorption and heat conduction, while the gap should provide high thermal resistance. In this high-level technique, one may use an evacuated gap for insulation of conducted heat loss. Besides, cooling from ambient degrades the conversion efficiency. The high ambient temperature and radiation in daytime causes the tendency of a higher conversion efficiency.

For verification of the characteristic model in Eq. (5) through experimental results, it is not practical to measure directly the temperature of the black-coated pipe. The modification by a factor is proposed in this study to make it more convenient to present the conversion efficiency in terms of a more accessible variable of the working fluid. The factor, called collector efficiency factor  $F$ , is the ratio of the actual useful heat collected along pipe in Eq. (3) to the equivalent useful heat, where the heat loss is determined by the temperature difference

between the mean temperature of the working fluid and the ambient temperature.

$$F = \frac{Q}{\alpha_{mp} \tau_{tp} IA - UA(T_{mf} - T_a)} \quad (6)$$

where  $T_{mf}$  is the mean temperature of the working fluid.

$$T_{mf} = \frac{T_o + T_i}{2} \quad (7)$$

It should be noted that the value of the collector efficiency factor is less than unity.

Applying the definition of the collector efficiency factor in Eq. (6), the conversion efficiency in Eq. (5) can be written as:

$$\eta = F \alpha_{mp} \tau_{tp} - FU \frac{(T_{mf} - T_a)}{I} \quad (8)$$

Now, it can be seen that the conversion efficiency is proportional to the ratio of the temperature difference between the mean temperature of the working fluid and the ambient temperature, to the solar radiation. Experimental results on solar collector testing are to verify the relation of the conversion efficiency in the following section.

## 4. Results and Discussion

The experiment is run so as to determine the conversion efficiency of the tube-type solar thermal collector. Including the ambient temperature and the incident solar radiation, the water temperatures at inlet and at outlet of the solar thermal collector and the water temperature in the storage tank are measured every 5 minutes, since the thermal response is quite slow in nature. All the measurements are from 8:30 a.m. to 5:30 p.m. Fig. 4(a)-4(e) illustrates the dotted lines of the water temperature at the inlet and at the outlet of the solar thermal collector and the water temperature in storage tank  $T_w$ , as well as the ambient temperature and the incident solar radiation.

The solid lines are determined by the least-squared-error method with polynomial interpolation for attaining the best possible estimation of trends. It is observed that all the temperatures, including the water temperature in the storage tank, increase while the incident solar radiation increases. This can be also interpreted that the thermal siphon circulates the hot water from the thermal collector filling up to the storage tank in a loop. When the solar radiation decreases at certain times, the water temperature at the outlet and the air temperature decrease accordingly, while the water temperature at the inlet and the water temperature in the storage tank remain steady, due to heat trapped in the storage tank by insulation. It should be noticed that the highest value of the water temperature is about 38 °C, which is feasible in various medium-temperature applications.

In the calculation of conversion efficiency, the density and the pressure specific heat of the water are 1000 kg/m<sup>3</sup> and 4.186 kJ/(kg.K), respectively. The useful heat collected along the black rubber pipe is to be determined in Eq. (1). However, it should be noted that it is not practical to measure a low flow rate of the water in the thermal collector for calculating the useful heat in Eq. (1). Intuitively, the useful heat collected along the black rubber pipe causes an increase of the internal energy and, in turn, the temperature of the water within the insulated tank increases. Therefore, the difference between the incoming enthalpy of the water and the outgoing enthalpy of the water in Eq. (1) is equal to the rate of change in the internal energy of the water contained within the insulated tank. For calculation of the useful heat, this yields:

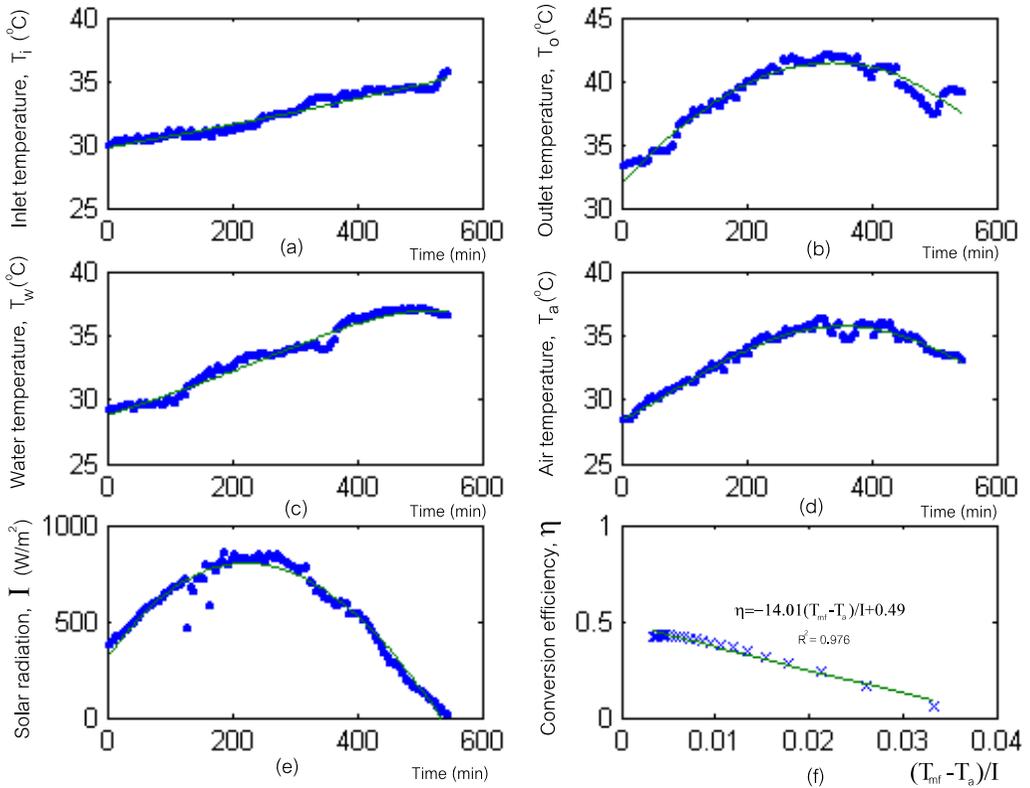
$$Q = \rho \dot{V} C_p (T_o - T_i) = \rho V C_p \dot{T}_w \quad (9)$$

where  $\dot{T}_w$  is the rate of temperature change of the water in the insulated tank and  $V$  is the volume of the water contained in the insulated tank.

Now, the derivatives of the water temperature with respect to time can be simply determined by a numerical difference approximation from the plots of the water temperature within the insulated tank against time from Fig. 4(c). It is straightforward to determine the useful heat from Eq. (9) instead of Eq. (1). From Eq. (2), the instantaneous conversion efficiency is calculated. Fig. 4(f) shows the conversion efficiency of the solar thermal collector designed in section 2 against the corresponding ratio of the temperature difference between the mean temperature of the working fluid and the ambient temperature, to the solar radiation. The ratios are obtained from the variation of the sampled data such as the mean temperature of the working fluid  $T_{mf}$ , the incident solar radiation of the sunlight  $I$ , and the ambient temperature  $T_a$ , during days of experiments. As the plots of the results, the characteristic values of the conversion efficiency, that is  $(F\alpha_{mp}\tau_p)$  and  $(FU)$  in Eq. (8), are obtained to be 0.49 (< 1) as expected, and -14.01, respectively. Experimental results in Fig. 4(f) are fitted well, in agreement of the characteristic model. With the same procedure, the experiments on the following days are performed to obtain the linear characteristics of conversion efficiency of the tube-type solar thermal collectors. It is observed that the experimental results have a similar tendency of all measurements in Fig. 4 on the first day. Table 1 shows the compared results of best fits to a linear regression in all three successive days of experiments. It can be deduced from Fig. 4(f) that the maximum conversion efficiency takes place when the ratio of the temperature difference between the mean temperature of the working fluid and the ambient temperature, to the incident solar radiation approaches zero. Hence, a high value of the conversion efficiency can be

obtained when the incident solar radiation is high ( $>1000 \text{ W/m}^2$ ), during working daytime for a given temperature difference

between the mean temperature of the working fluid and the ambient temperature.



**Fig. 4** Plots of temperatures, solar radiation and conversion efficiency.

**Table 1.** Results of linear regression in three-day experiments.

	<b>Slope</b>	<b>Intercept</b>
1 <sup>st</sup> day	-14.01	0.49
2 <sup>nd</sup> day	-15.05	0.54
3 <sup>rd</sup> day	-14.34	0.62
Average value	-14.47	0.55

## 5. Conclusion

Basically, the main contribution of this work is to present the thermal analysis on conversion efficiency of the tube-type solar thermal collector for engineering design. Furthermore, the solar thermal performance of a real low-cost tube-type collector is evaluated experimentally. The model of the conversion efficiency yields good agreement with experimental results. The results indicate the conversion efficiency is proportional to the ratio of the temperature difference between the mean temperature of the working fluid and the ambient temperature, to the solar radiation. It can be confirmed that material parameters of the solar thermal collectors are coefficient factors of the conversion efficiency. Therefore, in the optimal design on those parameters of the solar thermal collector, a transparent pipe might be selected with high radiation transmission and a black-coated metal pipe can be chosen with high radiation absorption and heat conduction as well as a gap should provide high thermal resistance. It should be noted that other types of solar thermal collectors can be analyzed for parametric characteristics on conversion efficiency by generalizing the concepts of this proposed experimental study for those related works.

## 6. Acknowledge

The author thanks Nunt Chaiyakarn and Thanawat Amornjaruchit for assistance on the experiments.

## 7. References

- [1] Ho, C.D. and Che, T.C., Collector Efficiency Improvement of Recyclic Double-pass Sheet-and-tube Solar Water Heaters with Internal Fins Attached, *Renewable Energy*, Vol. 33, No. 4, pp. 655-664, 2008.
- [2] Ersoy, H.K., Yalcin, S., Yapici, R. and Ozgoren, M., Performance Analysis of New-design Solar Air Collectors for Drying Applications, *Renewable Energy*, Vol. 32, No. 10, pp.1645-1660, 2007.
- [3] Grena, R., An Algorithm for the Computation of the Solar Position, *Solar Energy*, Vol. 82, No. 5, pp. 462-470, 2008.
- [4] Rubio, F.R., Ortega, M.G., Gordillo, F. and, Lopez-Martinez, M., Application of New Control Strategy for Sun Tracking, *Energy Conversion and Management*, Vol. 48, No. 7, pp. 2174-2184, 2007.
- [5] Morrison, G.L., Budihardjo, I., Behnia, M., Water-in-glass Evacuated Tube Solar Water Heaters, *Solar Energy*, Vol. 76, No.1-3, pp. 135-140, 2004.
- [6] Shah, L.J. and Furbo, S., Theoretical Flow Investigations of an All Glass Evacuated Tubular Collector, *Solar Energy*, Vol. 81, No. 6, pp. 822-828, 2007.
- [7] Zhang, X.R. and Yamaguchi, H., An Experimental Study on Evacuated Tube Solar Collector Using Supercritical CO<sub>2</sub>, *Applied Thermal Engineering*, Vol. 28, No. 10, pp. 1225-1233, 2008.
- [8] Charalambous, P.G., Maidment, G.G., Kalogirou, S.A. and Yiakoumetti, K., Photovoltaic Thermal (PV/T) Collectors: A Review, *Applied Thermal Engineering*, Vol. 27, No. 2-3, pp. 275-286, 2007.

# Stand-alone Photovoltaic-Based Reduction of Harmonic Current Using a 12-Pulse AC/DC Power Conversion System with Trapezoidal Voltage Waveform

**Narin Watanakul**

Department of Electrical & Computer Engineering, Thammasat University,  
Pathum Thani, Thailand.

**Singthong Pattanasethanon and Apinan Aurasopon**

Faculty of Engineering, Mahasarakham University, Mahasarakham, Thailand

## Abstract

This paper presents an application of stand-alone Photovoltaic (PV) systems for improving the total harmonic distortion (THD) of line input current in twelve-pulse AC/DC power conversion systems. In these control systems, the trapezoidal voltage is generated from a PV stand-alone and injected into line input voltage. The process control system uses a 12-pulse diode rectifier, using a conventional 3-phase bridge 6-pulse diode rectifier. The results are from experiments and simulation of the trapezoidal voltage waveform. The results of the process can reduce the total Harmonic current (THD) of line input current. The simulation results by MATLAB program are used as a guideline for analyzing and designing the parameters of a control system. The experimental results show, that the performance of the proposed control system in the case of THD of line input current and electrical system quality, are improved.

**Keyword:** Stand-alone Photovoltaic, Trapezoidal voltage waveform, 12-pulse AC/DC power conversion , power quality, harmonic current.

## 1. Introduction

Frequently found problems in the distribution system involve harmonics current. (IEEE 1159-1995) [1]. AI-Mathnani et al.[2] presented a simulation model of a twelve-pulse, parallel two-level three phase inverter, Dynamic Voltage Restore (DVR) using photovoltaics as a means of providing an alternative energy source for the DVR [2]. Simulations were carried out using the

PSCAD/EMTDC. Simulation results proved the capability of the photovoltaic-based DVR in mitigating voltage sag in a distribution system [2]. Sinusoidal pulse-width-modulation (SPWM) inverters are frequently found in variable-speed ac motor drives. Several modulation strategies have been developed. One of the most widely used, due to its simplicity, is the sinusoidal modulation [3]. Having many reported cases, harmonic current is found in a lot of

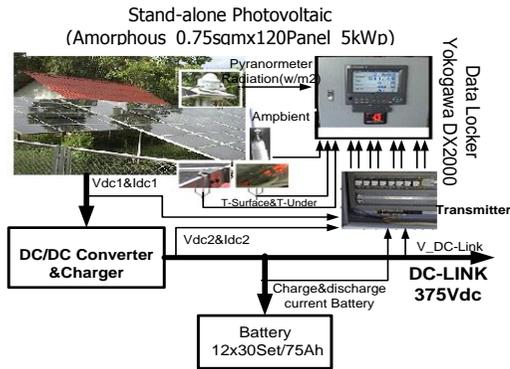
industrial factories. This indicates problems in the power quality of electricity used. Power loss in electric generators, power loss in transmission line systems and power loss in transformers, as well as electromagnetic interference (EMI) are produced as a result of harmonic current arising from installing devices such as power converters, power inverters, electroplating, welding machines, and induction furnaces. These devices normally contain power electronics such as power diode-rectifiers used to convert alternating current (AC) to direct current (DC). Theoretical input current harmonics for rectifier circuits are a function of pulse number and can be expressed as  $(h = (NP + 1))$  where  $N = 1, 2, 3, \dots$ . The isolation transformer with a delta primary, a Delta connected secondary, and a Wye secondary are connected to obtain the necessary phase shift. For a six-pulse rectifier, the input current will have harmonic components at the following multiples of the fundamental frequency 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, etc. For the twelve-pulse system shown in Fig.1, the input current will have theoretical harmonic components at the following multiples of the fundamental frequency: 11, 13, 23, 25, 35, 37, etc. The harmonic order of the 5th and 7th harmonics are absent in the 12-pulse system. Since the magnitude of each harmonic is proportional to the reciprocal of the harmonic number, the twelve-pulse system has a lower theoretical harmonic current distortion. The number of uses of such devices increases according to factories' technological advancements. Concerning harmonic current problems in Thailand, the electricity committee has considered power quality as an important part of an electricity system. Electric Generating Authority of Thailand (EGAT), Metropolitan Electricity Authority (MEA) and Provincial Electricity Authority (PEA) have installed inductors and capacitors as detained filters and it is found that harmonic current is not totally taken out from the

system. One alternative method to reduce harmonic current that could perform better than detained filters is to improve 6-pulse converters to be 12-pulse AC/DC power conversion systems. Twelve-pulse drives have provided a simpler and more cost effective approach to achieve higher current ratings than direct paralleling of power semiconductors. The authors propose an application by Stand-alone Photovoltaic-Based reduction of harmonic current using 12-Pulse AC/DC power conversion System with trapezoidal voltage wave form. For the PV Stand-alone system, one example of a station is at Baan Pukem School in Phetchaburi. The Measurement Diagram for PV stand-alone is shown in Fig.1 and installation site is shown in Fig.5 (a), (b). The system control unit is PWM battery charger DC/DC converter and power inverter unit.

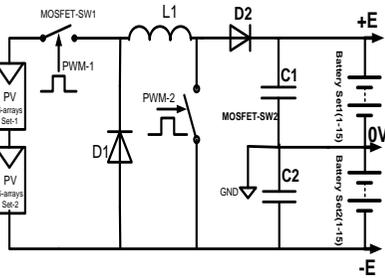
## 2. Stand-alone Photovoltaic and converter

A PV Stand-alone system has been installed at Baan Phukhem School, Phetchaburi province in the central region of Thailand. Fig.1 exhibits PV stand-alone [5] consisting of panels of amorphous silicon PV Panels. A total of 120 panels, 0.75square meter each with 42.11W/44.6/0.97A, are linked into 12 arrays providing 5 kW<sub>p</sub>/375Vdc. The charger controller is a dc-dc buck-boost converter. A total of 30 sets of sealed lead acid 12V/75Ah battery model UB12750 manufactured by Universal Power Group are installed. A Yokogawa recorder DX2000 has been used to continuously record data every 4 minutes. This supports various measurements such as irradiation ( $W/m^2$ ), temperature(c), dc-voltage (DC-link), dc-current (battery charge & discharge current) and energy consumption (kWh). Irradiation is detected with a Kipp&Zonen CMP11 pyranometer, with high accuracy. For the DC/DC converter

and battery charger (see Fig.1 (b)), details of non-inverting buck-boost converter can be found in [6], [7]. The converter used in this study contains battery (DC-Link) stabilized at 375Vdc with an average voltage: (+E=187.50Vdc) and (-E=187.50 Vdc).



(a)



(b)

**Fig. 1** PV stand-alone and PWM battery charger DC/DC converter (a) Measurement Diagram for PV stand-alone installed at Baan Phukhem School, Phetchaburi province.(b) Circuit diagram of Buck-boost converter for battery charger.

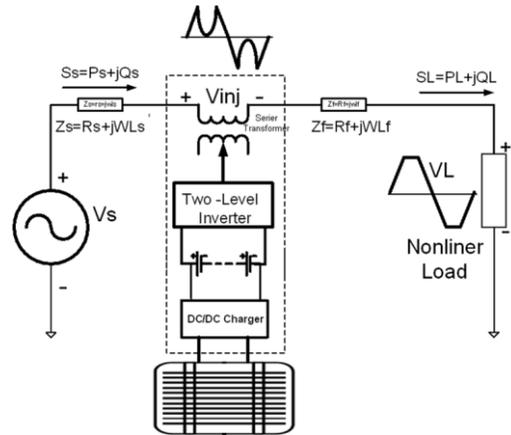
### 3. Technical design of Power supply Trapezoidal voltage waveform

The Signal-mixing technique between sine-waveform and trapezoidal waveform is performed and the result is a periodic waveform. However, when the utility power system recovers from an interruption, the command to the inverter is replaced with wave-shaping voltages to be a trapezoidal waveform. Equation (1) is

followed and corresponds to the block diagram in Fig.2.

$$f(\text{tra}(\omega t)) = f(\sin(\omega t)) - f(\text{Periodic}(\omega t)) \quad (1)$$

$$S_L(\text{VA}) = (P_{inj} + P_s) + j(Q_{inj} + Q_s) \quad (2)$$



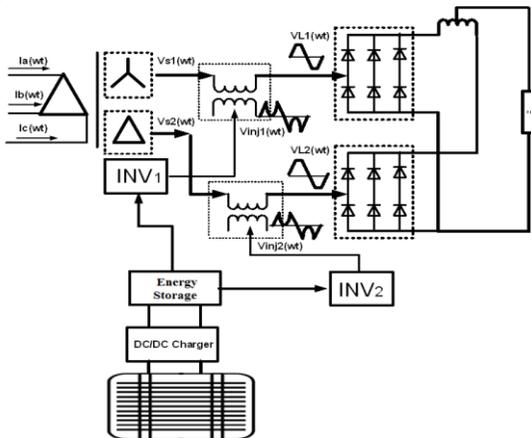
**Fig. 2** Signal mixing of the block diagram power supply trapezoidal voltage

When Equation. (2) is followed, the terms  $S_{inj}$ ,  $P_{inj}$ ,  $Q_{inj}$  are apparent, active, and reactive power compensation, respectively. When  $P_{inj}$ ,  $Q_{inj}$  are applied to series transformer,  $P_s$  and  $Q_s$  are obtained from the active and reactive power supply. Merging of power transmission  $P_{inj}$ ,  $Q_{inj}$ ,  $P_s$ , and  $Q_s$  supplied according to load requirements creates the total apparent power supply to the load  $S_L(\text{VA})$ .

### 4. Twelve -Pulse AC/DC power conversion system applied trapezoidal voltage

The ac supply is from a transformer having two secondaries, Wye-connected and Delta-connected. In this manner, the three-phase voltage supplying the two bridges is displaced by a phase angle of 30 degrees. Hence the two six-pulse output are

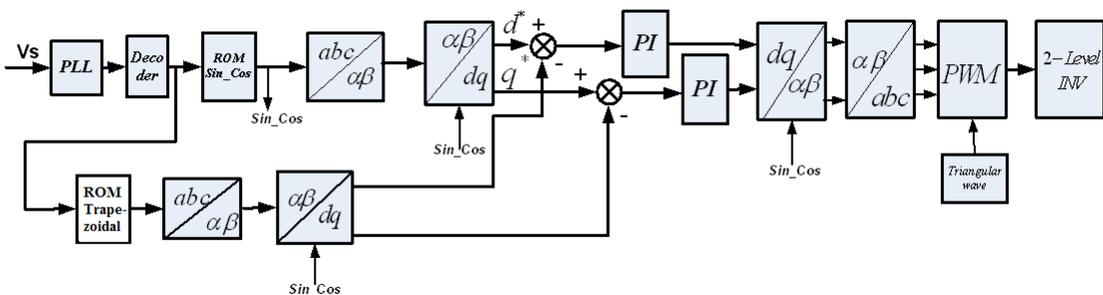
symmetrically displaced to give an overall twelve-pulse output [4]. A twelve-pulse system can also be constructed from two six-pulse rectifiers connected in parallel. It has been known that total harmonic current distortion from twelve-pulse converters is better than that from six-pulse converters. When a trapezoidal voltage is applied, total harmonic current distortion is further improved as shown in Fig.3. It consists of a combination of a double-series diode rectifier rated at 5kVA and SPWM inverter with a peak voltage and current rating of 5kVA.



**Fig.3** Twelve-pulse AC/DC power conversion system applied trapezoidal voltage waveform

### 5. Modeling of trapezoidal voltage waveform within MATLAB

Modeling of trapezoidal voltage source is proposed as a control strategy for reduction of harmonic current by applying a trapezoidal voltage source. The modeling blocks diagram consists of phase lock loop (PLL), decoder, sine-cosine-ROM and trapezoidal -ROM. The sine-cosine-ROM is vital for in-phase strategy. The applied signal is transformed from two phase coordinates stationary reference frame ( $\alpha\beta$ ) into two phase coordinates, rotating synchronously with a flux coordinate (dq command) as given in Equation. (1). For the other path, the dq censoring signal obtained from abc three phase coordinates and  $\alpha\beta$  is expressed in Equation. (3) and (4). Difference of signal from both paths is detected for a periodic voltage signal. The Output signal from the process then passes to PI-controller. Thereafter, the received signal in dq coordinate is transformed into abc coordinate as written in Equation. (7) and (8). Two-level inverter is further applied to the process.



**Fig.4** The proposed control strategy for reduction of harmonic current by applying trapezoidal voltage waveform.

$$Va(t)+Vb(t)+Vc(t)=0 \quad (3)$$

$$\begin{bmatrix} V\alpha(t) \\ V\beta(t) \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & \cos \frac{2\pi}{3} & \cos \frac{4\pi}{3} \\ 0 & \sin \frac{2\pi}{3} & \sin \frac{4\pi}{3} \end{bmatrix} \begin{bmatrix} Va(t) \\ Vb(t) \\ Vc(t) \end{bmatrix} \quad (4)$$

$$V(t) = V\alpha(t) + jV\beta(t) \quad (5)$$

$$V(t) = \sqrt{\frac{2}{3}} \left[ Va(t) + Vb(t)e^{j\frac{2\pi}{3}} + Vc(t)e^{j\frac{4\pi}{3}} \right] \quad (6)$$

$$\begin{bmatrix} Vd(t) \\ Vq(t) \end{bmatrix} = \begin{bmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{bmatrix} \begin{bmatrix} V\alpha(t) \\ V\beta(t) \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} Va(t) \\ Vb(t) \\ Vc(t) \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & 0 \\ \cos \frac{2\pi}{3} & \sin \frac{2\pi}{3} \\ \cos \frac{4\pi}{3} & \sin \frac{4\pi}{3} \end{bmatrix} \begin{bmatrix} V\alpha(t) \\ V\beta(t) \end{bmatrix} \quad (8)$$

**Table. 1** Main system parameters for the experimental & simulation trapezoidal voltage waveform.

Description	Rated-value	Per unit-value
Load rating	5kVA	1.0 pu
PV stand-alone	5kWp	1.0 pu
Load voltage	230V	1.0 pu
Input voltage	230V	1.0 pu
Trapezoidal	230V	1.0 pu
DC-link	375V	1.63 pu
Power-frequency	50Hz	1.0 pu
SW-frequency(SPWM)	10kHz	200 pu

Table 1. Describes main system parameters for the experimental and simulation trapezoidal voltage source setup for voltage supply  $V_s(V)$  equal to 1.0 pu and injected voltage  $V_{inj}(V)$  equal to 1.0 pu having battery energy storage rated 375V/75Ah, and containing load rating of 5 kVA.

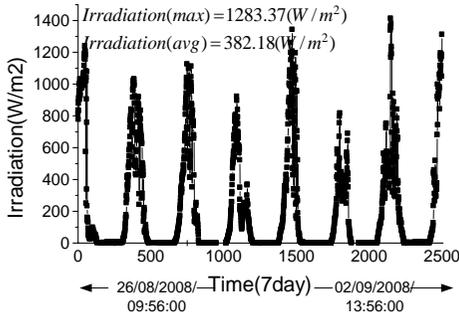
## 4. Study Results

### 4.1 Study results of stand-alone PV system

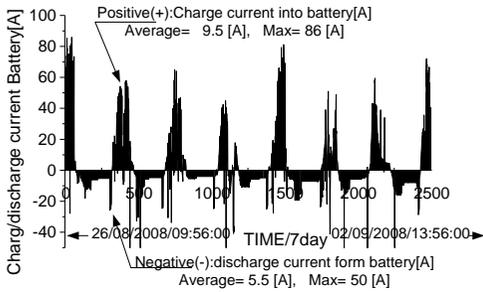
According to the test of PV stand-alone installed at Baan Pukem school (Fig.5 (a),(b)) the work performed on the collected and gathered data IEA PVPS task 2 [8], from seven consecutive days data is comprised of irradiation, average temperature of a solar panel consisting of T-under and T-surface, and ambient temperature [9]. A Yokogawa recorder DX2000 has been used to continuously record average data every four minutes with high accuracy. Irradiation ( $W/m^2$ ) is detected with Kipp&Zonen CMP11 pyranometer, with high accuracy. Consumption level of battery controlled by buck-boost converter has charge battery average current of 9.5A and discharge battery average current of 5.5A as in Fig.5 (d), producing output voltage of battery (DC-Link) stabilized at 375Vdc with the average, voltage (+E=187.50Vdc) and (E=187.50Vdc). Value of some parameters are as follows: Inductor  $L1=75\mu H$ , capacitor:  $C1=C2=470\mu F$ , MOSFET SW(360/750V), Diode: D1, D2 (60A/750V), duty cycle (D=0.30-0.95) and switching frequency 75 kHz as in Fig.1 (b) and result given in Fig.5 (c)-Fig.5 (e). The Average of measured irradiation is  $382 W/m^2$ . Ambient temperature is maximum at  $43.20^\circ C$  with an average of  $28.01^\circ C$ . Module temperature from T-under and T-surface is maximum at  $65.65^\circ C$  with an average of  $31.66^\circ C$ , Output power of PV-array is maximum 5.6 kWp and average 1.2 kW/day, Potential-energy 247.54 kWh.d. The efficiency of the PV-array is maximum of 19% with an average of 5.45% as in Fig.5 (f).



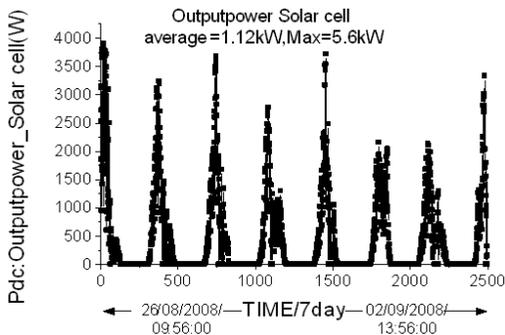
(a) (b)



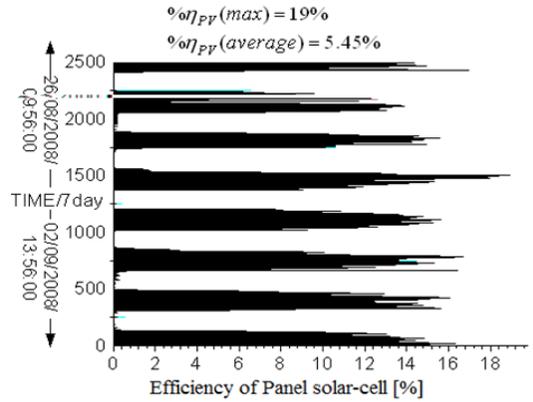
(c)



(d)



(e)

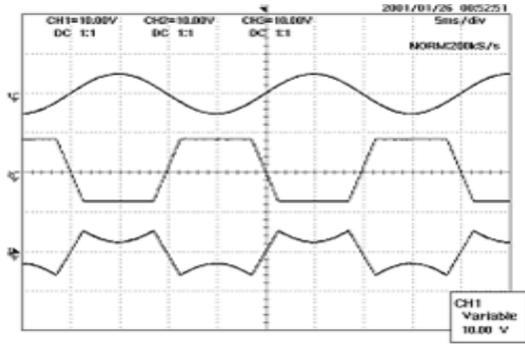


(f)

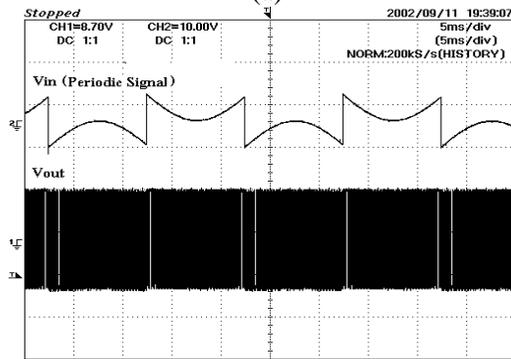
**Fig.5** PV Stand-alone system consumption level (a) installation site (b) Data logger Inverter and charger controller (c) irradiation data (d) battery charge & discharge current data (e) Output power of PV consumption (f) efficiency of PV-array system

## 4.2 Testing SPWM inverter having trapezoidal voltage

Trapezoidal voltage waveform is generated by power inverter circuit system that contains sine-waveform signal sensor, and a periodic voltage waveform, for the phase lock loop (PLL). Under normal state, PLL is employed to synchronize the wave-shaping voltage with the utility sinusoidal voltage. With freely run PLL, the output is a 50Hz trapezoidal voltage waveform. Notably, SPWM is then connected to power drive circuit, via power IGBT switch frequency which is equal to 10 kHz (200pu). Equations. (1)-(8) are followed and corresponds to the block diagram in Fig.4, Fig.6 and Fig.7 show experimental waveform, sine-waveform, trapezoidal-waveform, and periodic-waveform.

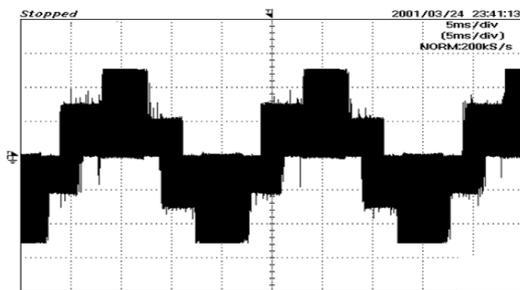


(a)

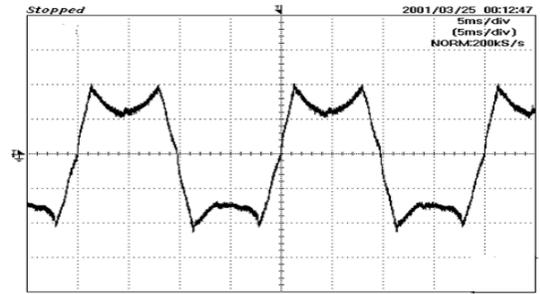


(b)

**Fig.6** (a) Experimental waveform Input-Output signal mixing (b) output signal SPWM when having periodic waveform input



(a)

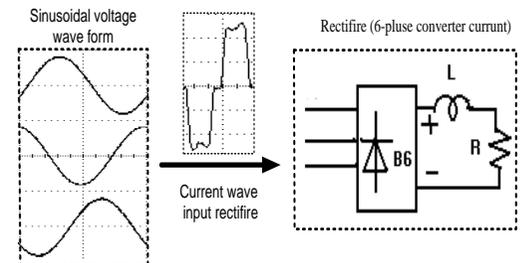


(b)

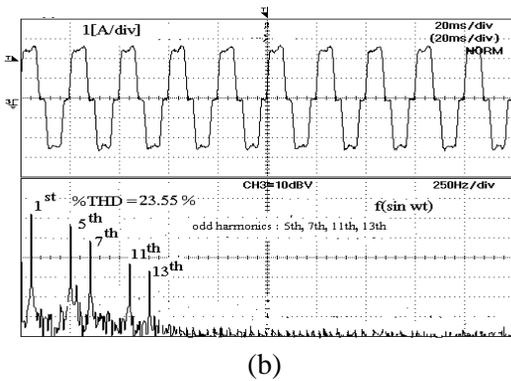
**Fig.7** Experimental waveform of Periodic voltage waveform (a) before and (b) after filtered

**4.3 Comparing harmonic current for Six-Pulse AC/DC power conversion**

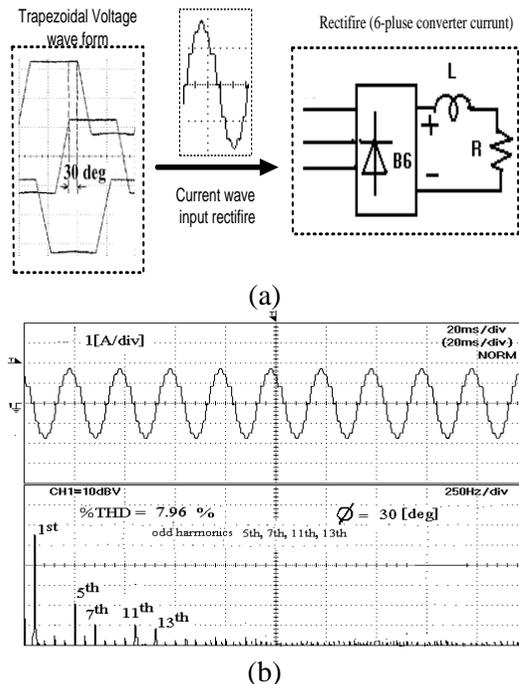
For experimental testing under the same load conditions, a comparison can therefore be made between using sinusoidal wave voltage source (Fig.8) and trapezoidal voltage source (Fig.9). From trial experiments, it is found that the best trapezoidal voltage source should have 1 ms rising time and 1 ms falling time with 30 degrees delay offset between phase-to-phase as shown in Fig.9 (a). The main reason for such conditions is that the input 6-pulse converter current will be nearly a sinusoidal waveform. Further, this causes total harmonic current distortion (%THDi) to be reduced by 15.6% (compare Fig.8 and Fig.9).



(a)



**Fig. 8** Experimental (a) Input current when using sine wave voltage source and (b) Spectrum that generates harmonic current when using sinusoidal voltage waveform

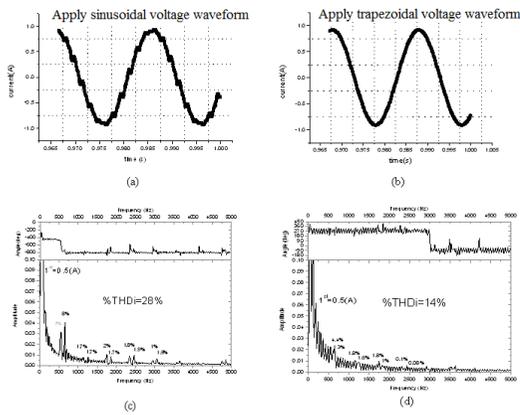


**Fig.9** Experimental (a) Input current when using trapezoidal voltage source and (b) Spectrum that generates harmonic current when using trapezoidal voltage waveform.

**4.4 Twelve-Pulse AC/DC power conversion system applied trapezoidal voltage source with MATLAB Simulation**

The ac supply is from a transformer having two secondaries, Wye-connected

and Delta-connected. In this manner, the three-phase voltage supplying the two bridges is displaced by a phase angle of 30 degrees. Hence, the two six-pulse output are symmetrically displaced to give an overall twelve-pulse output. It has been known that total harmonic current distortion from a twelve-pulse converter is better than that from a six-pulse converter. When trapezoidal voltage is applied, total harmonic current distortion is further improved as shown in Fig.3. It consists of a combination of a double-series diode rectifier rated at 5kVA. A clear advantage is from using DC link from PV stand-alone. The ac terminal of each sinusoidal SPWM inverter is connected in series with a power line of  $\Delta/\Delta/Y$  transformer rated 5kVA,  $\Delta/\Delta/Y$ , 380/220/220V,50Hz. This case study uses computer program MATLAB simulation, modeling of trapezoidal voltage waveform within MATLAB (Fig.3 and Fig.4). From this simulation, relevant parameters are given in the transformations, parameter of three winding power transformer. [ $\Delta$ /Primary ( $R1=0.0025pu$ , ( $L1=0.08pu$ ), secondary winding [ $\Delta$ /Secondary ( $R2=0.0025pu$ , ( $L2=0.08pu$ )] and [ $Y$ /Secondary ( $R3=0.0025pu$ , ( $L3=0.08pu$ )], DC-Link (375Vdc), two-level inverter using IGBT-devices of Fall-time  $T_f$  ( $1\mu s$ ) and tail time  $T_t$  ( $2\mu s$ ), frequency-SPWM ( $fsw=10kHz$ ) and LC-Filter ( $L=300\mu H$ ),( $C=20\mu F$ ). A comparison study, conducted by comparing results obtained from a conventional 12-pulse power converter, is nearly a sinusoidal waveform when applied trapezoidal voltage. From Fig.10, it is learnt that input current primary transformer is nearly a sinusoidal waveform (Fig.10 (a), against Fig.10 (b)). A comparison can easily be seen from the figures. For this simulation, total harmonic current distortion (THDi) can be reduced by 14 percent (comparing Fig. 10 (c) and Fig. 10 (d)) at harmonic spectrum in order 11<sup>th</sup>, 13<sup>th</sup>, 23<sup>th</sup>, 25<sup>th</sup>, 35<sup>th</sup>, 37<sup>th</sup>, 47<sup>th</sup> and 49<sup>th</sup>.



**Fig. 10** Simulation result waveform (a) before starting current into primary winding, when applying sinusoidal voltage waveform (b) after starting current into primary winding, when applying trapezoidal voltage waveform (c) Spectrum that generates harmonic current when using sinusoidal waveform voltage (d) Spectrum that generates harmonic current when using trapezoidal voltage waveform

### 5. Concluding Remarks

Application of stand-alone Photo-voltaic (PV) systems for improving the total harmonic distortion current (THDi) by experimental study involves design, testing, and simulation twelve-pulse diode rectifier AC/DC Power Conversion system applied trapezoidal voltage waveform. Study conclusions and remarks can be summarized as follows.

1. When applying trapezoidal voltage source to 6-pulse diode rectifier, total harmonic current distortion can be reduced by 15.6%. From Fig.8 and Fig.9, the best trapezoidal voltage source should have 1 ms rising time and falling time, with 30 degrees delay offset between phase-to-phase, shown in Fig.9 (a).

2. From simulation, by applying trapezoidal voltage, 12-pulse diode rectifier can decrease total harmonics current distortion (%THDi) of line input current of three winding power transformer ( $\Delta/\Delta/Y$ )

by 14%, shown in Fig.10 (d). This process can be repeated to get even lower total harmonic current distortion, i.e. 12 -pulse converter to 24-pulse converter.

3. Size of capacitor filter output DC voltage is smaller because trapezoidal voltage helps to lessen percent ripple factor.

4. The designed 12-pulse converter is very beneficial when line voltage from transmission line is interrupted. The reason is from signal-mixing technique as previously discussed (Fig.2, Fig.3).

5. According to the test of PV stand-alone system (amorphous silicon PV Panels (0.75x120 Panel (5 kWp)) the average of measured radiation is 382 W/m<sup>2</sup>. Average ambient temperature of PV cell is 28.01°C. Module temperature PV cell has an average of 31.66°C. The Output power of PV is a maximum of 5.6 kWp. The average of measured PV-Array (Pdc) is 1.2 kW/day. Potential-energy is 247.54 kWh.d. The efficiency of PV-array system is a maximum of 19% with an average of 5.45%.

6. This provides guidelines to further analyze and improve power quality in an electrical system by a PV stand-alone photovoltaic-based system for energy storage (DC-link).

### 6. Acknowledgements

The authors would like to express sincere thanks to the Department of Alternative Energy Development and Efficiency Ministry of Energy, Professor Dr. Toshihiko Noguchi for advice and suggestions, their cheerful invaluable support, discussions, and comments at Nagaoka University of Technology.

### 7. References

[1] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE Std 1159-1995, Jun, 1995, pp. 16-18.

- [2] A. Al-Mathnani, A Mohamed, and M. A. Mohd Ali, Photovoltaic Based Dynamic Voltage Restorer for Voltage Sag Mitigation, The 5<sup>th</sup> Student Conference on Research and Development-SCORED, Dec 11-12, 2007.
- [3] Carlos F. Christiansen, Maria ines Valla, and Claudio H. Rivetta, A Synchronization Technique for Static Delta-Modulated PWM Inverters, IEEE Transactions on Industrial Electronics, Vol. 35(4), November 1988, pp. 502-507 .
- [4] Sunt Srianthumrong, Hideaki Fujita, and Hirofumi Akagi, Stability Analysis of a Series Active Filter Integrated With a Double-Series Diode Rectifier, IEEE Transactions on Power Electronics, Vol. 17(1), January 2002, pp. 117-124.
- [5] N. Watanakul, A Analytical and Evaluated Study of the Photovoltaic System Efficiency, Thammasat University Research Project for the Department of Alternative Energy Development and Efficiency Ministry of Energy, 2007.
- [6] J-P. Lee, B-D. Min, D-W. Yoo, T-J. Kim, and J-Y. Yoo, A New Topology for PV DC/DC Converter with High Efficiency under Wide Load Range, Power Electronics and Applications, 12<sup>th</sup> European Conference on Power Electronics and Applications, Aalborg, Denmark, 2-5 Sept., 2007.
- [7] ST Microelectronics, An MCU-based Low Cost Non-inverting Buck-boost Converter for Battery Chargers, Application note AN2389. Aug, 2007.
- [8] International Energy Agency (IEA). Photovoltaic Power Systems TASK 2 –Performance, Reliability and Analysis of Photovoltaic System (IEA PVPS Task 2)
- [9] D. Mayer, and M. Heidenreich, Performance Analysis of Stand-alone PV System from a Rational Use of Energy Point of View, The 3<sup>rd</sup> world Conference on Photovoltaic Energy Conversion, Osaka, Japan, May 11-18, 2003, pp. 2155-2158.
- [10] A. Sannino, J. Svensson, and T. Larsson, Power-electronic Solutions to Power Quality Problems, Electric Power System Research, Vol. 66(1), pp. 71-82, Elsevier Science, 2003.
- [11] K. Haddad, G. Joos, and S. Chen, Control Algorithms for Series Static Voltage Regulators in Faulted Distribution Systems, 30th Annual IEEE Power Electronics Specialists Conference, Vol. 1, Jun 1999, pp. 418-423.
- [12] Kasuni Perera, Arulampalam Aputharajah, Sanath Alahakoon, and Daniel Salomonsson, Automated Control Technique for a Single Phase Dynamic Voltage Restorer, Proceeding of the International Conference on Information and Automation, December 15-17, 2006, pp. 63-68, Colombo, Sri Lanka.
- [13] M. Vilathgamuwa, A.A.D. Ranjith Perera and S.S. Choi. Voltage Sag Compensation with Energy Optimized Dynamic Voltage Restorer” IEEE Trans. on Power Delivery, Jul 2003, Vol. 18(3), pp. 928-936.
- [14] M.R. Banaei, S.H. Hosseini, and M.D. Khajee, Mitigation of Voltage Sag Using Adaptive Neural Network with Dynamic Voltage Restore, Intl. Power Electronics and Motion Control Conference, (IPEMC2006), CES/IEEE 5<sup>th</sup> Aug 14-16, 2006, Shanghai, China.
- [15] A. Ramasamy, V. K. Ramachandaramurthy, R. K. Iyer, and L.Z. Liu, Control of Dynamic Voltage Restorer Using TMS320F2812, 9<sup>th</sup> Intl Conference: Electrical Power Quality and

- Utilization, Barcelona, Oct 9-11, 2007.
- [16] H.P. Krug, T. Kume, and M. Swamy, Neutral-Point Clamped Three-level General Purpose Inverter - features, Benefits and Applications, IEEE 35th Annual Power Electronics Specialists Conference, Aachen, Germany, Jun 20-25, 2004, pp. 323 - 328.
- [17] A. Al-Mathnani, A Mohamed, and M. A. Mohd Ali, Photovoltaic Based Dynamic Voltage Restorer For Voltage Sag Mitigation, The 5<sup>th</sup> Student Conference on Research and Development-SCORED, Dec 11-12, 2007.
- [18] J-P. Lee, B-D. Min, D-W. Yoo , T-J. Kim, and J-Y. Yoo , A new Topology for PV DC/DC Converter with High Efficiency Under wide Load Range, Power Electronics and Applications, 12<sup>th</sup> European Conference on Power Electronics and Applications, Aalborg, Denmark, 2-5 Sept., 2007.

# HVAC's Chilled Water Flow and Temperature Prediction for Buildings in Tropical Zones

**Manwad Chokesuwattanaskul, Apichat Praditsmanont\*  
and Chakguy Prakasvudhisarn\***

School of Technology, Shinawatra University  
15th Floor, Shinawatra Tower III, Chatuchak, Bangkok, 10900 Thailand

**Bundit Limmeechokchai**

School of Manufacturing Systems and Mechanical Engineering,  
Sirindhorn International Institute of Technology, Thammasat University,  
Rangsit Campus, Pathum Thani, 12121, Thailand

## Abstract

The weather in a tropical region, which mainly consists of high temperature and humidity, is clearly different from those in cold and desert climate regions. These factors strongly affect the cooling load demand of heating, ventilating, and air conditioning (HVAC) systems in order to provide required thermal comfort in buildings. Thus, they should be taken into consideration together for determining cooling load demand. Traditional approaches such as physically based models and statistically based methods are extremely difficult to derive and time consuming to develop. Artificial neural networks (ANNs), which are a powerful modeling technique with robust, fast, and nonlinear modeling advantages, can flexibly and simply capture ambient conditions and cooling demand. In this study, the main objective is to investigate the performance of ANNs' predictive ability for HVAC systems. The ANNs are applied to predict flow and temperatures of chilled water in HVAC systems of a multifunctional building in Thailand. The obtained model can be used to effectively plan the energy use of such systems. In addition, thermal energy storage can be properly managed, including its capacity and size. The main features representing temperature and humidity are ambient temperature and relative humidity of indoor and outdoor conditions. The characteristics of cooling load demand are flow rate and temperatures of chilled water. The obtained prediction results show that a properly designed ANN model outperforms multiple linear regression (MLR). Moreover, they can easily be extended to predict multiple factors (such as temperatures) with satisfactory results.

**Keywords:** Artificial neural networks, Building, Cooling load, Temperature and humidity, Tropical zone

## 1. Introduction

World energy demand has considerably increased over the past decades. Due to the concern of energy shortage in the near future, the use of

renewable energy has gained a wide acceptance as an alternative solution. Energy conservation and management are a tangibly strategic approach to better energy consumption and utilization. For buildings, energy efficiency in subsystems such as in

HVAC and lighting systems is a key to reduce the cost in building operations. In Thailand, HVAC systems take the highest share of energy consumption in buildings at approximately 50-70% [1-3], which is a direct consequence of cooling load demand. Its energy use and utility cost can be reduced significantly by using Thermal Energy Storage (TES) systems [4]. TES is as an integrated system for optimal operation of a chiller plant to efficiently store the thermal load and control the distribution system. Peak load reduction can be managed by producing chilled water or ice during times of lower electricity rate, at night time. However, thermal storage systems are often found not to operate as efficiently as estimated during its design stage [5]. An equilibrium condition is necessary in which the required amount of thermal energy storage is equal to the available produced amount of energy storage. Operational management and energy use optimization are required to overcome this problem. The pattern of cooling load demand in buildings should be specifically identified to increase the efficiency of TES during the design stage and actual operation. Therefore, a proper predictive model pattern of cooling load demand is necessary for the accomplishment of an energy conservation goal. Based on these reasons, this research was conducted to study a predictive model of cooling load demand by using artificial neural networks (ANNs). ANNs are attractive because they can capture patterns of input factors regardless of their statistical distribution assumption. This assumption must be verified to validate the results obtained by traditional statistical methods. Furthermore, multiple output prediction can be simply developed by using ANNs if outputs are correlated. This is rather difficult and time consuming for other methods.

Cooling load is a requirement by HVAC systems to provide a thermal

comfort condition for occupancy. Heat is removed from the conditioned space to maintain a thermal comfort condition. Several parameters such as, outdoor air temperature, relative humidity, solar radiation and wind speed are outdoor environmental factors that affect the amount of cooling load requirement. Moreover, the number and activity of occupants in buildings also influence the amount of cooling load demand. Outdoor temperature is a key environment factor that has been selected as neural networks input to accurately predict cooling load for desert and subtropical regions in Kuwait and Japan, respectively [6-7].

For a tropical climate, as in Thailand, temperature and relative humidity ratios are high and different from those in other geographic zones. These two factors directly affect human comfort, resulting in increasing amounts of cooling load demand in buildings. In this paper, these two factors of outdoor and indoor conditions are selected as inputs for Cooling load prediction. Supply and Return Chilled Water Temperatures (SCWT and RCWT) and Chilled Water Flow (CWF) of air conditioning system are used as outputs. Cooling load demand could mainly be computed by the chilled water flow and the difference of chilled water temperatures.

## 2. Literature Review

ANNs are widely used in various areas of energy management, such as overall thermal transfer value, cooling load, air ventilation and thermal comfort in buildings. Demonstrated by several articles, ANNs have a better capability over traditional methods, such as time series and regression. Their advantages are non-linear modeling capability and faster development time.

Focusing on the predictive capabilities of ANNs, Kreider and Wang [8] studied the application of expert systems to

HVAC diagnostics in commercial buildings by using ANNs for determining the energy use of chillers based on hourly averaged data collected from the system. Karatasou et al. [9] implemented modeling and predicting a building's energy use with neural networks. The statistical procedures such as hypothesis testing, information criteria and cross validation were advantageously used in term of guidance to improve the performance of ANN for modeling and predicting a building's energy use. Kajl et al [10] proposed a fuzzy-neural assistant as a comparable method to the DOE-2 building analysis program for the simplified and detail estimation methods of a building's energy consumption. Three beneficial input parameters including orientation, insulation thickness, and transparency ratio were developed for the prediction of building energy consumptions by Ekici and Aksoy[11]. ANNs prediction for the energy consumption of passive solar, with faster development time than the dynamic simulation programs, has been studied by Kalogirou and Bojic[12]. Olofsson and Andersson have also developed ANNs to perform long-term energy demand prediction based on short-term measured data. The model parameters were indoor and outdoor temperature difference and energy for heating and internal use [13]. Moreover, prediction of a building's temperature using neural networks models for predictive control of air conditioning system has been proposed by Ruano et al. [14]. A neural network was also applied to the thermal load prediction case. Investigation of four predictive methods, namely Autoregressive Integrated Moving Average (ARIMA), Exponentially Weighted Moving Average (EWMA), Linear Regression (LR), and ANNs, was comparatively conducted for the use of hourly thermal load prediction by Kawashima et al. [15]. ANNs gave the highest thermal load prediction accuracy and clearly outperformed other methods.

This resulted in a decrease of operating cost without thermal energy shortage. Optimizations based on neural networks modeling have also been implemented to the energy management field. Curtiss et al. reported [16] ANNs could be used to optimize the energy consumption in a commercial scale HVAC system. Information from an actual system was used for training a network to optimize the energy consumption without sacrificing comfort by considering all the physical limitations of the system. On-line set-point resets in an actual HVAC control system were successfully performed by ANN based energy management. A variant of ANNs has been applied to energy management as well [7]. General regression neural networks (GRNN) are a powerful instrument for optimizing thermal energy storage in buildings based only on the use of external temperature. External hourly temperature readings for a 24-hour period were used as network inputs to predict an hourly cooling load for the next day.

The application of ANNs was introduced to the system identification and the intelligent control of an air handling unit by Albert and Wai [17]. ANN traced the online parameters relative to the air handling unit as an identifier and then controlled the system. Atthajariyakul and Leephakpreeda [18] studied a practical approach to determine human thermal comfort quantitatively via neural computing. The feedforward neural network model allowed a real time determination of a thermal comfort index, the predicted mean vote (PMV) index. In contrast, a major obstacle of the conventional method for PMV calculation is its long computational time and hence it cannot be calculated in real time.

The literature above has confirmed the performance of neural networks in prediction. However, there is still no study about the performance of neural networks in cooling load prediction for tropical regions

by using both temperature and humidity of indoor and outdoor conditions as inputs. Therefore, this study is focused on the feasibility of using ANNs for tropical cooling load prediction.

### 3. Methodology

Section 3.1 and 3.2 discuss the fundamentals of a machine learning technique, ANNs, and widely used traditional method, multiple linear regression (MLR).

#### 3.1 Multiple outputs artificial neural networks regression

Artificial Neural Networks (ANNs) imitate the learning process of human brain. They eliminate the need of using complex mathematically explicit formulas, computer models, and impractical and costly physical models. ANNs can capture relationships between input and output by adjusting weights on each link while learning from data. Their advantages are robustness, speed, and nonlinear modeling. Furthermore, they can perform both single and multiple output predictions. In this study, a feedforward backpropagation neural network was attempted to predict flow and temperatures of chilled water of an air conditioning system. A neural network normally has two elementary components, processing elements and connection weights. A feedforward network has no loops as opposed to a feedback type. A classic learning algorithm, backpropagation, was used by propagating errors backward to train and update the weights on each link of a neural network with training examples. These weights capture the pattern of multivariable functions through learning. In other words, they were used to capture the relationship between temperature and humidity of indoor and outdoor conditions, and flow and temperatures of chilled water of air conditioning systems. Weight adjustment between processing nodes in

backpropagation is carried out according to the difference between the target and the output values of the neural network. This difference is measured by mean squared error shown below [19]:

$$E = \frac{\sum_{p=1}^P \sum_{k=1}^K (d_{pk} - o_{pk})^2}{pk} \quad (1)$$

where  $d_{pk}$  is the  $k^{th}$  desired value of the  $p^{th}$  data and  $o_{pk}$  is the actual output.

The weights (W) are adjusted toward the gradient direction that produces a smaller approximation error as follows:

$$\mathbf{W}(t+1) \leftarrow \mathbf{W}(t) + \eta \delta(t) \mathbf{y}(t) \quad (2)$$

where  $\eta$  is a positive constant called learning rate,  $\delta$  is the gradient of the difference between the desired and actual neuron's responses, and  $\mathbf{y}$  is the input vector. The weight matrix adapted at time  $t$  becomes equation (2) at the next instant.

In regression problems, the following set of data  $\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_p, \mathbf{y}_p)\} \subset \mathcal{R}^m \times \mathcal{R}^n$  can be approximated by using ANNs. The  $\mathbf{x}_i$  is the vector set of temperature and humidity of indoor and outdoor conditions and  $\mathbf{y}_j$  is the output vector which consists of service and return chilled water temperatures and chilled water flow conditions. ANNs prediction model was first implemented with a single output to investigate its effectiveness as compared with a traditional method like the MLR. Theoretically, the advantages of ANNs are robustness, nonlinear modeling ability and nonparametric concept. However, ANNs have a major disadvantage as the physical relationship among input and output cannot be explained. A proper architecture of ANNs must be chosen from the split data sets among training and validation sets to avoid overfit problems. In addition, care must be taken while selecting a proper architecture of ANNs to avoid overfit problems. Data must be split into three sets,

training, validation, and test sets, to help select such architecture.

### 3.2 Multiple linear regression

Multiple linear regression analysis is a statistical technique which is very useful for exploring the relationships between two or more variables  $(\mathbf{x}_i, y_i)$ .  $\mathbf{x}_i$  represents the independent variables which contain a set of temperature and humidity variables of indoor and outdoor conditions.  $y_i$  is an interesting dependent variable consisting of the set of chilled water flow of an air conditioning system. Chilled water flow output was chosen for this pilot experiment due to its variation in operation. Normally, a chiller system is operated in accordance with the cooling load demand by varying the amount of chilled water flow and fixing the value of chilled water temperatures at some level ranges. That means the amount of chilled water flow can consistently reflect the characteristic of cooling load demand.

Suppose that there are  $m$  independent variables and  $p$  observations  $(x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}, y_i); i = 1, 2, \dots, p$ . The fitted regression model can be described as:

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} \quad i = 1, 2, \dots, p; \quad (3)$$

and  $j = 1, 2, \dots, m$ .

The parameters  $\beta_0$  and  $\beta_j, j = 1, 2, \dots, m$  are called the regression coefficients which will be determined by the method of least squares.

The difference between the observation  $\{y_i\}$  and the fitted value  $\hat{y}_i$  is a residual,  $e_i = y_i - \hat{y}_i$ .  $\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^m \hat{\beta}_j x_{ij}$  and  $\hat{\beta}_0$  and  $\hat{\beta}_j$  are the estimators of the regression coefficients. The criterion used is the sum of squared error:

$$S_r = \sum_{i=1}^p e_i^2 = \sum_{i=1}^p (y_i - \hat{y}_i)^2 \quad (4)$$

The quality of the models is estimated by considering the correlation coefficient  $R$  between the actual and predicted outputs. It can be described as:

$$R = \frac{p \sum y_{ij} \hat{y}_i - (\sum y_{ij})(\sum \hat{y}_i)}{\sqrt{p \sum y_{ij}^2 - (\sum y_{ij})^2} \sqrt{p \sum \hat{y}_i^2 - (\sum \hat{y}_i)^2}} \quad (5)$$

MLR is the most widely used regression method because it can describe the relationship between input and output. Moreover, it is quite fast and simple due to the use of a closed form solution for the determination of regression coefficients.

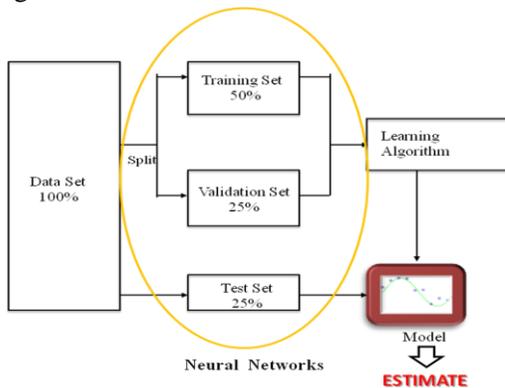
## 4. Results and discussions

The objective of this study is to investigate the feasibility of using neural networks to predict the cooling load demand. The causal method was used for such a task. The data collection was performed between November 2008 and January 2009 at the main campus of Shinawatra University in Pathum Thani. The input parameters are temperature and relative humidity of indoor and outdoor conditions. The outputs are service and return chilled water temperatures and chilled water flow data. Temperature and relative humidity of indoor and outdoor conditions were recorded by temperature and humidity measuring devices and a weather station at the main campus. The Building Automation System (BAS) monitoring program was used to collect data of supply and return chilled water temperatures and chilled water flow. Those data of chilled water temperatures are in Celsius with the magnitudes of ones and tens. The data of chilled water flow are in gallons per minute (GPM) with the magnitude of hundreds. All data were recorded every fifteen minutes. ANNs and

MLR predictions were implemented in MATLAB 7 with neural networks and multiple linear regression toolboxes to develop the cooling load demand model.

**4.1 Chilled water flow (CWF) prediction**

A pilot experiment for MLR’s prediction has been implemented to compare its predictive ability with that of ANNs. The prediction was performed only with CWF as single output because, as mentioned earlier, CWF is a major factor of cooling load determination. The experiment was set up by using large and small sizes of data sets with 3520 and 200 selected data points in order to see the characteristic of MLR prediction on different sizes of data sets. These selected data were randomly divided into training and test subsets for building and verifying the MLR predictive model. The training and test sets contained 75% and 25% of data sets. The selected data of ANNs were randomly divided into three subsets for training 50%, validation 25%, and testing 25%, as demonstrated in Figure 1.



**Figure 1.** Model selection method diagram.

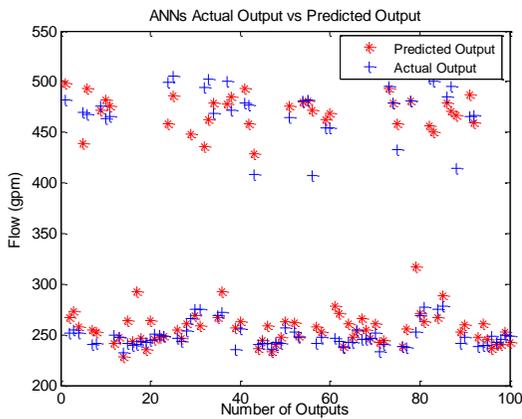
Details for the procedures of ANNs single output prediction are discussed in section 4.2. The performance of each technique was evaluated by using the average R-value. Comparative accuracy results between the average R-values of MLR’s and ANNs’ predictions are presented in Table 1. In comparison

between these two regression methods, ANNs show an excellent prediction performance due to the high level of R-values for both small and large size data sets. Furthermore, when applied to a much larger data set, ANNs produce a slight decrease of performance, although they have to deal with higher variation. By increasing the data size from 200 to 3520 data sets, the R-value results drop from 0.930 to 0.908 for training set and 0.899 to 0.892 for test set. For MLR prediction, it performs fairly well for a small size data set. Conversely, it makes a poor prediction for a large size of data sets. By increasing the data size from 200 to 3520 data sets, the R-value results decrease from 0.807 to 0.645 for training sets and 0.797 to 0.641 for test sets.

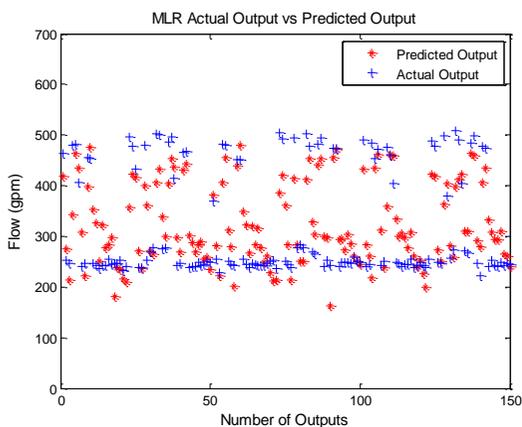
**Table 1.** Comparative accuracy results between the average R-values of MLR and ANNs for CWF.

Predictive Model	Average R-value			
	200 data sets		3520 data sets	
	Training Set	Test Set	Training Set	Test Set
ANNs	0.930	0.899	0.908	0.892
MLR	0.807	0.797	0.645	0.641

The predicted outputs by ANNs and MLR regression are plotted against the actual outputs as shown in Figures 2 and 3. The predicted outputs from ANNs regression are quite close to the actual outputs, whereas the predicted outputs from MLR vary widely from the actual outputs. These obviously illustrate the superiority of ANNs over MLR, numerically and graphically.



**Figure 2.** ANNs actual output vs predicted output.



**Figure 3.** MLR actual output vs predicted output.

In summary, the MLR technique was outperformed by ANNs. Higher prediction accuracy for large size and high variation data could be expected from ANNs. As addressed before, there are more than one output parameter to be used for cooling load calculation. This prediction requirement of multiple outputs, chilled water flow (CWF), return chilled water temperature (RCWT), and service chilled water temperature (SCWT) can easily be handled by ANNs. Section 4.2 initially describes single output ANNs and subsequently generalizes to multiple outputs.

## 4.2 ANNs' prediction for CWF, RCWT and SCWT

In this section, the procedures of ANNs' prediction for CWF, RCWT, and SCWT are described for each and every cooling load parameter and all of them simultaneously.

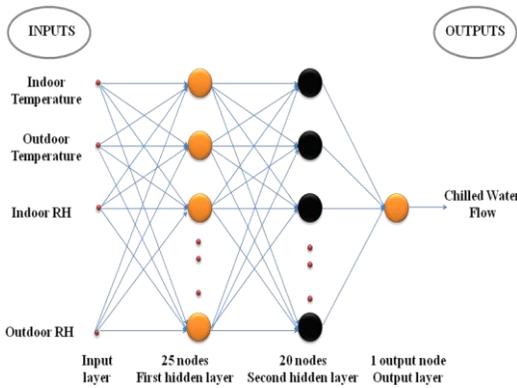
Generally, input parameters of the target function are composed of various magnitudes. The one with higher magnitude may dominate others with lower magnitudes. Therefore, preprocessing should be applied to raw data before training. Thus, the raw data were normalized to  $[-1,1]$  for every factor: temperature of indoor ( $T_i$ ), temperature of outdoor ( $T_o$ ), relative humidity indoor ( $RH_i$ ), and relative humidity outdoor ( $RH_o$ ).

In this study, 3520 data were randomly selected from 8832 for developing the cooling load model. Due to the large selected data set, the holdout method was chosen as a validation technique for model selection and performance estimation of the constructed model. The data were randomly divided into three subsets for training, validation, and testing subsets as illustrated in Figure 1.

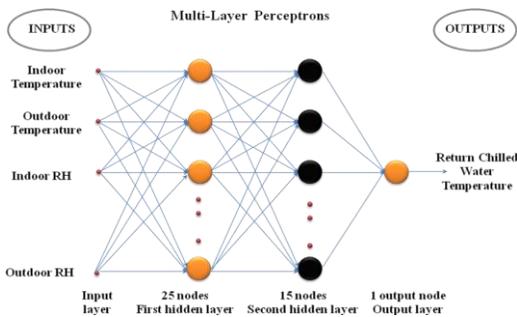
Training neural networks with training set is done to determine optimal weights. Then, a validation set is used for tuning the parameters and estimating the optimal number of hidden units or a stopping point of the training algorithm. The testing set is used to assess the performance of properly trained and validated model. The procedure and randomization eliminates bias of ANNs, while increasing their generalization ability.

For single output prediction, ANNs were designed with 4 inputs and 1 output. There are three single output models for the predictions of CWF, RCWT, and SCWT. Each predictive model was separately developed for each output. Temperature and relative humidity of indoor and outdoor conditions were used as inputs. The proper architecture, the number of hidden layers,

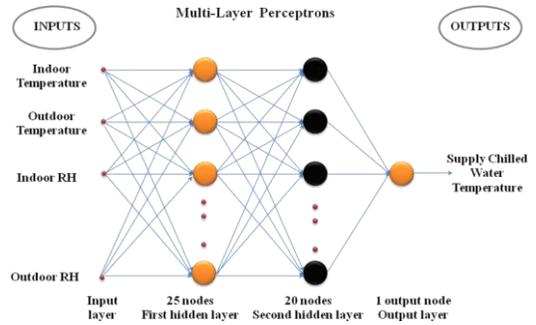
the number of hidden nodes in each layer, and the number of iterations was selected based on generalization performance indicators by using trial and error approach. The combination of these ANNs parameters that provided the lowest training error, and shared the same trend as validation error, was experimentally found. The experiment was conducted for 10 runs for each structure from 5 to 50 nodes in the first hidden layer. Five nodes were increased each time. The second hidden layer was also attempted in the same fashion. The final architectures with minimum error are 4-25-20-1, 4-25-15-1 and 4-25-20-1 for chilled water flow prediction, return and service chilled water temperature predictions, respectively, as shown in Figures 4- 6.



**Figure 4.** Final architecture, 4-25-20-1, of ANNs for CWF prediction.

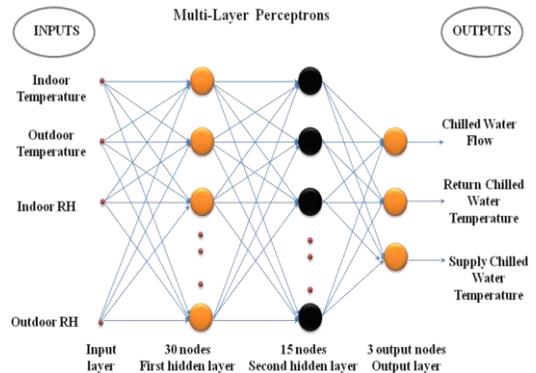


**Figure 5.** Final architecture, 4-25-15-1, of ANNs for RCWT prediction.



**Figure 6.** Final architecture, 4-25-20-1, of ANNs for SCWT prediction.

For multiple outputs prediction, a neural network was created with 4 inputs and 3 outputs. These four inputs and three outputs were the same as those in the previous single output case but they would be used together to develop a prediction model. The selection for the best architecture was also the same. The final architecture with minimum error is 4-30-15-3 as shown in Figure 7.



**Figure 7.** Final architecture, 4-30-15-3, of ANNs for multiple outputs prediction.

The activation functions used were the hyperbolic tangent sigmoid transfer functions or “tansig” for all hidden nodes. Because it is differentiable and covers the bipolar continuous range (-1,1), it is commonly used in backpropagation networks. The linear transfer function or “purelin” was used for the output node(s) in

the last layer since the network output(s) could take on any value. The speed-up optimizer, Levenberg and Marquardt backpropagation, was used to train neural networks by minimizing mean squared error.

Five measures of accuracy were chosen to evaluate the performance of ANNs. They are coefficient correlation (R), the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Mean Bias Error (MBE).

R-value can be determined by using linear regression analysis between the predicted outputs and the desired corresponding targets. MSE and RMSE are commonly used to evaluate models based on two reasons. Firstly, they penalize large forecasting errors proportionately. Most users of forecasts prefer a model that produces consistently moderate errors to one that produces some small errors and some very large errors. Secondly, the mean squared error can be used to estimate the variance of the random error component. MSE and RMSE can be described as:

$$MSE = \frac{\sum_{i=1}^p (y_i - \hat{y}_i)^2}{p} \quad (6)$$

and

$$RMSE = \sqrt{\frac{\sum_{i=1}^p (y_i - \hat{y}_i)^2}{p}} \quad (7)$$

where  $\hat{y}_i$  is the predicted value,  $y_i$  the measured value.

MAE is an average of the absolute errors. It is one of the most popular and simplest for measuring the forecast errors. This measure provides a better intuitive feel for how much error is likely to occur when using a forecast from the model. Its measure is:

$$MAE = \frac{\sum_{i=1}^p |y_i - \hat{y}_i|}{p} \quad (8)$$

MBE provides information on the long term performance of the correlations by allowing a comparison of the actual deviation between actual and predicted outputs term by term. The ideal value of MBE is zero which implies a lack of bias. A positive result or low bias indicates the condition that predicted outputs are consistently lower than the actual outputs. Whereas, a negative result or high bias demonstrates the condition that predicted outputs are consistently higher than the actual outputs. MBE can be described as:

$$MBE = \frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_i) \quad (9)$$

From 10 replications of the best architectures of the single and multiple outputs, the results from each measures of accuracy are presented in Tables 2-6. After Training, the validation set was grouped with the training set to determine the performance of the developed predictive model for training. The actual assessment was also done with the unseen test set.

**Table 2.** Accuracy of average R-values.

ANNs Predictive Models	Final Architecture	Average R-value					
		Training Set			Test Set		
		Output No.			Output No.		
		1	2	3	1	2	3
		CWF	RCWT	SCWT	CWF	RCWT	SCWT
Single Output (CWF)	[4-25-20-1]	0.908	-	-	0.892	-	-
Single Output (RCWT)	[4-25-15-1]	-	0.901	-	-	0.879	-
Single Output (SCWT)	[4-25-20-1]	-	-	0.762	-	-	0.724
Multiple Outputs	[4-30-15-3]	0.904	0.876	0.726	0.890	0.860	0.704

Table 2 demonstrates the comparative accuracy results between average R-values of ANNs predictive models. The average R-values of training and test sets of all models are in the range of 0.70- 0.91 which can reflect the high performance of neural networks in cooling load prediction. The average R-values of training sets of all models are slightly higher than the average R-values of test sets of all models by about 2-3%. In comparison between ANNs predictive models for single and multiple outputs, ANNs predictive models for single output perform a better prediction as they produce the higher results of R-values. For a training set, the average R-value of each single output prediction is slightly greater than the average R-value of multiple outputs prediction, 0.908 and 0.904 for

CWF, 0.901 and 0.876 for RCWT and 0.762 and 0.726 for SCWT. For a test set, the average R-value of each single output prediction is slightly higher than the average R-value of multiple outputs prediction with 0.892 and 0.890 for CWF, 0.879 and 0.860 for RCWT and 0.724 and 0.704 for SCWT. Notably, the average R-values of the training and test sets of SCWT are markedly lower than the average R-values of the other two outputs.

The slight differences between average R-values of training and unseen test sets show that ANNs can predict key parameters of cooling load with high generalization. With large size and high variation of data tested, ANNs are quite robust for cooling load prediction.

**Table 3.** Accuracy of average MSE values.

ANNs Predictive Models	Final Architecture	Average MSE					
		Training Set			Test Set		
		Output No.			Output No.		
		1	2	3	1	2	3
		CWF	RCWT	SCWT	CWF	RCWT	SCWT
Single Output (CWF)	[4-25-20-1]	1720.247	-	-	1993.042	-	-
Single Output (RCWT)	[4-25-15-1]	-	0.281	-	-	0.345	-
Single Output (SCWT)	[4-25-20-1]	-	-	0.319	-	-	0.365
Multiple Outputs	[4-30-15-3]	1787.338	0.348	0.361	2004.598	0.393	0.384

**Table 4.** Accuracy of average RMSE values.

ANNs Predictive Models	Final Architecture	Average RMSE					
		Training Set			Test Set		
		Output No.			Output No.		
		1	2	3	1	2	3
		CWF	RCWT	SCWT	CWF	RCWT	SCWT
Single Output (CWF)	[4-25-20-1]	41.476	-	-	44.643	-	-
Single Output (RCWT)	[4-25-15-1]	-	0.530	-	-	0.587	-
Single Output (SCWT)	[4-25-20-1]	-	-	0.565	-	-	0.604
Multiple Outputs	[4-30-15-3]	42.277	0.590	0.601	44.773	0.627	0.620

**Table 5.** Accuracy of average MAE values.

ANNs Predictive Models	Final Architecture	Average MAE					
		Training Set			Test Set		
		Output No.			Output No.		
		1	2	3	1	2	3
CWF	RCWT	SCWT	CWF	RCWT	SCWT		
Single Output (CWF)	[4-25-20-1]	26.494	-	-	28.653	-	-
Single Output (RCWT)	[4-25-15-1]	-	0.361	-	-	0.391	-
Single Output (SCWT)	[4-25-20-1]	-	-	0.374	-	-	0.396
Multiple Outputs	[4-30-15-3]	27.634	0.405	0.408	29.346	0.425	0.416

The accuracy results of the average MSE, RMSE and MAE values of ANNs predictive models are depicted in Tables 3, 4 and 5, respectively. The average MSE and RMSE values of all ANNs predictive models demonstrate a similar tendency to

the average MAE values of all ANNs predictive models. Results of training sets for all outputs in every model are slightly better than those of unseen test sets for all outputs.

**Table 6.** Accuracy of average MBE values.

ANNs Predictive Models	Final Architecture	Average MBE					
		Training Set			Test Set		
		Output No.			Output No.		
		1	2	3	1	2	3
CWF	RCWT	SCWT	CWF	RCWT	SCWT		
Single Output (CWF)	[4-25-20-1]	0.398	-	-	-1.271	-	-
Single Output (RCWT)	[4-25-15-1]	-	-0.00027	-	-	0.0036	-
Single Output (SCWT)	[4-25-20-1]	-	-	-0.0021	-	-	0.0049
Multiple Outputs	[4-30-15-3]	0.470	-0.0025	0.0026	-1.690	-0.0065	0.0060

Table 6 demonstrates the comparative accuracy results by using average MBE of ANNs predictive models. The average MBE results indicate a good performance of ANNs' predictions as they are entirely close to zero. All results of average MBE are between -0.0025 and 1.69. The results of average MBE for training sets of both types of outputs are slightly better than those of the unseen test sets for all corresponding outputs. Results of average MBE of CWF are slightly higher than the results of average MBE of RCWT and SCWT due to the differences in magnitude between CWF and RCWT and SCWT.

In summary, accuracy evaluations of all measures are consistent as the average results of training set and are superior to those of test set. This illustrates the distinct strengths of ANNs with high robustness and good generalization capabilities.

Comparing ANNs predictive models for single and multiple outputs, better results of all measures, R, MSE, RMSE, MAE and MBE, imply better performance in prediction. The less error in ANNs' prediction could be expected from single output predictive model when compared with multiple outputs model.

All three separated single output models slightly outperform the multiple outputs model due to the less complicated calculation required during input-output mappings of ANNs. However, the development time of the former was much large than that of the latter. The times it took to develop each single output model and the combined outputs model were about the same but three different outputs models were needed for the former and only one model was enough for the latter. Consequently, the selections of ANNs' best architectures including training, validation, and testing processes need to be implemented separately three times. The models accuracy is not much different, but the development time becomes more critical. Therefore, ANNs' multiple outputs

prediction is more appropriate to be applied for practical work in this study

## 5. Conclusions

This work demonstrated that ANNs can be effectively applied to predict cooling load demand. As a result from a pilot experiment, ANNs have outperformed MLR in prediction. In comparison of ANNs and MLR, ANNs have superior performance in prediction as they could handle huge different sizes of data sets with high accuracy, about 0.9 for R-values. MLR showed fair performance in prediction for small data sets and quite poor predictive performance for large data sets, about 0.8 and 0.65 for R-values, respectively.

ANNs show excellent performance in prediction of flow and temperatures of chilled water of air conditioning system by using ambient temperatures and humidity rates of indoor and outdoor conditions with both large and small sizes of data sets. Good generalization of ANNs for unseen test set can be obtained. ANNs also show robustness in prediction with large and high variation data sets for all measures. Clearly, ANNs can be applied to capture implicit relationships between input and output factors for cooling load demand calculations. The results from ANNs single output prediction are more accurate than those from ANNs multiple outputs prediction with a slight difference. Development time of ANNs multiple outputs prediction is much less as opposed to the combined development time of all three separated single output models. As a result, ANNs' multiple outputs prediction can be further applied to real-time energy planning. They can effectively be utilized for optimization and management of energy use in HVAC systems. In addition, thermal energy storage can suitably be properly managed, including its capacity and size, by using this intelligent technique.

## 6. References

- [1] Asian Institute of Technology, The Development of Building Energy Code for Control Building in Thailand, Department of Alternative Energy Development and Efficiency, Ministry of Energy, pp. 8, 2004.
- [2] Wibulswas, P. and Sopitpagdeepong, S., Energy Management in a Medical School, Asean-EC Energy Conservation Seminar Bangkok, Thailand, pp. 182-191, 1987.
- [3] Chokesuwattanaskul, M., Energy Management in Main Hall Building. Bangkok: Shinawatra University, 2006.
- [4] Henze, G. P., Felsmann, C., and Knabe, G., Evaluation of Optimal Control for Active and Passive Building Thermal Storage, International Journal of Thermal Sciences, Vol. 43, pp. 173-178, 2004.
- [5] Nakajima, Y., Sugi, J. Saito, M., Harmagishi, H., Hattori, D., and Matsumoto, T., Hierarchical Bayesian Neural Nets for Air-conditioning Load Prediction: Nonlinear Dynamics Approach, Proceeding of IEEE World Congress on Computational Intelligence, Vol. 3, pp. 1948-1953, 1998.
- [6] Kashiwag, N., and Tobi, T., Heating and Cooling Load Prediction Using a Neural Network System, Proceedings of 1993 International Joint Conference on Neural Networks, pp. 939-942, 1993.
- [7] Ben-Nakhi, A. E. and Mahmoud, M. A., Cooling Load Prediction for Buildings Using General Regression Neural Networks, Energy Conversion and Management, Vol.45, pp. 2127-2141, 2004.
- [8] Kreider, J. F., & Wang, X. A., Artificial Neural Network Demonstration for Automated Generation of Energy use Predictors for Commercial Buildings, In J. S. Haberl, R. M. Nelson, & C. C. Culp (Eds.), The Use of Artificial Intelligence in Building Systems, pp. 193-198, 1995.
- [9] Karatasou, S., Santamouris, M. and Geros, V., Modeling and Predicting Building's Energy Use with Artificial Neural Networks: Method and Results, Energy and Buildings, Vol. 38, pp. 949-958, 2006.
- [10] Kajl, S., Roberge, M. A., Lamarche, L., and Malinowski, P., Evaluation of Building Energy Consumption Based on Fuzzy Logic and Neural Networks Applications, Proceedings of CLIMA 2000 Conference on CDROM, 2000.
- [11] Ekici, B. B. and Aksoy, U. T., Prediction of Building Energy Consumption by Using Artificial Neural Networks, Advances in Engineering Software, Vol. 40, pp. 356-362, 2009.
- [12] Kalogirou, S. A. and Bojic, M., Artificial Neural Network for the Prediction of the Energy Consumption of a Passive Solar Building, Energy, Vol. 26, pp. 479-491, 2000.
- [13] Olofsson, T. and Andersson, S., Long-term Energy Demand Prediction Based on Short-term Measured Data, Energy and Building, Vol. 33, pp. 85-91, 2001.
- [14] Ruano, E. A., Crispim, M. E., Conceicao, E. Z. E. and Lucio, M. M. J. R., Prediction of Building Temperature Using Neural Networks Models, Energy and Buildings, Vol. 38, pp. 682-694, 2006.
- [15] Kawashima, M., Dorgan, C. E., & Mitchell, J. W., Hourly Thermal Load Prediction for the Next 24 Hours by ARIMA, EWMA, LR, and an Artificial Neural Network. ASHRAE Trans, 1, pp. 186-200, 1995.
- [16] Curtiss, P., Kreider, J. F. and Brmdemuehl, M. J., Energy

- Management in Central HVAC Plants Using Neural Networks. ASHRAE Trans, 100, pp. 476- 493, 1994.
- [17] Albert, T. P., and Wai, L., A Neural-Network-based Identifier/Controller for Modern HVAC Control. ASHAE Trans, Vol. 10,(2), pp.11-31, 1995.
- [18] Atthajariyakul, S., and Leephakpreeda, T., Neural Computing Thermal Comfort Index for HVAC Systems. Energy Conservation and Management, Vol.46, pp. 2553-2565, 2005.
- [19] Prakasvudhisarn, C., Kunnapapdeert, S. and Yenradee, P., Optimal Cutting Condition Determination for Desired Surface Roughness in End Milling, International Journal of Advanced Manufacturing Technology, Vol.41, pp. 440-451, 2009.

# Stage-Warping Load Sharing Strategy for Fine Grain Applications over Grid Environments

**Natthakrit Sanguandikul and Natawut Nupairoj**

Department of Computer Engineering,  
Chulalongkorn University, Bangkok, 10110 Thailand  
E-mail: Natthakrit.S@student.chula.ac.th

## Abstract

A Load sharing strategy is one of important keys to improve the performance of computing systems. Nowadays, large scale computing system can be created by aggregating multiple computing clusters from different organizations using Grid technology. However, it is difficult to define a practical load sharing strategy due to the computing heterogeneity and dynamic behavior of Grid resources. In this work, we introduce a load sharing strategy for distributing workloads among participating clusters. The proposed strategy implements a new job-stealing technique called “stage-warping”, for dynamically adjusting the amount of assigned workloads for each cluster during an execution. In our strategy, the entire workloads are divided into stages which enable the total control of workload assignment during an execution, while still being highly robust against performance fluctuation and information inaccuracy of the computing resources. During execution, faster-than-expected clusters which will finish the assigned workloads during each stage before other clusters will steal left-over workloads from other clusters and let them skip or warp to the foremost stage. This will make all clusters to be fully utilized by finishing their assigned workloads almost at the same time near the end of each stage, resulting in a better overall parallel performance of load sharing strategy. We evaluate our proposed strategy using a set of simulation experiments based on the parameters from the available computing resources in ThaiGrid, as well as, fine-grain applications which can serve as an example of computationally intensive applications. The results show that our proposed strategy can achieve better parallel runtimes when being compared to other existing methods, especially when the estimator of the underlying system is not accurate.

**Keywords:** Distributed Computing, Grid Technology, Load Sharing Strategy

## 1. Introduction

The current trend of parallel processing is to create a low-cost supercomputer by aggregating multiple PCs together. For example, Grid computing

focuses on aggregating the computing resources geographically distributed across different organizations [1]. Therefore, the number of computing nodes and the complexity of the underlying system will be dramatically increased. In addition, the

performance of Grid resources can change abruptly because they are not dedicated resources and are subject to the local resource policy. In order to utilize this vast amount of computing power effectively, we must employ a load sharing strategy in order to keep every computing resource busy until the end of an execution. Although many load sharing strategies have been proposed in the past [2-3], these strategies are aimed for single cluster computing environments only. Therefore, they fail to address new characteristics in multiple-cluster environments [4] such as communication structure between clusters, large overhead over WAN, high computing heterogeneity, and the dynamic behavior of non-dedicated resources. Moreover, most load sharing strategies proposed for large scale computing environments [5-6] still require some specific performance indicators of computing resources for making load decisions. These indicators are difficult to derive as there are various parameters which can affect the overall runtime such as processor speed, system architecture, communication bandwidth, and submitted applications. Thus, it is very difficult to collect all necessary internal information within each cluster and define a sophisticated resource model that can truly predict the performance of the computing system [7-8]. Given the growing complexity in computing environments and applications, the traditional methodologies are no longer practical.

To address this foreseeable problem, we propose a load sharing strategy with new job-stealing technique for distributing workloads among participating clusters. Our strategy first divides total workloads into different stage sizes before further assigning workloads allocated within each stage to the participating clusters according to their performance indicators. By implementing our new job-stealing technique called "stage-warping", faster-than-expected clusters can execute

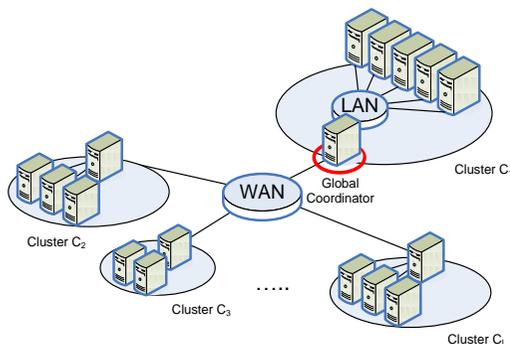
additional tasks by stealing workloads to be executed by other clusters. Hence, those clusters can catch up with the faster-than-expected clusters by skipping or warping to the same current stage. This behavior will give additional robustness against the performance fluctuation of highly heterogeneous resources in Grid environments. In addition, our performance indicator is just a rough performance estimator of the underlying system, which a global coordinator can calculate during an execution. Therefore, there is no need to implement any monitoring services to gather information from the computing clusters at all. To increase the accuracy of this performance indicator, our strategy consists of increasing stages and decreasing stages. The increasing stages are for obtaining accurate performance indicators of the underlying systems and the decreasing stages are for reducing load imbalance near the end of an execution.

Although the proposed strategy in [9] makes load decision based on the performance indicators obtained during an execution like our proposed strategy, it uses job replication to address dynamic behavior of Grid resources, which is more suitable for coarse grain applications. Moreover, given a fine grain computing application, our proposed strategy can balance the workloads between each cluster dynamically which eliminates load imbalance near the end of an execution. Therefore, we do not have to implement job replication, and we can also save some computing cycles by doing so. The organization of this paper is as follows. Section 2 provides the details about the models within our work. Section 3 describes other related works. In section 4, we propose our stage-warping load sharing strategy. The performance evaluations are shown in section 5. And finally, we conclude our work in section 6.

## 2. Models

### 2.1 Grid Model

We assume that the computing system is a multiple-cluster computing environment. It consists of  $N$  computing nodes which will be grouped together into  $L$  clusters  $\{C_1, C_2 \dots C_L\}$ . These computing clusters communicate with each other over WAN while the intra-cluster communication will be made over LAN. The computing nodes within the same cluster will be considered homogeneous and have the same computing power. With this assumption, we can increase the computing heterogeneity within our system by specifying some clusters to have more total computing power than the others. Within each cluster, there is one local gateway which is responsible for distributing workloads submitted by Grid users to other computing nodes in the same cluster and also handles the inter-cluster communications. In addition, one of the local gateways will also serve as the global coordinator which manages submitted jobs and assigns workloads to the other clusters. Global strategy will be used for assigning workloads among participating clusters while local strategy is for assigning workloads within each cluster. Note that the local strategy can be specified differently depending on the local administrator of that cluster.



**Figure 1** The system model of Grid environment.

### 2.2 Application Model

We define an application model based on fine-grain computationally intensive application which consists of  $U$  unit tasks where the computation and communication size of each task can be specified differently. This model represents parameter sweeping applications which are considered as the killer applications for Grid technology [10]. The examples of real life applications that belong to this type are radiation equipment calibration analysis, searching for extra-terrestrial intelligence, protein folding, molecular modeling for drug design, human-genome sequence analysis, brain activity analysis, high-energy physics event analysis, ad hoc network simulation, crash simulation, tomography, financial modeling, and M-cell simulations. Moreover, we also further divide the fine-grain applications into four distinct classes of applications, including uniform, increasing, decreasing, and random distribution [11]. These classes can represent popular applications where the computation size of each unit task can be different during an execution such as Matrix Multiplication, SOR, Reverse Adjoint Convolution, LU Decomposition, and Gauss Jordan Elimination.

## 3. Related Works

The Self-Scheduling strategy (SS) [12] dynamically assigns only one unit task per each request for an idle computing resource. With this behavior, it can achieve almost perfect load balancing because every computing resource will finish within one task of each other. However, this strategy also suffers from high communication overhead. To address this problem while keeping its simplicity, many variations of SS have been proposed [13-14]. One of them, which is famous for robustness is called "Factoring" (FSS) [15]. This strategy assigns workloads into multiple stages. In the first stage, FSS distributes the largest

chunk and decreases the chunk size in the subsequence stages proportionally. During each stage, every processor will receive an equal chunk size of workloads. FSS can reduce communication overhead by sending large chunks at the beginning while it achieves sub-optimal runtime by sending small chunks near the end of computation. To address heterogeneity within the computing system, “Weighted Factoring” (WFSS) [16] is proposed as an extension of FSS. In this strategy, the amount of total unit tasks allocated during each stage is the same as in FSS. However, unlike FSS, WFSS utilizes pre-execution information of the computing resources as weighted values to assign workloads allocated within each stage. We will call a strategy that utilizes pre-execution information an “explicit strategy”.

One of the major weaknesses of explicit strategies is that these strategies rely on static knowledge. Thus, they perform quite poor in a dynamic environment like Grid computing. One of the descendants of FSS called “Adaptive Weighted Factoring” (AWF) [17], addresses this problem by further extending WFSS with an adaptive weighted value called “Weighted Average Performance” (WAP). This weighted value will be re-calculated every stage using the newly obtained computing rates of each resource. Therefore, the explicit information will be used as a weighed value during the first stage only. With this average value, AWF can address the dynamic behavior of the heterogeneous computing system. However, since AWF assigns half of the available workloads during the first stage, the problem of inaccurate explicit information can still affect the performance of this strategy.

#### 4. Proposed Load Sharing Strategy

In this work, we propose a new strategy called “stage-warping load sharing

strategy” (SWS). Our proposed strategy is aimed to address the computing heterogeneity and dynamic behavior of Grid resources by dynamically adjusting workload assignment based on the estimation of the computing power of the participating clusters. However, the estimation can be inaccurate since the performance of Grid resources can change dramatically. To address this problem, our proposed strategy requires a job-stealing technique in order to minimize the inefficiency from inaccurate estimation. To define a practical job-stealing technique, we divide the entire workloads into different stages. The workloads in each stage will be further divided for the participating clusters according to their computing power. In an ideal case, all computing clusters are supposed to finish their assigned workloads nearly at the same time during the end of each stage. This will fully utilize all computing resources and leads to the minimal execution time. In reality, however, some clusters will finish their assigned workloads before other clusters and request more workloads by entering the next stage immediately. This behavior will result in differences of stage numbers of these clusters which will make clusters complete their execution at different times. Thus, this will lead to underutilization of the computing resources and eventually decrease overall parallel performance. To address this problem, we introduce the stage-warping technique that allows faster-than-expected clusters to execute additional tasks by stealing workloads to be executed by other clusters. Slower-than-expected clusters can then catch up with the other clusters by skipping or warping to the same current stage. With this approach, the completion time of each cluster can be controlled to be finished almost at the same time near the end of each stage despite the dynamic behavior of the underlying resources. We can see this behavior as another form of job-stealing technique since

the faster-than-expected clusters steal workloads which are supposed to be executed by other clusters.

Our proposed strategy uses rough performance estimation, called “consuming rate” ( $cr$ ), which is calculated from the amount of assigned workloads and the interval time between requests. Thus, overall execution aspects including the relationship between the submitted application and the underlying system, in term of performance, can be evaluated through this performance indicator during the execution. Therefore, the problem of misleading parameters like assigning I/O intensive applications to the computing cluster with fast processing speed can be avoided. Moreover, this indicator can be calculated at the coordinator node that is responsible for assigning workloads. Hence, our strategy can make load decisions without implementing the monitoring service within the participating clusters.

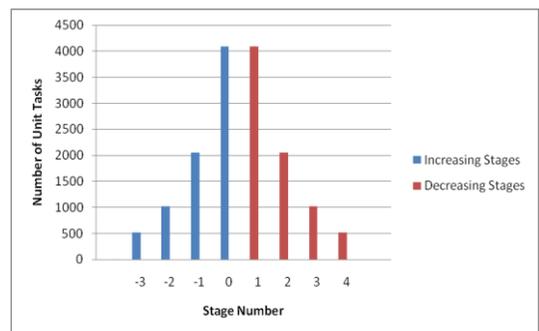
#### 4.1 Stage Size Assignment

In our proposed strategy, application workloads are divided into stages. The amount of workloads assigned to the participating clusters during each stage will be further adjusted with respect to the performance of the requesting cluster and the performance of other clusters. Moreover, we also divide entire computing stages into two groups, which are the increasing stages and the decreasing stages. The purpose of the increasing stages is to obtain accurate performance indicators before entering the decreasing stages. Although we can obtain accurate consuming rate from the beginning of an execution by sending a large chunk of workloads, it is also too risky to assign workloads without having an accurate estimator of the requesting cluster first. To address this problem, our strategy sends small workloads during the first stage and increases the stage size exponentially. After that, the remaining half of the total

workloads will be used to balance the workloads between each cluster. The stage number of every cluster starts from a negative value specified as  $-\left\lceil \log_2 \left( \frac{U}{2} \right) \right\rceil$ .

Then the stage number of each cluster will be gradually increased by one value every time that cluster requests more workloads until it becomes zero, which is the last stage of the increasing stages. After that, the cluster will enter the decreasing stages, and its stage number will still be increased by one value for every request until the end of an execution. An example of stage number sequence from one participating cluster during the entire execution can be given as  $\{-3,-2,-1,0,1,2,3,4\}$ . Equation (1) describes how we allocate workloads during both increasing and decreasing stage  $m$  ( $u_m$ ) given the total number of tasks ( $U$ ) and constant ratio ( $\delta$ ) which we specify as 2 throughout this work. An example of how our strategy allocates workload for each stage is shown in Fig. 2.

$$u_m = \begin{cases} \left\lceil \frac{U}{2 * \delta^{|m|+1}} \right\rceil & \text{if } m \leq 0 \\ \left\lceil \frac{U}{2 * \delta^m} \right\rceil & \text{otherwise} \end{cases} \quad (1)$$



**Figure 2** The number of unit tasks allocated for each stage.

Given the stage size of both increasing and decreasing stages, the global coordinator can determine  $k_{i,m}$ , which is the chunk size to be assigned for cluster  $C_i$

during stage  $m$ . This value will be calculated from the performance indicator of the requesting cluster ( $cr_{i,m}$ ), compared with those of the other clusters, together with the predefined stage size ( $u_m$ ) as shown in Eq. (2). Note that the workloads allocated in the first stage will be assigned equally to all clusters because there is no consuming rate available yet.

$$k_{i,m} = \frac{cr_{i,m}}{\sum_{q=1}^L cr_{q,m-1}} * u_m \quad (2)$$

Keeping all clusters to execute at the same stage for the entire execution period is very important as it prevents a load imbalance, which can lead to poor performance. In an ideal case where the estimator of each cluster is always accurate, tasks will be assigned to all clusters, such that all clusters will enter the same stage, complete all tasks in the predefined stage chunk, and then move to the next stage at the same time throughout the entire execution. In reality, this behavior will never happen. Given the inaccuracy, some clusters will be over-estimated (or slower than expected) while the others will be under-estimated (or faster than expected). Under-estimated clusters will enter the following stage and continue requesting workloads for the new stage while over-estimated clusters are still in the previous stage. This behavior will create additional load imbalance in the system especially when there are some clusters still in the increasing stages at the end of an execution. Moreover, this problem will become worse in a multiple-cluster environment as the underlying resources can be highly heterogeneous.

#### 4.2 Stage-Warping Technique for Grid Environment

In order to keep every cluster to be at the same stage throughout an execution, our strategy introduces a job-stealing technique called “stage-warping”. This

technique allows the clusters whose stage numbers are behind the others to skip (or warp) from their current stages to the foremost stage. The newly derived stage will include the remaining workloads in the previous stages into the foremost stage. With this behavior, the leftover workloads in the previous stages will be re-assigned again to every cluster and the effect of inaccurate estimators, which causes some clusters to stay behind, can be reduced. In other words, the faster-than-expected clusters absorb the load imbalance effects by stealing workloads from those clusters that warp to catch up to the foremost stage. Equation (3) illustrates how we can calculate the new stage size ( $u'_m$ ) with stage-warping technique, by first finding the number of assigned tasks during the previous stages from the remaining tasks at the beginning of the current stage  $m$  ( $\omega'_m$ ), before combining the leftover tasks with the pre-allocated tasks. An example of this behavior can be seen in Fig. 3 where the tasks inside stage -1 and 0 are moved into stage 1 with our stage-warping technique.

$$u'_m = \begin{cases} 2 \left\lceil \frac{U}{2 * \delta^{|m|+1}} \right\rceil + \omega'_m - U & \text{if } m < 1 \\ \left\lceil \frac{U}{2 * \delta^m} \right\rceil + \left( \frac{U}{2} + \omega'_m - U \right) & \text{if } m = 1 \\ \left\lceil \frac{U}{2 * \delta^m} \right\rceil + \left( \frac{U}{2} + \sum_{q=1}^{m-1} \left\lceil \frac{U}{2 * \delta^q} \right\rceil + \omega'_m - U \right) & \text{if } m > 1 \end{cases} \quad (3)$$

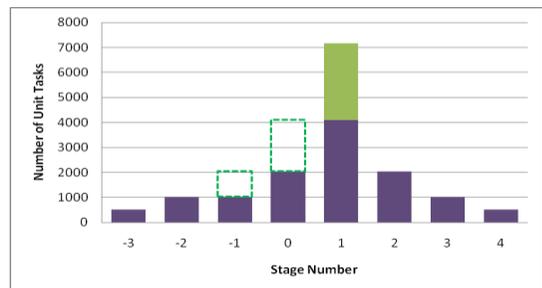


Figure 3 Behavior of stage-warping technique.

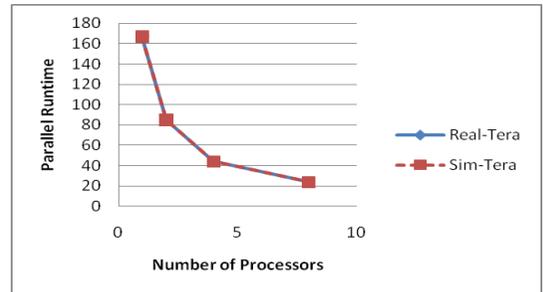
## 5. Performance Evaluation

In this section, we will evaluate the performance of load sharing strategies by simulating a real computing environment with NS2 [18]. We choose two computing clusters in Thaigrd[19], TERA and PLUTO, as representatives of computing clusters. TERA cluster belongs to Kasetsart University while PLUTO cluster is belonged to Chulalongkorn University. We collect the computing power of each cluster by using a simple Matrix-Multiplication program with different matrix sizes. The communication-related parameters are defined based on our preliminary tests and the available specifications. The parameters that we use to create our test environment from the real environment are shown in Table 1. Note that the unit time represents the computation time for one computing node in each cluster to execute only one multiplication. Hence, the entire execution will consist of  $M^2$  multiplications where  $M$  is the matrix size during each run.

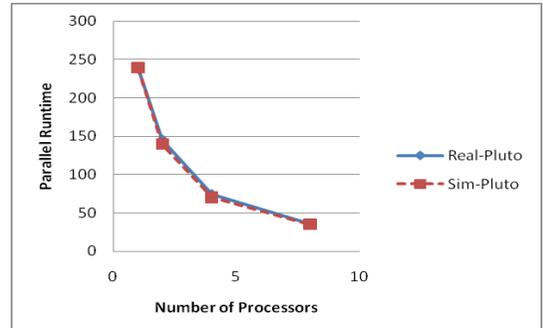
Figure 4 and 5 illustrate the accuracy of our simulation by comparing the parallel runtimes from both real and simulated single cluster environment using SS as load sharing strategy with matrix size specified as 2000 x 2000.

**Table 1** The parameters from real environment

Variables	Values
Unit time (TERA)	20.834ns
Unit time (PLUTO)	29.223ns
LAN Latency	30 $\mu$ s
LAN Bandwidth	1000Mbps
WAN Latency	25ms
WAN Bandwidth	1.5Mbps



**Figure 4** The comparison of parallel runtime over TERA.

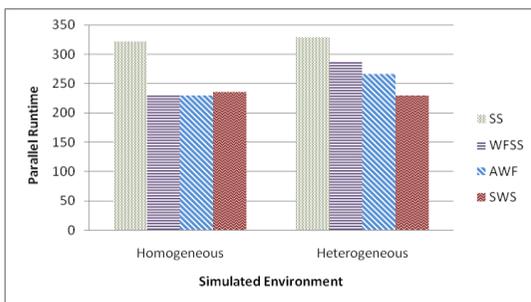


**Figure 5** The comparison of parallel runtime over PLUTO.

To simulate heterogeneous computing environment, we create four clusters where one of them consists of 64 nodes while the other clusters will consist of only 16 nodes. The first cluster will use the computing power collected from TERA cluster and the rest will be from PLUTO cluster. Since this work focuses on the global strategy which assigns workloads among participating clusters, the local strategy within each cluster is defined based on SS. As for the submitted application, the matrix size for the simulated experiments will be specified as 10000 x 10000 and the communication size of each unit task is defined based on data within the Matrix Multiplication program, which is in floating point. Therefore, our submitted application can be considered as a computing-intensive application where the communication overhead will not affect overall performance of load sharing strategies, except SS, which assigns only single task per request for obtaining better performance stability.

## 5.1 Effect of the Computing Heterogeneity

First, we evaluate the effect of computing heterogeneity within an underlying system by comparing the parallel runtimes of load sharing strategies over both homogeneous and heterogeneous systems. We create a homogeneous system within our simulation by assigning every cluster to have the same number of computing nodes with the same computing power. Note that the homogeneous system will have the same total computing power as in the heterogeneous system. We choose SS to represent the simplest form of load sharing strategy that can assign workload without using any performance indicators about an underlying system. WFSS can be considered as an example of explicit strategy that uses only the information obtained from an external source, where AWF will also adjust to the dynamic of the computing system by re-calculating the weighted value throughout an execution. Figure 6 shows the performance of load sharing strategies over computing systems with different computing heterogeneities.



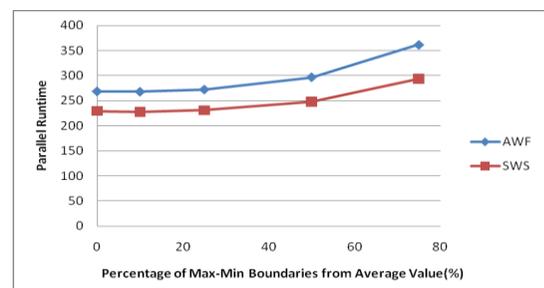
**Figure 6** Parallel runtimes over different computing systems.

Experiment results show that SS has the worst results on both homogeneous and heterogeneous system because of an excessive communication overhead for each request. As for WFSS and AWF, we can see that AWF performs better than WFSS in both cases. This behavior is the result of how AWF can adjust the performance

indicator of each computing resource during an execution. This is the reason why AWF is considered to be one of the best load sharing strategies for grid computing systems. However, while SWS achieves a comparable parallel performance with AWF over a homogeneous system, its parallel performance will be a lot better than AWF over a heterogeneous system. The reason behind this behavior is how our strategy can reduce the load imbalance between the fast and slow clusters better than AWF by allowing under-estimated clusters to steal workload of other over-estimated clusters. Therefore, our strategy can avoid a bad situation when there are some tasks left over in the slow clusters, while other clusters idle near the end of an execution.

## 5.2 Effect of the Dynamic Behavior of the Underlying Resources

Since Grid technology allows local administrators to have control over their resources, the available computing power of Grid resources can change abruptly throughout an execution because the local policy allows other jobs to be executed on the same computing nodes. To simulate this behavior, we specify the computing power of each computing node with normal distribution where the max-min boundaries are specified as a percentage of the average value as mentioned in [20].



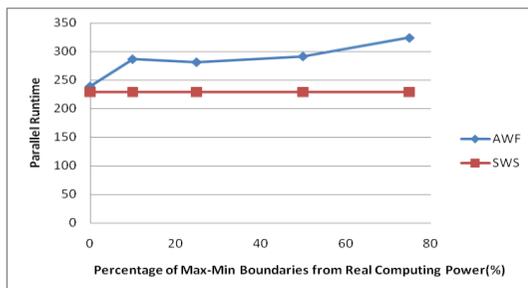
**Figure 7** Parallel runtimes over different fluctuation in available computing power.

We can see from the results in Fig. 7 that while the parallel performance of both strategies decrease when the

fluctuation of the computing power increases, our proposed strategy can achieve better parallel runtime than AWF. This behavior shows that our strategy can tolerate a fluctuation of computing power more than AWF because we define our strategy so that it can address the changes in computing system throughout an entire execution with our stage-warping technique.

### 5.3 Effect of Inaccurate Information

Our strategy can obtain the available computing power by itself. Thus, it does not need an estimated value of the underlying system prior to execution. This behavior not only eliminates the need for implementing the monitoring service over participating clusters, but also makes our strategy independent from information inaccuracy of each resource when we start the execution. On the contrary, most load sharing strategies, such as AWF, rely on the estimated performance of the underlying systems. Thus, their performances are subjected to the accuracy of the estimators. To illustrate this effect, we define a random variable which represents the estimated computing power used by AWF during the first stage. The value of this variable will be selected from a normal distribution, where different max-min boundaries are specified as a percentage of the real computing power. The effect of inaccurate information over the performance of load sharing strategies is shown in Fig. 8.

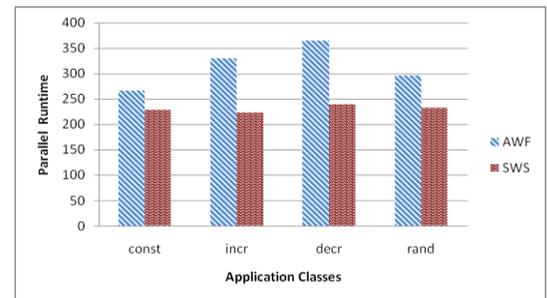


**Figure 8** Parallel runtimes over different fluctuation with information inaccuracy.

The experiment results show that our strategy is not affected by inaccurate information like AWF. This behavior is the results of how we define our strategy to collect and adjust the estimated computing power of an underlying resource by itself without using an estimated value from other sources. We can see that the performance of AWF will decrease greatly when the estimated information is not accurate.

### 5.4 Effect of Application Classes

Different types of applications can also affect the performance of load sharing strategy. The computation size of each task can be uniform, random, increasing, or decreasing throughout the entire execution.



**Figure 9** Parallel runtimes over different application classes.

Figure 9 shows that our strategy can achieve better parallel runtimes than AWF on every application, especially when the computation size of each task decreases during an execution. This is due to the fact that for applications with decreasing computation size, the correct task assignment is very crucial for the first stage, as it includes larger tasks, which contribute the majority of the execution time. Thus, this will greatly impact the system using AWF to distribute workloads, as it assigns half of the entire workloads during the first stage.

## 6. Conclusion

In this work, we propose a robust load sharing strategy for large-scale cluster-based computing systems. By implementing a stage-warping technique, our strategy can make load decisions based on a rough performance estimator which can be collected at the global coordinator during an execution, and steal jobs from slower-than-expected clusters to prevent a load imbalance near the end of the execution. Hence, there is no need to implement a monitoring service or worry about its accuracy at all. From our experiments, we show that our proposed strategy can address computing heterogeneity, fluctuation in available computing power, and different classes of submitted applications. Therefore, our proposed global strategy is simple yet efficient in a large scale computing environment which definitely will serve as the computing platform for the next generation.

## 7. Reference

- [1] Foster, I. et al., The Anatomy of the Grid, *Int. J. Supercomput. Appl. High Perform. Comput.*, Vol. 15, pp. 200-222, 2001.
- [2] Kruskal, C. P. and Weiss, A., Allocating Independent Subtasks on Parallel Processors, *IEEE Trans. Software Eng.*, Vol. 11, pp. 1001-1016, 1985.
- [3] Balasubramanian, J. et al., Evaluating the Performance of Middleware Load Balancing Strategies, *EDOC*, pp. 135-146, 2004.
- [4] Shen, K. et al., Cluster Load Balancing for Fine-grain Network Services, *IPDPS*, pp. 51-58, 2002.
- [5] Chronopoulos, A. T. et al., Scalable Loop Self-scheduling Schemes for Heterogeneous Clusters, *Int. J. Comput. Eng. Sci.*, pp. 353-359, 2002.
- [6] Shih, W. C. et al., A Performance-based Parallel Loop Self-scheduling on Grid Environments, *Lect. Notes Comput. Sci.*, Vol. 3779, pp. 48-55, 2005.
- [7] Biswas, R. et al., Tools and Techniques for Measuring and Improving Grid Performance, *Lect. Notes Comput. Sci.*, Vol. 2571, pp. 45-54, 2002.
- [8] Wolski, R., Experiences with Predicting Resource Performance On-line in Computational Grid Settings, *Perform. Eval. Rev.*, Vol. 30, pp. 41-49, 2003.
- [9] Lee, Y. C. and Zomaya, A. Y., Practical Scheduling of Bag-of-Tasks Applications on Grids with Dynamic Resilience, *IEEE Trans. Comput.*, Vol. 56, pp. 815-825, 2007.
- [10] Buyya, R. et al., Scheduling Parameter Sweep Applications on Global Grids: a Deadline and Budget Constrained Cost-time Optimization Algorithm, *Software Pract. Ex.*, Vol. 35, pp. 491-512, 2005.
- [11] Fann, Y. W. et al., IPLS: An Intelligent Parallel Loop Scheduling for Multiprocessor Systems, *ICPADS*, pp. 775-782, 1998.
- [12] Tang, P. and Yew, P. C., Processor Self-scheduling for Multiple-nested Parallel Loops, *ICPP*, pp. 528-535, 1986.
- [13] Polychronopoulos, C. D. and Kuck, D. J., Guided Self-scheduling: A Practical Scheduling Scheme for Parallel Supercomputers, *IEEE Trans. Comput.*, Vol. 36, pp. 1425-1439, 1987.
- [14] Tzen, T. H. and Ni, L. M., Trapezoid Self-scheduling: A Practical Scheduling Scheme for Parallel Compilers, *TPDS*, Vol. 4, pp. 87-98, 1993.
- [15] Hummel, S. F. et al., Factoring: A Method for Scheduling Parallel Loops, *Comm. ACM*, Vol. 35, 1992.

- [16] Hummel, S. F. et al., Load-sharing in Heterogeneous Systems via Weighted factoring, SPAA, pp. 318-328, 1996.
- [17] Banicescu, I. and Velusamy, V., Performance of Scheduling Scientific Applications with Adaptive Weighted Factoring, IPDPS, pp. 84, 2001.
- [18] McCanne, S. and Floyd, S., VINT Network Simulator - ns (version 2), <http://www.mash.CS.Berkeley.EDU/> ns/.
- [19] Varavithya, V. and Uthayopas, P., ThaiGrid: Architecture and Overview, NECTEC Tech. J., Vol. 2, 2000.
- [20] Casanova, H., Simgrid: A Toolkit for the Simulation of Application Scheduling, CCGrid, pp. 430-441, 2001.

# The Clearness Index Model for Estimation of Global Solar Radiation in Thailand

**Jompob Waewsak and Chana Chancham**

Solar and Wind Energy Research Laboratory (SWERL)

Renewable Energy System Research and Demonstration Center (RERD)

Department of Physics, Faculty of Science, Thaksin University,

Phatthalung, 93110, THAILAND

E-mail: jompob@tsu.ac.th

## Abstract

The aim of this paper is to present a novel model to estimate monthly global solar radiation in Thailand. The novel model was investigated based upon the regression technique. Meteorological data sets obtained from the ground-based measurement between 1995-2004 at Khon Kaen, Ubon Ratchathani, Nakhon Sawan, Bangkok and Songkhla were used in modeling. The monthly clearness index and normalized monthly sunshine duration were correlated in order to search the regression coefficients using 5 different models i.e. linear, quadratic, linear-logarithmic, logarithmic, and power equations. The MBE, MPE and RMSE were used as the main criteria for selecting the best model for estimation of global solar radiation. Results showed that, the linear equation and power equation were the best fit for Nakhon Sawan whereas the quadratic equation was the best fit for Ubon Ratchathani and Bangkok, while the linear-logarithmic was the best fit for Khon Kaen and Songkhla. Finally, the quadratic equation performed the best fit for all regions compared to other models, as the RMSE was minimum.

**Keywords:** Clearness Index, Regression, Solar Radiation, Sunshine Duration

## 1. Introduction

Global solar radiation is the key data for design, development, application and assessment of outdoor solar energy systems. Generally, this data can be obtained accurately from routine ground-based measurement over a particular area. However, the limitation of the measurement is both cost and time consuming. Another means is the use of a model to predict the global solar radiation. Presently several techniques have been presented in modeling solar radiation e.g. forecasting [1] and ANN based modeling [2-3]. However, classical

techniques such as regression are still effective. Many results using this technique are presented [4-7]. The model developed by this technique is very useful for simulation and mathematical modeling of such solar energy systems. At present the Meteorological Department has set up solar radiation measurement at 9 main meteorological stations. During the past few years, the Department of Alternative Energy Development and Efficiency (DEDE) has set up a network for measuring global solar radiation. Twenty five stations were set up and installed throughout the country. Results from the collected data by this

network are very useful in the prediction of solar radiation in Thailand [8]. It seems to be sufficient, but the number of stations is still limited. To avoid the spatial distribution problem, many researchers have proposed several mathematical models to estimate the global solar radiation from ground-based measurement. The extensive investigations have been conducted and presented [9-12]. Among these models, the sunshine duration was best parameter and is often used to correlate to the global solar radiation.

In Thailand, the estimation of global solar radiation from the number of sunshine hours has been investigated by Kirtikara and Siriprayuk [4]. Most of these models used data sets obtained during the last two decades. Recently, Janjai and Tohsing have proposed a new model for estimating the global solar radiation from sunshine duration in Thailand [10]. In this paper the data set of global solar radiation and sunshine duration between 1995-2004 was used in the analysis for modeling.

**2. Data Preparation and Modeling**

Prior to analyzing and correlating the clearness index and normalized sunshine duration, the extraterrestrial radiation was firstly computed [13]. In this calculation, the data of location, i.e., latitude and longitude are required. Table 1 shows the geographical locations of five meteorological stations in Thailand. The clearness index, which is the ratio of global solar radiation and extraterrestrial radiation, was then correlated to the normalized sunshine duration (the ratio between the number of sunshine hours and length of the day) based on the regression technique. Five different models as shown in Eq. 1-Eq. 5 were applied to search for the best model. In order to select the best model, the mean bias error (MBE) as shown in Eq. 6, the average absolute percentage error (MPE) as shown in Eq. 7, and the root mean square error (RMSE) as given in Eq. 8., were used as the

main criteria. Not only that, the coefficient of determination ( $r^2$ ) was also used. Finally, contour maps of monthly mean global solar radiation of Thailand were created using Surfer computer software in order to display the spatial distribution of global solar radiation over Thailand. An interval of 0.2 MJ/m<sup>2</sup>/day was used to create the contour map.

**Table 1** Geographical coordinate of 5 meteorological stations.

Station	Symbol	Latitude (°N)	Longitude (°E)
Khon Kaen	KK	16.28	102.47
Ubon Ratchathani	UB	15.25	104.87
Nakhon Sawan	NSW	15.80	100.02
Bangkok	BK	13.73	100.57
Songkhla	SK	7.20	100.60

Linear Model  $K_T = a + bR_s$  (1)

Quadratic  $K_T = a + bR_s + cR_s^2$  (2)

Linear-Logarithmic  $K_T = a + bR_s + c \log R_s$  (3)

Logarithmic  $K_T = a + b \log R_s$  (4)

Power  $K_T = e^a R_s^b$  (5)

$MBE = \frac{1}{n} \sum_{i=1}^n (H_i - \hat{H}_i)$  (6)

$MPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{H_i - \hat{H}_i}{H_i} \right) \times 100$  (7)

$RMSE = \sqrt{\sum_{i=1}^n (H_i - \hat{H}_i)^2 / (n - k)}$  (8)

where  $K_T$  is the monthly clearness index

$R_s$  is the normalized monthly sunshine hours

$MBE$  is the mean bias error

$MPE$  is the average absolute percentage error

$RMSE$  is the root mean square error

a,b,c is the regression coefficient

- $H_i$  is the measured global solar radiation
- $\hat{H}_i$  is the predicted global solar radiation
- $n$  is the number of valid data points
- $k$  is the number of parameters estimated in the model.

The goodness of fit was judged by the size of the coefficient of determination. The *MBE*, *MPE*, and *RMSE* were computed as a further check on the stability of the models [12].

### 3. Results and Discussion

The regression coefficients are shown in Table 2, whereas the errors of each model are displayed in Table 3. Results show that, the linear equation and power equation are the best fit for Nakhon Sawan, whereas the quadratic equation is the best fit for Ubon Ratchathani and Bangkok, while the linear-logarithmic is the best fit for Khon Kaen and Songkhla. Furthermore, the quadratic equation performed the best fit for all regions compared to other models, as the RMSE is at a minimum. Table 2 shows that the  $r^2$  is in the range of 0.819-0.981. Table 3 shows that the MBE is in the range of -0.06-0.47, the MPE is in the range of 5.12-5.49 and the RMSE is in the range of 0.47-1.59. The estimated monthly global solar radiation using the quadratic model for Thailand is shown in Fig. 1-Fig. 12. It is found that the monthly intensity of incident global solar radiation over Thailand is in the range of 14.25-23.76 MJ/m<sup>2</sup>/day.

**Table 2** Regression coefficients of 5 different models for each station.

Station	Model	a	b	c	r <sup>2</sup>
KK	1	0.32	0.45	-	0.98
	2	0.29	0.56	-0.10	0.98
	3	0.34	0.42	0.01	0.98
	4	0.71	0.23	-	0.98
	5	-0.31	0.42	-	0.98
UB	1	0.26	0.49	-	0.88
	2	0.55	-0.66	1.07	0.91
	3	-1.97	0.91	-1.13	0.92
	4	0.69	0.24	-	0.85
	5	-0.32	0.49	-	0.87
NKW	1	0.39	0.27	-	0.83
	2	0.55	-0.66	1.07	0.91
	3	0.29	0.38	-0.06	0.83
	4	0.63	0.14	-	0.82
	5	-0.45	0.27	-	0.82
BK	1	-0.25	0.51	-	0.91
	2	0.43	-0.21	0.71	0.92
	3	-0.26	1.10	-0.29	0.92
	4	0.69	0.24	-	0.89
	5	-0.32	0.49	-	0.90
SK	1	0.25	0.53	-	0.89
	2	-0.05	1.63	-0.96	0.90
	3	1.22	-0.55	0.6	0.90
	4	0.73	0.30	-	0.89
	5	-0.28	0.54	-	0.89

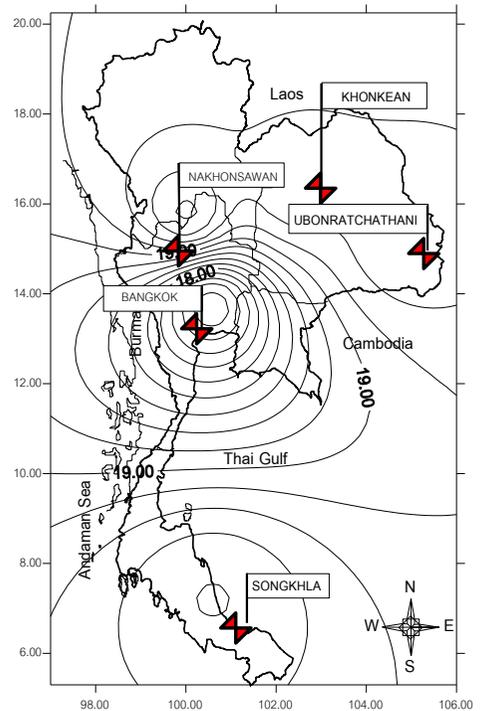
### 4. Conclusion

This paper investigated a novel model for predicting global solar radiation in Thailand. The meteorological data set between 1995-2004 was analyzed using a regression technique in order to correlate the clearness index and normalized sunshine duration using 5 different models. Results indicate that, the linear and power models are suitable for estimating global solar radiation for Nakhon Sawan, whereas the quadratic model is suitable for Ubon Ratchathani and Bangkok while the linear-

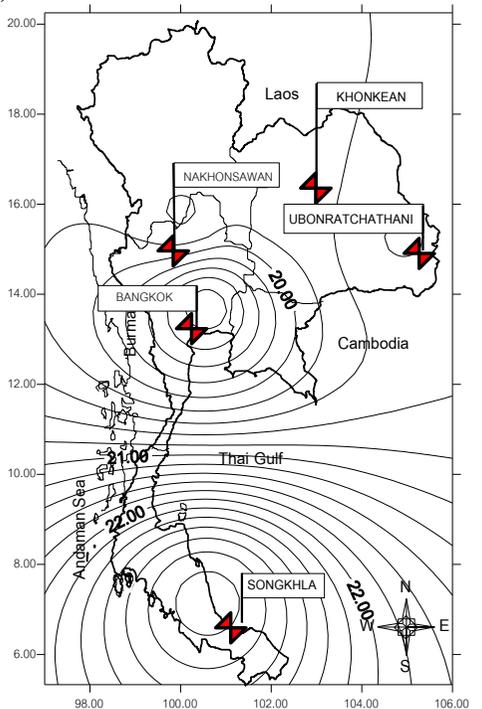
**Table 3** The *MBE*, *MPE* and *RMSE* of 5 different models.

Station	Model	<i>MBE</i>	<i>MPE</i>	<i>RMSE</i>
KK	1	0.05	5.13	0.47
	2	-0.01	5.13	0.49
	3	-0.01	5.13	0.49
	4	-0.02	5.12	0.53
	5	-0.02	5.12	0.48
UB	1	-0.04	5.49	1.39
	2	-0.01	5.49	1.22
	3	-0.01	5.49	1.22
	4	-0.06	5.50	1.59
	5	-0.04	5.49	1.49
NKW	1	-0.05	5.33	0.85
	2	0.47	5.21	1.82
	3	-0.05	5.33	0.89
	4	-0.06	5.33	0.89
	5	-0.05	5.33	0.88
BK	1	-0.05	5.49	1.14
	2	-0.05	5.49	1.07
	3	-0.05	5.49	1.08
	4	-0.06	5.49	1.26
	5	-0.05	5.49	1.20
SK	1	0.02	5.33	1.35
	2	0.01	5.32	1.29
	3	0.01	5.32	1.28
	4	0.02	5.32	1.30
	5	0.02	5.32	1.33

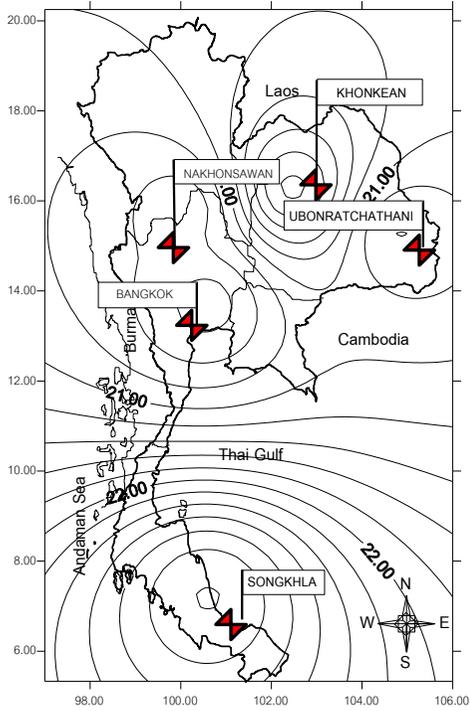
logarithmic model is suitable for Khon Kaen and Songkhla. Finally, the quadratic model has the best results for all regions of Thailand. Using the quadratic model, the estimated global solar radiation in Thailand is in the range of 14.25-23.76 MJ/m<sup>2</sup>/day. However, these maps are simply indicative of the distribution of solar radiation and can be of limited use for finding out the utilization potential of solar energy.



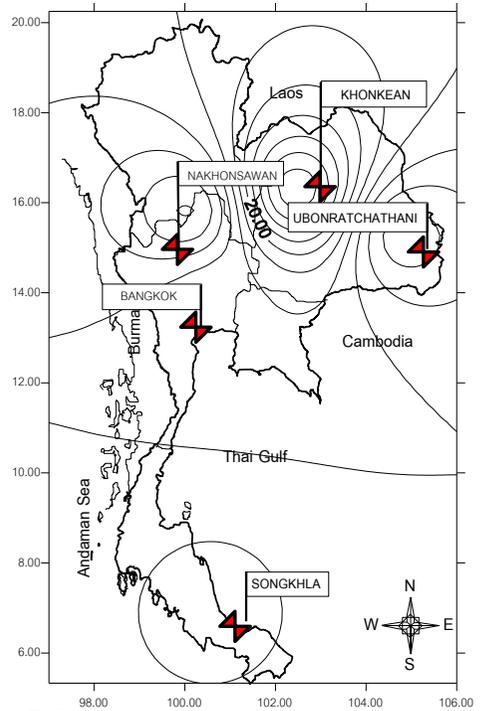
**Fig. 1** The contour map of monthly mean global solar radiation in January (MJ/m<sup>2</sup>/day).



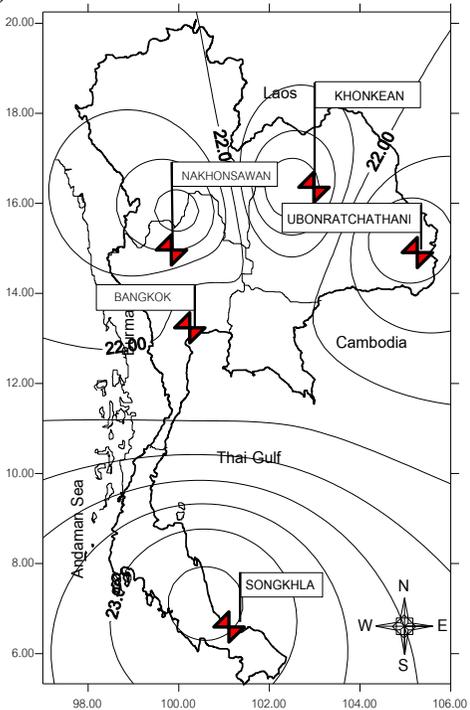
**Fig. 2** The contour map of monthly mean global solar radiation in February (MJ/m<sup>2</sup>/day).



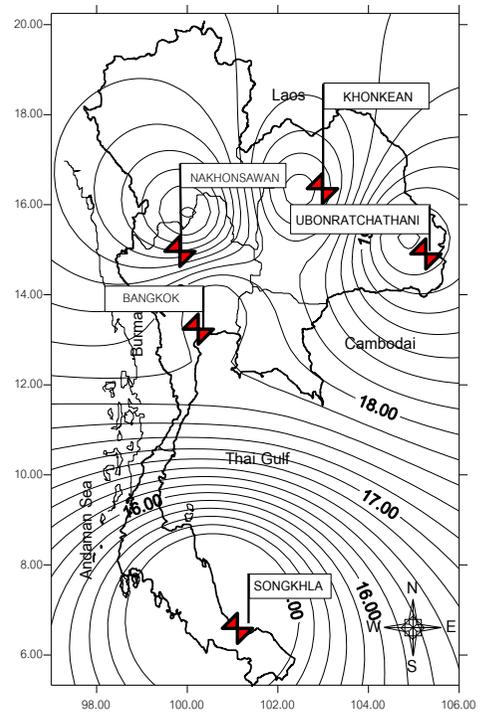
**Fig. 3** The contour map of monthly mean global solar radiation in March ( $\text{MJ}/\text{m}^2/\text{day}$ ).



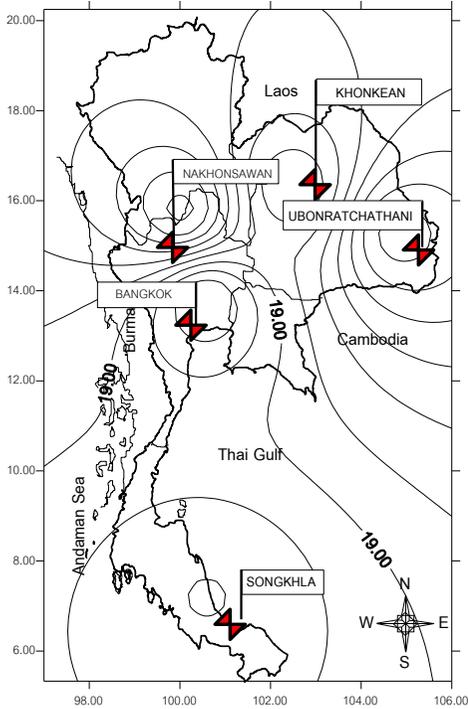
**Fig. 5** The contour map of monthly mean global solar radiation in May ( $\text{MJ}/\text{m}^2/\text{day}$ ).



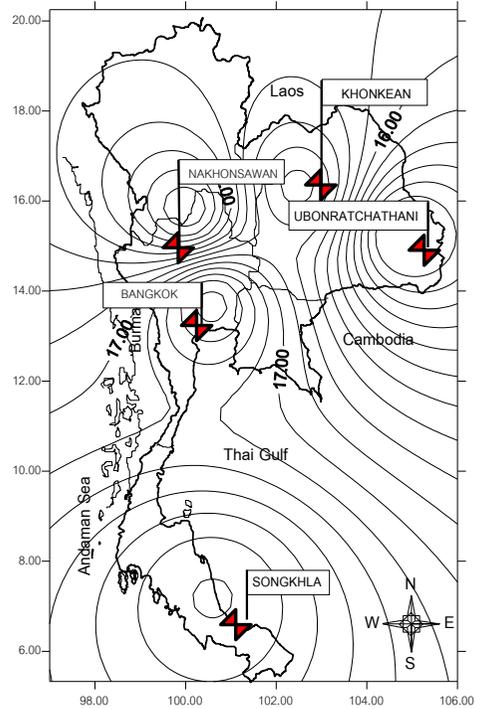
**Fig. 4** The contour map of monthly mean global solar radiation in April ( $\text{MJ}/\text{m}^2/\text{day}$ ).



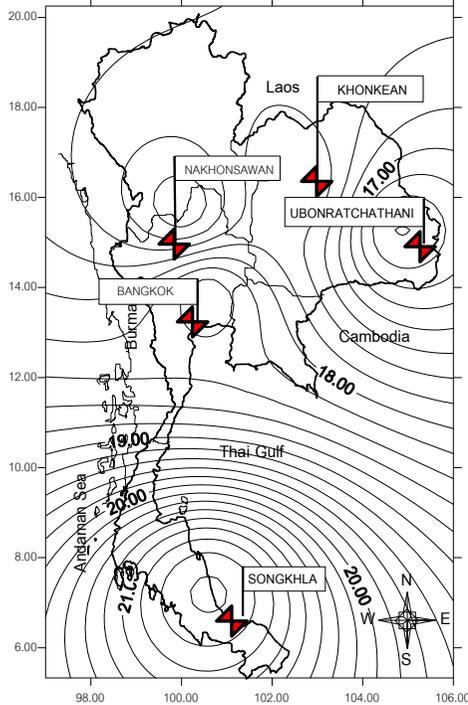
**Fig. 6** The contour map of monthly mean global solar radiation in June ( $\text{MJ}/\text{m}^2/\text{day}$ ).



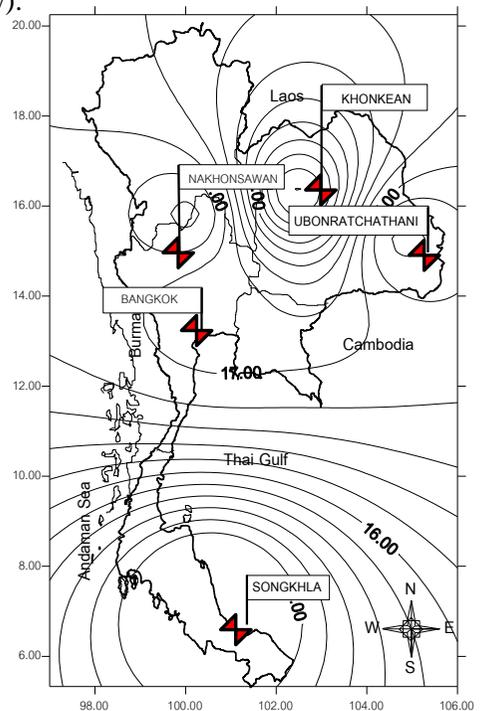
**Fig. 7** The contour map of monthly mean global solar radiation in July ( $\text{MJ}/\text{m}^2/\text{day}$ ).



**Fig. 9** The contour map of monthly mean global solar radiation in September ( $\text{MJ}/\text{m}^2/\text{day}$ ).



**Fig. 8** The contour map of monthly mean global solar radiation in August ( $\text{MJ}/\text{m}^2/\text{day}$ ).



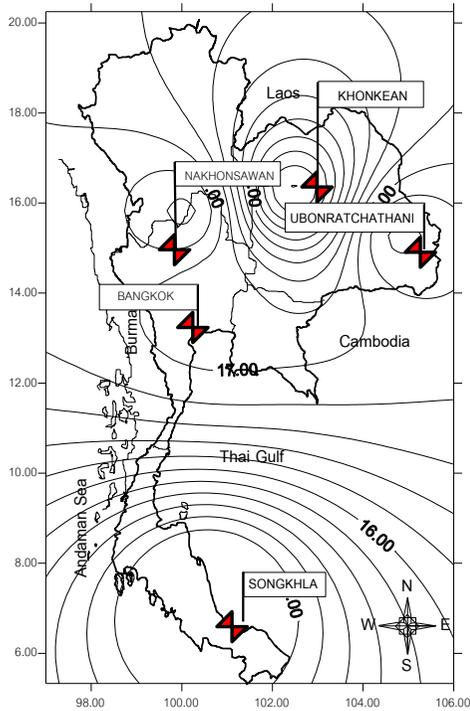
**Fig. 10** The contour map of monthly mean global solar radiation in October ( $\text{MJ}/\text{m}^2/\text{day}$ ).

### 5. Acknowledgement

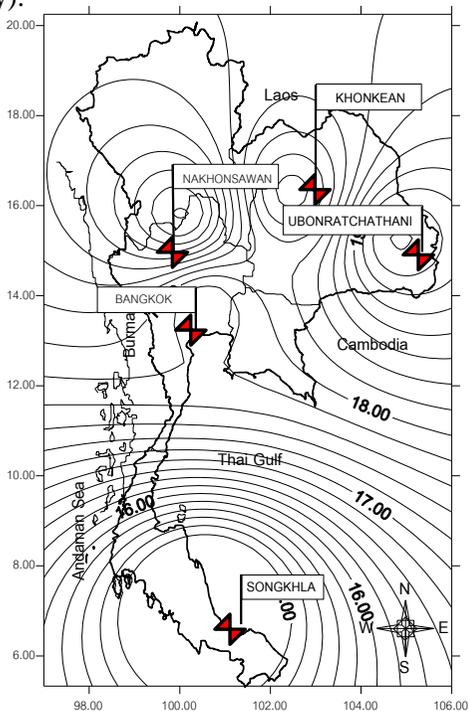
The authors gratefully acknowledge the Thai Meteorological Department for providing the data set of the global solar radiation and sunshine duration in this analysis. The authors would like to thank the Solar and Wind Energy Research Unit, Thaksin University for financial support.

### 6. References

- [1] Reikard, G. Predicting Solar Radiation at High Resolution: A Comparison of Time Series forecasts, *Solar Energy*, Vol. 83, pp. 342-349, 2009.
- [2] Benghamen, M., Mellit, A. and Alamri, S. N., ANN-Based Modeling and Estimation of Daily Global Solar Radiation Data: A Case Study, *Energy Conversion and Management*, Vol. 50, pp. 1644-1655, 2009.
- [3] Azadeh, A., Maghsoudi, A., and Sorabkhani, An Integrated Artificial Neural Networks Approach for Predicting Global Solar Radiation, *Energy Conversion and Management*, Vol.50, pp. 1497-1505, 2009.
- [4] Kirtikara, K. and Siriprayuk, T., 1980, "Relationships between Solar Radiation and Some Meteorological Data of Thailand, Proc. Symposium of Solar Science and Technology, 3-5 Nov, Bangkok, Thailand, pp.481-484, 1980.
- [5] Exell, R. H. B. and Santibuppakul, P., New Estimates of Mean Daily Diffuse Solar Radiation in Thailand, *Journal of the Science Society of Thailand* 10, pp. 37-41,1984.
- [6] Janjai, S., Praditwong, P. and Keawkiew, J, Mathematical Models for Computing Global and Diffuse Solar Radiation for Thailand, Proc. The Second Asean Renewable Energy Conference, 5-7 Nov, Phuket, Thailand, pp. 472-478, 1997.



**Fig. 11** The contour map of monthly mean global solar radiation in November ( $\text{MJ}/\text{m}^2/\text{day}$ ).



**Fig. 12.** The contour map of monthly mean global solar radiation in December ( $\text{MJ}/\text{m}^2/\text{day}$ ).

- [7] Chirarattananon, S., Chaiwiwatworakul, P. and Pattanasethanon, S., Daylight Availability and Models for Global and Diffuse Horizontal Illuminance and Irradiance for Bangkok, Proc. The First Regional Conference on Energy Technology Towards a Clean Environment, 1-2 Dec, Chiang Mai, Thailand, pp. 31-36, 2000.
- [8] Janjai, S., Pankaew, P. and Laksanaboonsong, J., A Model for Calculating Hourly Global Solar Radiation from Satellite Data in the Tropic, Applied Energy 86, pp. 1450-1457, 2009.
- [9] Hirunlabh, J., Santisirisomboon, J. and Namprakai, P., Assessment of Solar Radiation for Thailand, International Workshop: Calculation Methods for Solar Energy System, COMPLES, pp.29-30 September, University de Perpignan, France, p.14, 1994.
- [10] Janjai, S. and Tohsing, K., A Model for the Estimation of Global Solar Radiation from Sunshine Duration for Thailand, The Regional Conference on Energy and Environment, Phuket, Thailand, 2004.
- [11] Phethuayluk, S., Waewsak, J. and Kaew-On, J., Estimation and Contour Mapping of Global Solar Radiation under Tropical Climate of Southern Thailand, The Solar World Congress, 6-12 August, Florida, U.S.A., 2005.
- [12] Ampratwum, D. B. and Dorvlo, A. S. S., Estimation of Solar Radiation from the Number of Sunshine Hours, Applied Energy 63, pp.161-167, 1999.
- [13] Duffie, J. A. and Beckman, W. A., Solar Engineering of Thermal Processes, New York, John Wiley & Sons, 1991.

# Effects of Deposition Period on the Properties of FeS<sub>2</sub> Thin Films by Chemical Bath Deposition Method

**K. Anuar\***, W.T. Tan, Md. Jelas, S.M. Ho and S.Y. Gwee

Department of Chemistry, Faculty of Science, Universiti Putra Malaysia,  
43400 Serdang, Selangor, Malaysia.

\*Corresponding author Email: anuar@science.upm.edu.my

**N. Saravanan**

Department of Bioscience and Chemistry, Faculty of Engineering and Science,  
Universiti Tunku Abdul Rahman, 53300 Kuala Lumpur, Malaysia.

## Abstract

Iron sulfide thin films were deposited onto microscope glass substrates by chemical bath deposition technique from aqueous solutions containing iron nitrate and sodium thiosulfate. The effects of deposition time toward the properties of the thin films were investigated. The deposited thin films were characterized with X-ray diffraction, scanning electron microscopy and UV-Vis Spectrophotometer. The X-ray diffraction patterns indicate that the films have orthorhombic structure with better crystalline phase for the films deposited for 90 min. According to SEM micrographs, the sizes of the grains were noticed to increase as the deposition time was increased to 90 min. The values of band gap energy, calculated from the absorption spectra, decreased from 2.65 to 1.85 eV as the deposition time was increased from 30 to 90 min. Therefore, deposition for 90 min was suitable to produce good quality FeS<sub>2</sub> thin films.

**Keywords:** Iron sulfide, X-ray diffraction, Chemical bath deposition, Thin films

## 1. Introduction

Recently, much attention has been focused to study properties of various metal chalcogenide thin films. This is due to their possible application to the manufacture of large area photodiode arrays, solar selective coatings, solar cells [1-4] and sensors. The main advantage of chalcogenide semi-conductors is their promise of lower costs, since less energy for processing and relatively lower costs for the materials are required, and large scale production is feasible.

FeS<sub>2</sub> thin film has attracted considerable attention in recent years, as it is composed of very abundant, cheap, nontoxic elements and possesses semi-conducting properties. Various techniques have been used to prepare FeS<sub>2</sub> thin films. These include flash evaporation [5], metal organic chemical vapour deposition [6], sputtering [7], chemical vapour transport [8], electrodeposition [9] and molecular beam deposition [10].

However, to our knowledge, a study on the properties of the chemical bath deposited iron sulfide thin film has not been

reported so far. In the present investigation, we report the growth and characterization of FeS<sub>2</sub> thin films deposited using this method. Chemical bath deposition as a method for thin film preparation is a good approach with respect to economic considerations. An important advantage of this method is that thin films with a large area can be prepared without vacuum, using simple and low-cost equipment. The chemical bath deposition method has been proved as a suitable method of preparing chalcogenide thin films such as CdS [11], ZnS [12], CdSe [13], PbS [14], Cd<sub>1-x</sub>Zn<sub>x</sub>Se [15], Cd<sub>0.5</sub>Zn<sub>0.5</sub>Se [16] and CuInSe<sub>2</sub> [17].

In this work, we prepare iron sulfide thin films using the chemical bath deposition technique from aqueous medium. We investigate the influence of the deposition time (30, 60 and 90 min) on structural, morphological and optical properties of thin films. The thin film has been characterized by X-ray diffraction (XRD) for structure determination, scanning electron microscopy (SEM) analysis for surface morphology study and optical properties for band gap calculation.

## 2. Materials and Methods

### 2.1 Sample preparation

All the chemicals used for the deposition were analytical grade. They include iron nitrate (Fe(NO<sub>3</sub>)<sub>3</sub>·9H<sub>2</sub>O), sodium thiosulfate (Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>·5H<sub>2</sub>O), triethanolamine [(HOC<sub>2</sub>H<sub>4</sub>)<sub>3</sub>N] and hydrochloric acid [HCl]. All the solutions were prepared in deionised water (Alpha-Q Millipore). During the deposition, the triethanolamine (TEA) is used as a complexing agent while iron nitrate and sodium thiosulfate act as a source of iron and sulfide ion, respectively.

The microscope glass substrates of the dimension 25.4 mm x 76.2 mm x 1 mm were degreased in ethanol for 10 min and then ultrasonically cleaned with distilled water for another 15 min before deposition

of films. Deposition of FeS<sub>2</sub> thin film was carried out by using the following procedure: 20 ml of 0.15 M iron nitrate solution was put in a 100 mL beaker. Then, 10 mL of triethanolamine solution was added. Following that, 20 mL of 0.15 M sodium thiosulfate solution was added into the above mentioned mixture. By adding drop-wise hydrochloric acid solution, the pH of resultant solution was adjusted to 2.5. The beaker containing resultant solution was kept in the water bath. The temperature of the bath was then allowed to increase slowly up to 80 °C. The clean microscope glass slides were placed in the bath, supported vertically on the wall of the beaker. In order to determine the best conditions for the deposition process, the films were deposited at different deposition periods, ranging from 30 to 90 min. During the deposition period the beaker was kept undisturbed. After completion of film deposition, the glass slide was removed, washed several times with distilled water and dried naturally in desiccators for further characterization.

### 2.2 Characterization method

In order to investigate the crystallographic properties of the FeS<sub>2</sub> thin films, we carried out the X-ray diffraction analysis using a Philips PM 11730 diffractometer with CuK<sub>α</sub> (λ=1.5418 Å) radiation. The surface morphology was observed by scanning electron microscopy (JEOL, JSM-6400). The optical absorption measurement was carried out in the wavelength range from 350 to 800 nm by using a Perkin Elmer UV/Vis Lambda 20 Spectrophotometer. The film-coated glass substrate was placed across the sample radiation pathway while the uncoated glass substrate was put across the reference path. The absorption data were studied for the determination of the band gap energy.

### 3. Results and discussion

Figure 1 is the X-ray diffractogram of the iron sulfide thin films deposited at different deposition periods. The film deposited for 30 min shows only a single peak at  $2\theta = 33.8^\circ$  corresponds to interplanar distances of  $2.65 \text{ \AA}$  (Fig. 1a). As the deposition time was increased to 60 min, three prominent diffraction peaks were observed at  $2\theta$  values =  $31.9^\circ$ ,  $33.5^\circ$  and  $47.5^\circ$  corresponding to (011), (101) and (121) planes. Comparison between the films deposited for 60 and 90 min revealed that the number of peaks increased to six peaks, indicating a better crystalline phase for the film deposited in longer time (Fig. 1c). In addition, the preferred orientation is changed from (011) to (101) plane at this period. The peaks were identified by comparing the  $d$ -spacing values obtained from the XRD patterns with Joint Committee on Powder Diffraction Standards (JCPDS) data for  $\text{FeS}_2$  (reference No.: 00-024-0074). The lattice parameters of the orthorhombic structure are equal to  $a = 4.436 \text{ \AA}$ ,  $b = 5.414 \text{ \AA}$  and  $c = 3.381 \text{ \AA}$ . The peaks observed at  $2\theta$  values of  $37.5^\circ$  ( $d$ -spacing =  $2.38 \text{ \AA}$ ) and  $41.3^\circ$  ( $d$ -spacing =  $2.17 \text{ \AA}$ ) are caused by the glass substrate. The data indicated the formation of orthorhombic phase of  $\text{SiO}_2$  (JCPDS reference No.: 01-074-0201).

Scanning electron microscopy (SEM) is a useful technique analysis of the surface morphology of a film. It can give the grain size and structure of samples. The examination of SEM micrographs from Figure 2 brings a comparison of the surface grains from the deposited films under different deposition periods. All the samples were examined at 20 kV with 250 X magnification.

The morphology study shows that the  $\text{FeS}_2$  thin films deposited for 30 min are not compact and do not have good coverage of glass substrate (Fig. 2a). These films have smaller grains compared to the other

films. This observation is supported by data obtained from XRD pattern. This can be explained by the broadening of the peaks, which can be due to the small size of the crystals. The sizes of the grains were noticed to increase as the deposition time was increased to 90 min (Fig. 2c). The films are composed of largely irregular-shaped grains of diameter 100-150  $\mu\text{m}$ .

The optical properties of  $\text{FeS}_2$  thin films are determined from absorbance measurement in the range of 350 to 800 nm. Figure 3 shows absorbance spectra for  $\text{FeS}_2$  thin films deposited at different deposition periods. All the films show a gradually increasing absorbance throughout the visible region, which makes it possible for this material to be used in a photo-electrochemical cell. As it can be seen, the thin films deposited for 90 min (Fig. 3c) display high absorption characteristics in the visible region when compared with the other films. This confirmed that the thicker film has higher absorption properties. This response also is associated with the fact that more  $\text{FeS}_2$  materials are formed at this period.

Band gap energy and transition type can be derived from mathematical treatment of data obtained from optical absorbance versus wavelength with the Stern relationship of near-edge absorption (Equation 1):

$$A = \frac{[k(h\nu - E_g)^{n/2}]}{h\nu} \quad (1)$$

where  $\nu$  is the frequency,  $h$  is the Planck's constant,  $k$  equals a constant while  $n$  carries the value of either 1 or 4. The value of  $n$  is 1 and 4 for the direct transition and indirect transition, respectively. The plots of  $(Ah\nu)^2$  versus  $h\nu$  are shown in Figure 4. The linear nature of the plots indicated the existence of direct transitions. The band gap energy is obtained by extrapolating the linear portion of  $(Ah\nu)^{2/n}$  versus  $h\nu$  to the energy axis at  $(Ah\nu)^{2/n} = 0$ . The results reveal that the band gap energy decreases linearly from 2.65 to

2.60 and 1.85 eV when the deposition time was increased from 30 to 60 and 90 min, respectively. Clearly, the grain size increases gradually with deposition time. This statement is also supported by the data obtained from XRD and SEM analysis.

#### 4. Conclusions

FeS<sub>2</sub> thin films can be chemically deposited by using iron nitrate and sodium thiosulfate solutions. The triethanolamine solution was used as a complexing agent during the deposition process. The film deposited for 90 min showed larger grain size and more peaks attributed to FeS<sub>2</sub>, as could be observed from SEM and XRD results, respectively. The band gap energy was decreased from 2.65 to 1.85 eV as the deposition time was increased from 30 to 90 min. Therefore, deposition for 90 min was suitable to produce good quality of FeS<sub>2</sub> thin films.

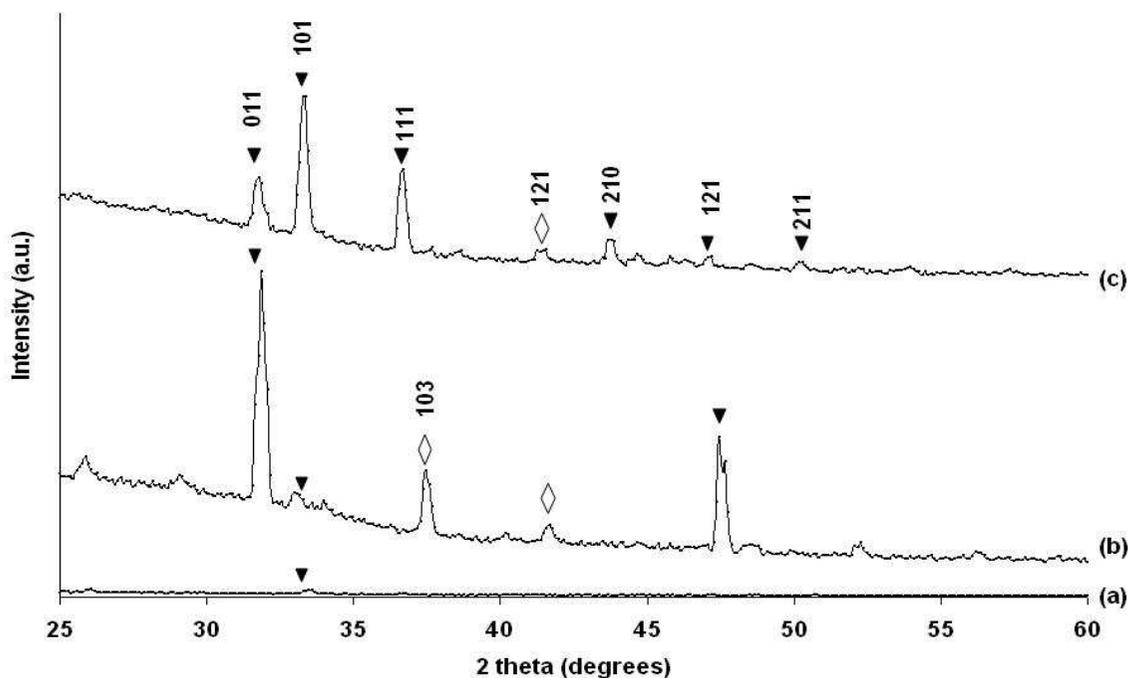
#### 5. Acknowledgements

The authors would like to thank the Department of Chemistry, University Putra Malaysia (UPM) for the provision of laboratory facilities.

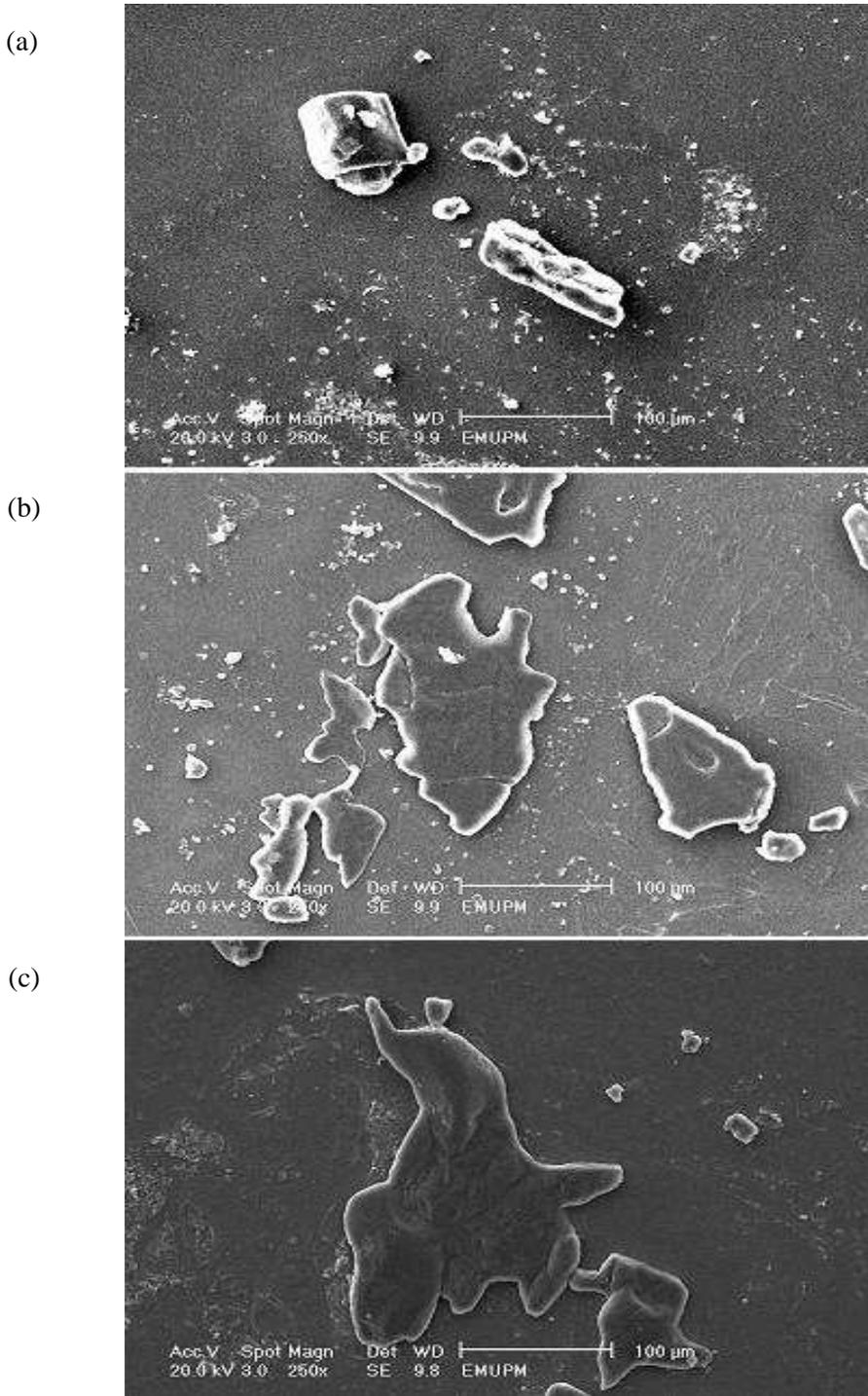
#### 6. References

- [1] Anuar, K., Ho, S.M., Tan, W.T., Atan, S., Zulkefly, K., Md. Jelas, H. and Saravanan, N., Cathodic Electrodeposition of Chalcogenide Thin Films Cu<sub>4</sub>SnS<sub>4</sub> for Solar Cells, *Chiang Mai University Journal of Natural Sciences*, Vol. 7, pp. 317-326, 2008.
- [2] Messina, S., Nair, M.T.S. and Nair, P.K., Antimony Sulfide Thin Films in Chemically Deposited Thin Film Photovoltaic Cells, *Thin Solid Films*, Vol. 515, pp. 5777-5782, 2007.
- [3] Ristov, M., Sinadinovski, G., Mitreski, M. and Ristova, M., Photovoltaic Cells Based on Chemically Deposited P-type SnS, *Solar Energy Materials and Solar Cells*, Vol. 69, pp. 17-24, 2001.
- [4] Fernandez, A.M. and Merino, M.G., Preparation and Characterization of Sb<sub>2</sub>Se<sub>3</sub> Thin Films Prepared by Electrodeposition for Photovoltaic Applications, *Thin Solid Films*, Vol. 366, pp. 202-206, 2000.
- [5] Heras, C. and Sanchez, C., Characterization of Iron Pyrite Thin Films Obtained by Flash Evaporation, *Thin Solid Films*, Vol. 199, No. 2, pp. 259-267, 1991.
- [6] Thomas, B., Ellmer, K., Muller, M., Hopfner, C., Fiechter, S. and Tributsch, H., Structural and Photoelectrical Properties of FeS<sub>2</sub> (Pyrite) Thin Films Grown by MOCVD, *Journal of Crystal Growth*, Vol. 170, No. 1, pp. 808-812, 1997.
- [7] Birkholz, M., Lichtenberger, D., Hopfner, C. and Fiechter, S., Sputtering of Thin Pyrite Films, *Solar Energy Materials and Solar Cells*, Vol. 27, No. 3, pp. 243-251, 1992.
- [8] Willeke, G., Blenk, O., Kloc, Ch. and Bucher, E., Preparation and Electrical Transport Properties of Pyrite (FeS<sub>2</sub>) Single Crystals, *Journal of Alloys and Compounds*, Vol. 178, No. 1-2, pp. 181-191, 1992.
- [9] Nakamura, S. and Yamamoto, A., Electrodeposition of Pyrite (FeS<sub>2</sub>) Thin Films for Photovoltaic Cells, *Solar Energy Materials and Solar Cells*, Vol. 65, No. 1, pp. 79-85, 2001.
- [10] Bronold, M., Kubala, S., Pettenkofer, C. and Jaegermann, W., Thin Pyrite (FeS<sub>2</sub>) Films by Molecular Beam Deposition, *Thin Solid Films*, Vol. 304, No. 1-2, pp. 178-182, 1997.
- [11] Wagh, B.G. and Bhagat, D.M., Some Studies on Preparation and Characterization of Cadmium

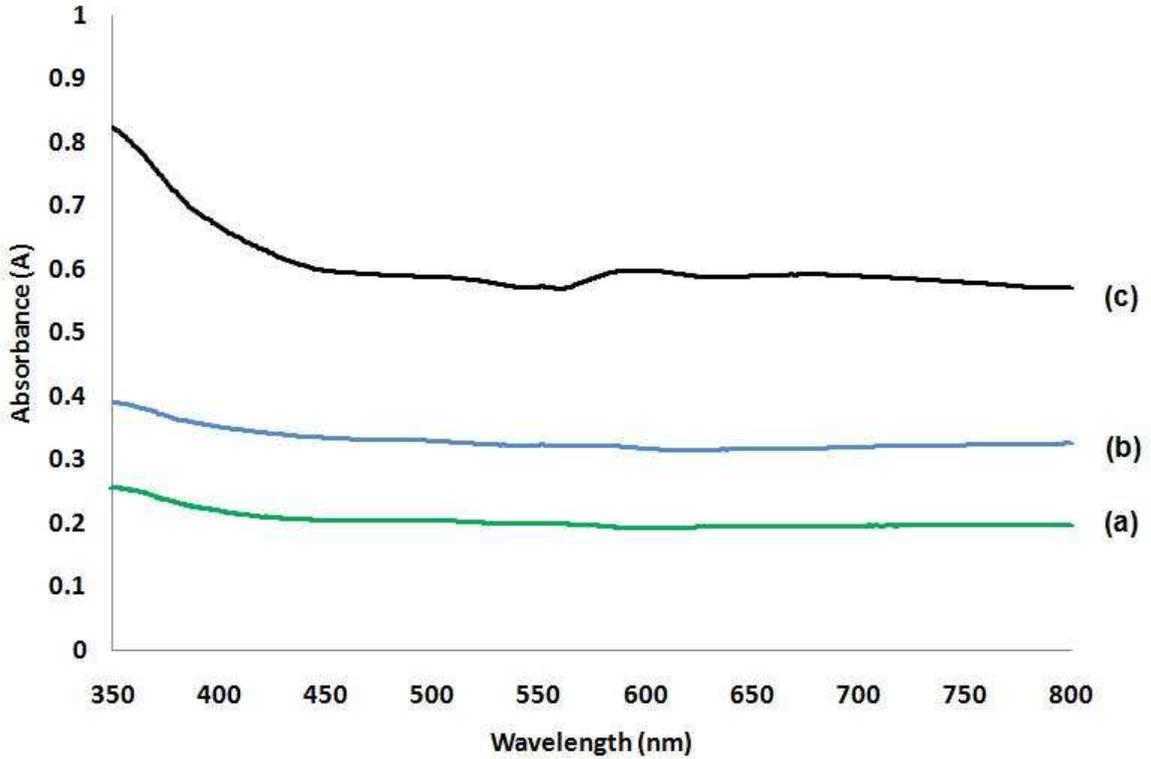
- Sulphide Films, *Current Applied Physics*, Vol. 4, pp. 259-262, 2004.
- [12] Long, F., Wang, W.M., Cui, Z.K., Fan, L.Z., Zou, Z.G. and Jia, T.K., An Improved Method for Chemical Bath Deposition of ZnS Thin Films, *Chemical Physics Letters*, Vol. 462, pp. 84-87, 2008.
- [13] Simurda, M., Nemeč, P., Formanek, P., Nemeč, I., Nemcova, Y. and Maly, P., Morphology of CdSe Films Prepared by Chemical Bath Deposition: The role of Substrate, *Thin Solid Films*, Vol. 511-512, pp. 71-75, 2006.
- [14] Larramendi, E.M., Calzadilla, O., Arias, A.G., Hernandez, E. and Garcia, J.R., Effect of Surface Structure on Photosensitivity in Chemically Deposited PbS Thin Films, *Thin Solid Films*, Vol. 389, pp. 301-306, 2001.
- [15] Hankare, P.P., Chate, P.A., Asabe, M.R., Delekar, S.D., Mulla, I.S. and Garagkar, K.M., Characterization of  $Cd_{1-x}Zn_xSe$  thin Films Deposited at Low Temperature by Chemical Route, *Journal of Materials Science - Materials in Electronics*, Vol. 17, pp. 1055-1063, 2006.
- [16] Kale, R.B., Lokhande, C.D., Mane, R.S. and Han, S.H.,  $Cd_{0.5}Zn_{0.5}Se$  Wide Range Composite Thin Films for Solar Cell Buffer Layer Application, *Applied Surface Science*, Vol. 253, pp. 3109-3112, 2007.
- [17] Bari, R.H., Patil, L.A., Sonawane, P.S., Mahanubhav, M.D., Patil, V.R. and Khanna, P.K., Studies on Chemically Deposited  $CuInSe_2$  Thin Films, *Materials Letters*, Vol. 61, pp. 2058-2061, 2007.



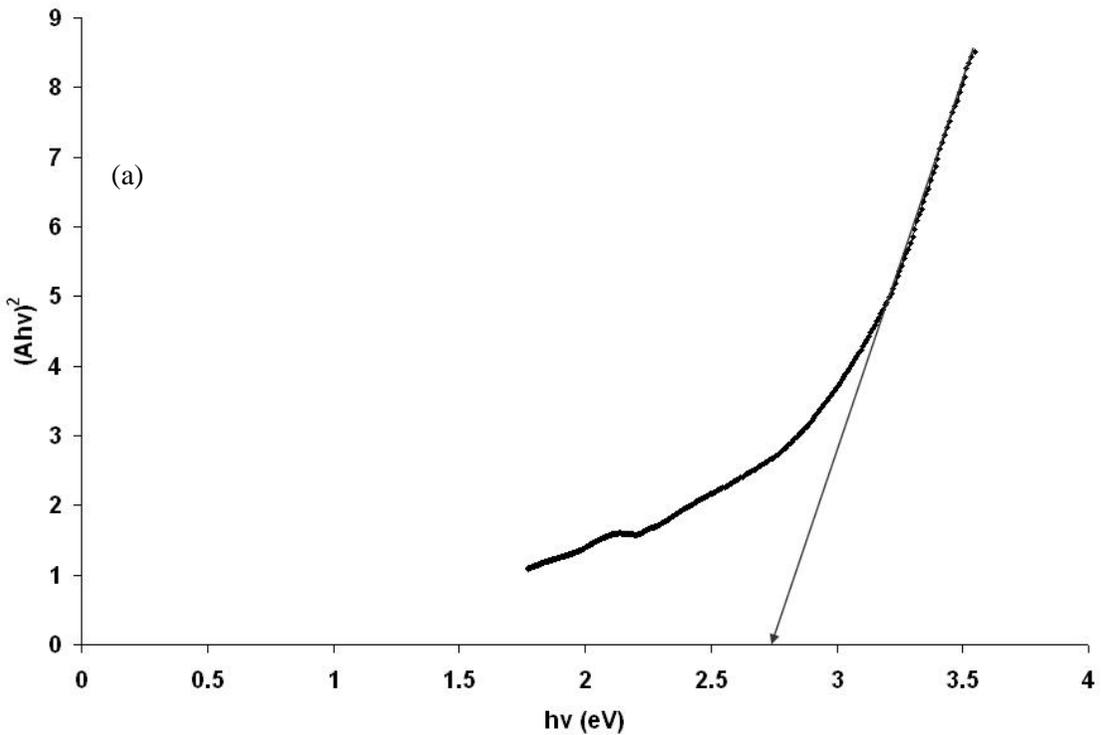
**Fig. 1** X-ray diffraction pattern of  $FeS_2$  thin films deposited at different deposition periods. (a) 30 min (b) 60 min (c) 90 min (  $\blacktriangledown$   $FeS_2$ ;  $\diamond$   $S_xO_2$  )

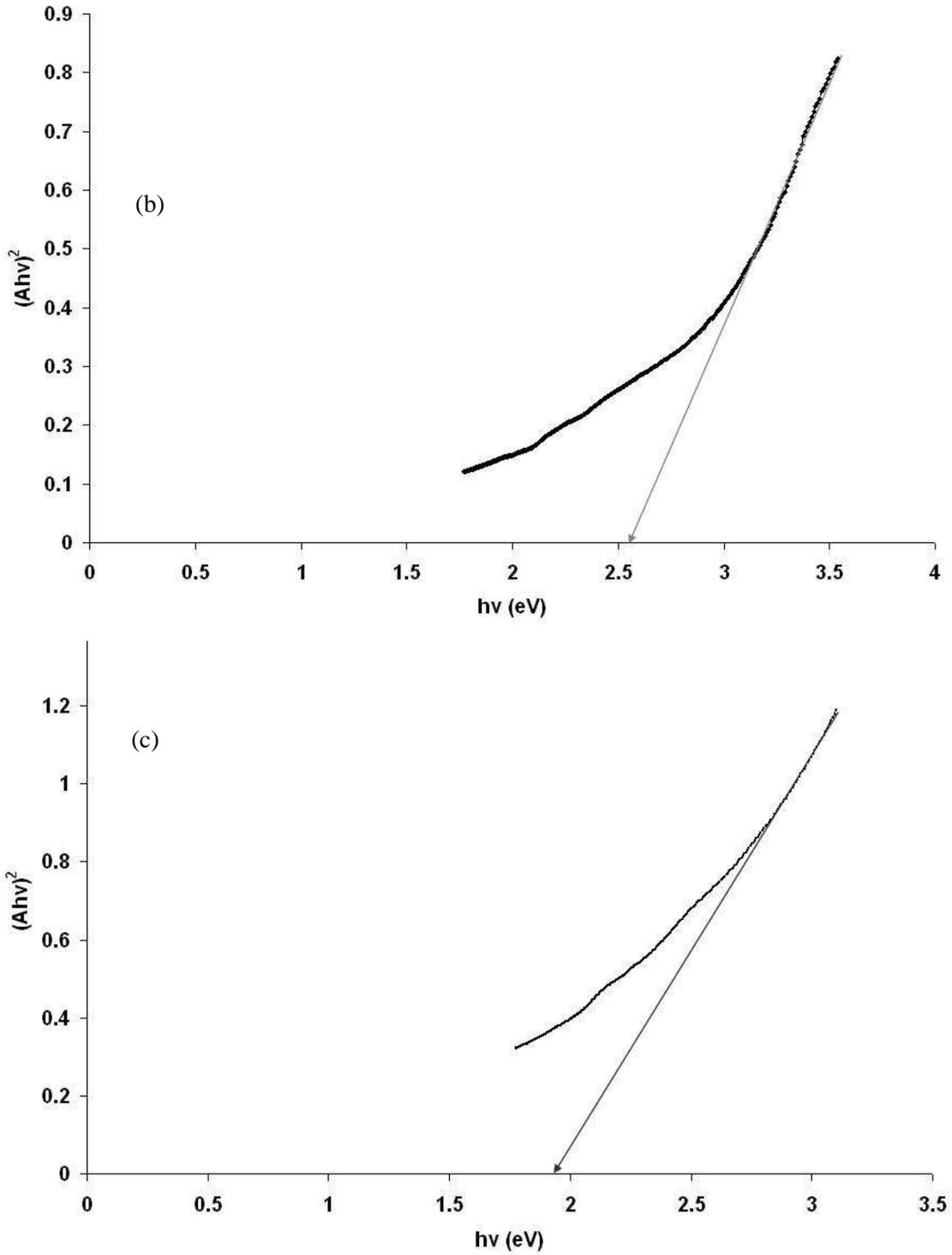


**Fig. 2** SEM micrograph of FeS<sub>2</sub> thin films deposited at different deposition periods  
(a) 30 min (b) 60 min (c) 90 min



**Fig. 3** Absorbance versus wavelength spectra of FeS<sub>2</sub> thin films deposited at different deposition periods (a) 30 min (b) 60 min (c) 90 min





**Fig. 4** Plot of  $(Ahv)^2$  versus  $h\nu$  band gap of  $FeS_2$  thin films deposited at different deposition periods (a) 30 min (b) 60 min (c) 90 min

# A Customized Tabu Search for the Vehicle Routing Problem with Simultaneous Pick-up and Delivery

**Thanchuda Phannikul and Sombat Sindhuchao**

Department of Industrial Engineering

Ubonratchathani University

E-mail: thanchuda@hotmail.com

## **Abstract**

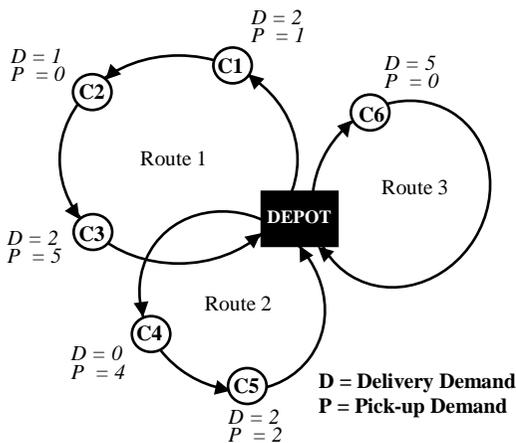
The Vehicle Routing Problem with Simultaneous Pick-up and Delivery (VPSPD) is classified as one of the most complicated and challenging Routing Problems. Having both pick-up and delivery loads at the same customer makes total loads uncertain on the vehicles along all their service routes. This paper presents an implementation of the Savings method for finding initial solutions of 72 benchmark instances. After that, an adaptive metaheuristic, Customized Tabu Search was applied in order to improve those initial solutions. The unique aspects of this search are the acceptance of the move which is tabu with some 'allowance' value and the extension of the tabu tenure with some 'penalty' charges. The allowance and penalty values used in each problem are varied by the problem's size. The results indicate a number of new best known solutions found by this proposed method compared with previous research.

**Keywords:** Tabu Search; Pick-up and Delivery; Vehicle Routing Problem.

### 1. Introduction

Nowadays, manufacturing, trading, services, as well as tourism sectors play an important role in Thailand’s economy. In a competitive market, these activities require a right quantity with a best quality at the lowest cost in a rapid time. All in all, transportation and logistics are the vital key success factors.

In this paper, the Vehicle Routing Problem with Simultaneous Pick-up and Delivery (VRPSPD) is studied. The unique characteristic of this problem is that there are together, pick-up and delivery demands at each point of customers. Pick-up demands must be collected from each customer and sent to a depot. In the meantime, delivery demands must be loaded from the depot and carried to each customer. A vehicle has a limited capacity. With the nature of the problem, a fluctuation of vehicle load will occur all the time. Fig.1 presents the example of VRPSPD.



**Fig. 1** Example of the Vehicle Routing Problem with Simultaneous Pick-up and Delivery (VRPSPD)

In a real situation, we can see many activities that can be applied with the VRPSPD model. The best practical sample is soft drink bottles distribution. Full bottles of soft drink must be supplied from a central depot, loaded on trucks in a sufficient quantity and delivered to grocery stores. After sending each order to each shop, empty bottles (in different quantities) must be collected and carried back to the depot.

However, every Vehicle Routing Problem including VRPSPD is categorized in NP-Hard combinatorial optimization problems. This means we cannot find the exact solution for medium and large scale instances in a reasonable time. Having only 15 customers in a problem can cause more than one thousand billion feasible solutions to evaluate. Heuristics and metaheuristics are always considered to handle these types of situations. The solution gained from heuristic methodology cannot be guaranteed for an optimal solution, but the solution’s quality is good enough to use, considering the computation saving from solving for the exact solution. More importantly, in most real cases, solving for the exact solution is impossible, thus, using heuristics becomes the only feasible choice in practice.

In this article, the Savings method is a heuristic we propose to find initial solutions. After that, the modified Tabu Search called ‘Customized Tabu Search’ is developed to find final solutions. The rest of the paper is previous studies, mathematical modeling, an illustration of Savings and Customized Tabu Search in detail, four types of neighborhood search, comparison of our results with others and finally the summary of the research with a future study suggestions.

### 2. Literature Review

The first person who defined the VRPSPD problem is widely known as Min [1]. The author solved a 22-customer

problem by three-phase heuristics. Firstly, customers were clustered by the average linkage method. In the second phase, vehicles are assigned to these clusters and the Traveling Salesman Problem (TSP) algorithm was used to find routing solutions in the last phase. The initial phase of Halse [2] also began with clustering customers, but then 3-Opt and exchange moves are used to improve those routes.

Later, Dethloff [3] presented an insertion-based procedure applied from cheapest insertion but considered three metrics: travel distance, capacity and radial surcharge with four different criteria. Both real and randomly generated instances were tested on these methods.

In 1999, Salhi and Nagy [4] solved these type of problems by constructing routes from partial customers and then inserting the remaining customers into the existing routes. Then in 2005, they developed another heuristic by constructing initial solutions first and improving them with VRP Routine to become feasible solutions [5]. Solving VRPSPD with multiple depots and the maximum distance constraint has made their research very interesting. These two papers also proved that the VRPSPD can be extended for use with other types of VRP problem. Wassan and Nagy [6] studied the old sets of instances but they moved from their own heuristic to study a general metaheuristic, Reactive Tabu Search. In this article, the initial solutions were formed by the modified Sweep method, and then the searching mechanism of Tabu Search was implemented. The special characteristic of the Reactive Tabu Search is the changing of tenure (period of tabu-move) due to the repetition of answers found. The solution's quality of research was improved in many instances.

In the same year, Montane and Galvao [7] also used a Tabu Search with four types of neighborhood search: relocation, interchange, crossover and all

those three movements combined. Two strategies of intensification and diversification are implemented together with frequency penalization. Chen and Wu [8] presented a hybrid heuristic based on record-to-record travel as well as an insertion-based procedure to solve VRPSPD. Chen and Chia [9] introduced a hybrid metaheuristic between Simulated Annealing (SA) and Tabu List to solve VRPSPD. The quick solutions are obtained by a parallel-insertion procedure before their hybrid search was used.

Bianchessi and Righini [10] solved the VRPSPD by merging arc-exchange-based with node-exchange-based in Tabu List construction. They also compared the results from Constructive Heuristics and Local Search with Tabu Search.

The most recent papers (published in 2009) about VRPSPD are Zachariadis et al. [11] and Ai and Kachitvichyanukul [12]. The first paper constructed the initial solutions by Cost Savings method then the combination between Tabu Search and Guide Local Search was applied to improve the results. The latter paper used another Metaheuristic which had not been implemented for this type of problem before Particle Swarm Optimization (PSO). This method is inspired by social behavior of bird flocking or fish schooling. A population of random solutions are searched and they are approved by following the current optimum particles. However, only a few of the results from PSO in this article are better than previous studies.

### 3. Mathematical Model

#### 3.1 Problem Characteristics

The assumptions of problem are set as follows:

- All pick-up and delivery demands are deterministic.
- Deliveries must be supplied from depot only.

- No interchanges of goods between customers.
- No restricted pick-up and delivery orders.
- Delivery and pick-up demands cannot be split
- Each customer is visited only once.
- Number of available vehicles is sufficient.
- Every vehicle has equally limited capacity.
- No time window requirement.
- No maximum distance constraints.

**3.2 Notations**

- $i, j$  index for depot and customers (depot = 0, customer = 1, 2, ...,  $n$ )
- $k$  vehicle index
- $C_{ij}$  distance (cost) between customers  $i$  and  $j$
- $n$  total number of customers
- $P_j$  pick-up demand of customer  $j$  where  $j = 1, 2, \dots, n$
- $D_j$  delivery demand of customer  $j$  where  $j = 1, 2, \dots, n$
- $Q$  vehicle capacity
- $TP_{ij}$  total pick-up demand from beginning to customer  $i$  and traveling in arc  $(i,j)$
- $LD_{ij}$  demand left to deliver to customers after customer  $i$  and traveling in arc  $(i,j)$
- $x_{ij}^k$   $\begin{cases} 1, \text{ if vehicle } k \text{ travels from customer } i \text{ to } j \\ 0, \text{ otherwise} \end{cases}$

**3.3 Mathematical Formulation**

$$\text{Minimize } \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij}^k \quad (1)$$

Subject to

$$\sum_{i=1}^n \sum_{j=1}^n D_j x_{ij}^k \leq Q \quad \forall k \quad (2)$$

$$\sum_{i=1}^n \sum_{k=1}^m x_{ij}^k = 1 \quad \forall j > 0 \quad (3)$$

$$\sum_{i=1}^n x_{ij}^k - \sum_{i=1}^n x_{ji}^k = 0 \quad \forall j, k \quad (4)$$

$$\sum_{i=1}^n TP_{ji} - \sum_{i=0}^n TP_{ij} = P_j \quad \forall j > 0 \quad (5)$$

$$\sum_{i=1}^n LD_{ji} - \sum_{i=0}^n LD_{ij} = D_j \quad \forall j > 0 \quad (6)$$

$$TP_{ij} + LD_{ij} \leq Q \sum_{k=1}^m x_{ij}^k \quad \forall i, j \quad (7)$$

$$x_{ij}^k \in \{0, 1\} \quad (8)$$

$$n, C_{ij}, P_j, D_j, Q, TP_{ij}, LD_{ij} \geq 0 \quad (9)$$

The objective function (1) aims to minimize the total cost or total distances that occur in all routes. Constraints (2) ensure that all delivery demands in each route are not greater than vehicle capacity. Constraints (3) count the number of vehicles that serve each customer and guarantee that each customer is visited by only one vehicle (except depot). Constraints (4) ensure that the same vehicle arrives and departs from each customer while constraints (5) and (6) dictate the pick-up and delivery demand constraints, respectively. Constraints (7) verify that at each customer node, both types of demands must not be greater than the vehicle capacity. Constraints (8) define the possible members in the set of decision variables that is only 1 or 0 and the last constraints are the non-negativity constraints.

**4. Initial Solution**

In general, solving procedures of the Vehicle Routing Problem and other combinatorial problems are frequently divided into two phases: an initial phase and an improvement phase. The initial phase begins with finding a basic feasible solution from any heuristic method. Then, some iterated Local Search methods will be employed in order to find a better solution from the initial one in the improving step. A

proper starting solution certainly leads to a better final result. According to our study, we found that the Savings method outperforms other heuristics and gave the most appropriate initial solution. Thus, in this research, the Savings method is selected to be an initial solution constructive algorithm.

The Savings method was first introduced by Clarke and Wright in 1964. We apply this heuristic to our VRPSPD problem with the following steps.

**Step 1** For each possible pair of customers, calculate the savings value between customers  $i$  and  $j$  ( $S_{ij}$ ) from the distance saved if they are combined in the same route instead of visited by a vehicle separately. The savings can be computed using Equation (10); where  $C_{io}$  represents distance from customer  $i$  to the depot,  $C_{oj}$  represents distance from the depot to customer  $j$  and  $C_{ij}$  is the distance from customer  $i$  to customer  $j$ .

$$S_{ij} = C_{io} + C_{oj} - C_{ij} \quad (10)$$

**Step 2** Arrange customers in a decreasing order of savings.

**Step 3** Choose the highest savings and then consider if both customers can be possibly linked together. If so, join them in the same route. If not, discard them and select the next savings.

**Step 4** Continue with step 3 until no more pairs of customers can be combined with positive savings.

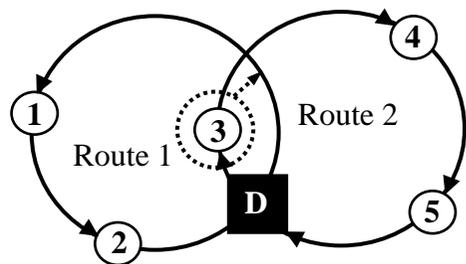
### 5. Neighborhood Search

After an initial solution is obtained from the Savings method, the improvement phase will be implemented. The concept of route improvement comes from customer

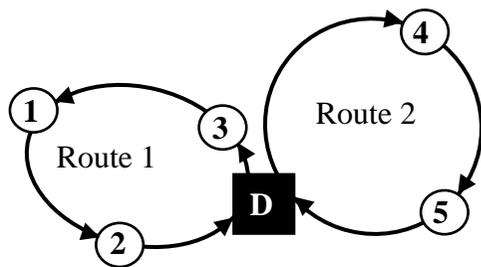
exchange, both intra-route and inter-route. Moving a customer from one position to another creates many possible alternative routes. The number and the order of neighborhood structures have direct effect on search space. Too few structures would result in less exploration, whereas too many structures applied can cause too large search areas, and result in uncontrollable solutions. Our Customized Tabu Search applies four types of neighborhood search (given below).

#### 5.1 Insertion move

An insertion move begins by selecting a customer, and then moving it to a new position. This move considers all possible positions, within current and new routes, that each customer can be relocated to. Please see Fig. 2.



*Insert C3 into route 1 before C1*



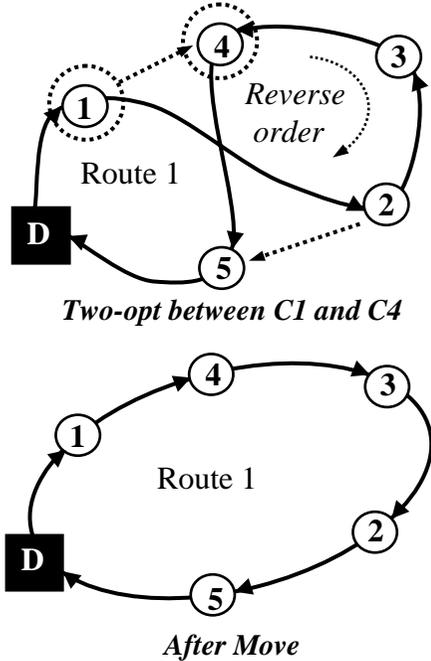
*After Move*

**Fig. 2** Example of Insertion move

#### 5.2 Two-opt move

This type of move considers only intra-route moving. In this improvement method, two non-adjacent arcs are selected and then swapped, if this exchange results

in a better solution. All exchanges of two edges are tested until no better solution can be obtained. Please note that, a Two-opt move can be applied only with a route that has more than two customers. Fig. 3 presents the example of this move.



**Fig. 3** Example of Two-opt move

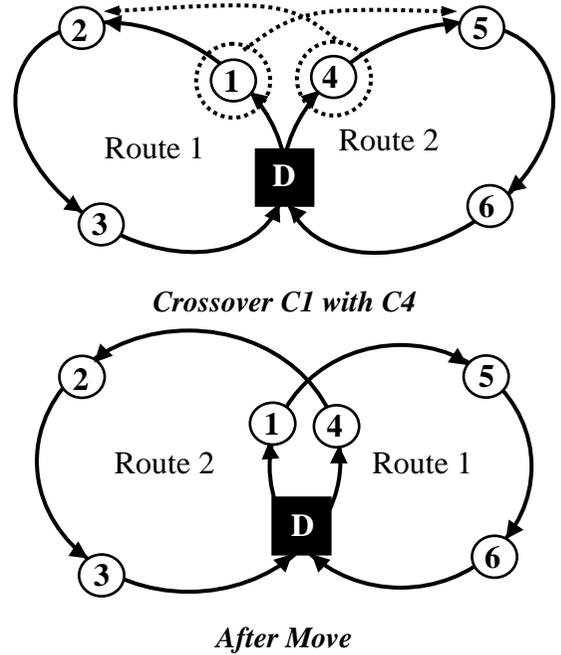
**5.3 Crossover move**

An example of a Crossover move shown in Fig.4. All customers after the selected customers will be swapped. In other words, two arcs from two routes are removed and two new ones are connected. This allows the initial section of the first route to connect to the final section of the second route and vice versa.

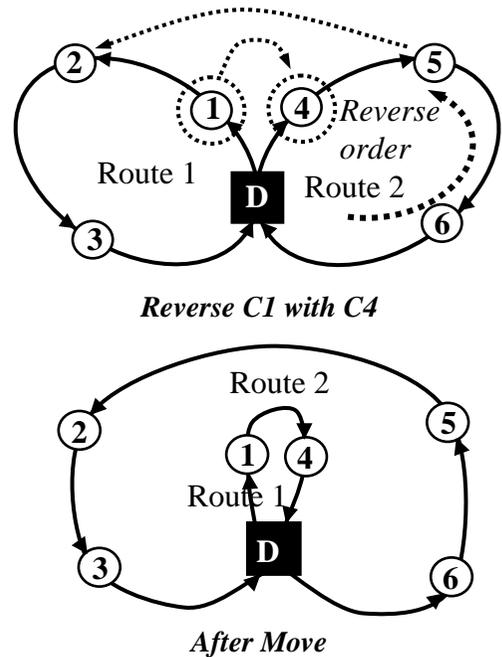
**5.4 Reverse Move**

A reverse move is similar to a crossover move in that two arcs from two routes are removed and two new ones are connected, but in a different way. This type of move will connect both initial sections of the first and the second routes together. Consequently, the final parts of those two routes will be joined. The reversal of the

customer order occurs within the second route. Please see Fig. 5 for more understanding.



**Fig. 4** Example of Crossover move



**Fig. 5** Example of Reverse Move

The four types of moves stated above are done with the Best Improvement principal. During these moves, if any move results in an infeasible solution (excess vehicle capacity), that solution is neglected and the next move is considered instead. Only feasible solutions are recorded. All possible movements are explored before the best feasible one that gives the lowest cost is selected.

## 6. Customized Tabu Search

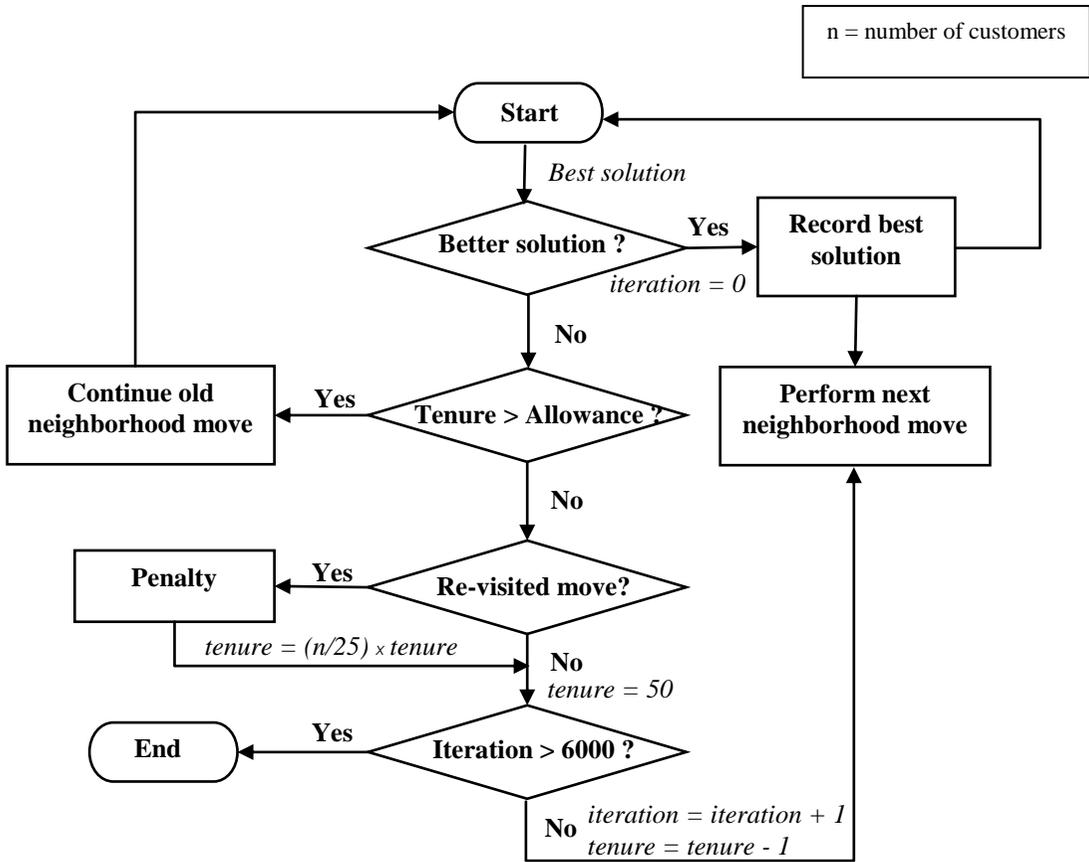
According to the article of Gendreau [13], Tabu Search is attributed to Fred Glover (1986), while some researchers assumed that this methodology is from the publication named 'Tabu Search' by Glover and Laguna in 1997. However, the main rule of Tabu Search is the memory structures used to record determined solutions, that will be marked in the 'Tabu' or 'Taboo' list, to prevent cycling phenomena.

By using the same concept as the standard Tabu Search, we modify our own technique called 'Customized Tabu Search' to solve VRPSPD. Its name is taken from its characteristic in controlling tenure with Penalty and Allowance values, which are set by the number of customers in each problem. Consequently, the algorithm can fit problems of any size.

After conducting some tests, we prefer to set all tabu tenure to begin at 50. If the selected move in any iteration is repeatedly visited (had been recorded in tabu list), the tenure number will be multiplied by a Penalty value. After trials, we discovered that the ratio of 1/25 or four percent of the number of customers gave preferable results.

Allowance value is a special feature of our Customized Tabu Search. In our approach, any move that has a tenure period less than an allowable bound (set as 1/5 or 20 percent of the number of customers) can be selected if it is the best answer found in that iteration. This Allowance value gives an opportunity for a good answer to be selected, even it is still in the tabu period. However, this permission differs from reduction of the tenure with allowance value due to the penalty process. For example, setting Tenure at 50, Penalty at 2, and Allowance at 10, if the move re-occurs, the tenure after penalty will be 100, but if we decrease tenure with allowance (tenure =  $50 - 10 = 40$ ), the tenure after penalty will be 80. This causes different final solutions.

In order to compare the results, we decided to use the same stopping criteria from the latest best known paper (Zachariadis et al. [11]). The searching process will be terminated when 6,000 iterations are reached without any improvement in the objective function.



**Fig. 6** The working procedure of the Customized Tabu Search used after each neighborhood.

Fig. 6 explains the working procedure of the Customized Tabu Search. An initial solution obtained from the Savings method is searched with the first neighborhood move (e.g. Two-opt move). If the best answer found from this move is better than other best solutions, it will be recorded. This answer is then used to be an initial solution to perform the next move (e.g. Crossover Move). But if the best answer found is not better and its Tenure is greater than the Allowance value (being in tabu period), the next best answer from the first neighborhood move is searched after that. In case the Tenure is less than Allowance, this solution is accepted, Tenure is given and recorded in Tabu list. If the answer is selected before, the tenure is

extended longer by penalty process. The iteration and tenure period are counted at this point. Because the tenure is counted after each neighborhood move, totalling four types of move, tenure is set at a higher value when compared with usual Tabu Search.

This Customized Tabu Search approach is applied after each type of neighborhood move until 6,000 iterations are reached without any improvement.

### 7. Computational Results

In order to test the performance of our proposed algorithm, we use VRPSPD instances from three papers. There are 14 problems from Salhi and Nagy [4], 18

problems from Montane and Galvao [7] and 40 problems from Dethloff [3].

A comparison is made between our solutions and those from previous studies. The best known solutions in each dataset are marked with bold characters.

The implementation of the Savings method and the Customized Tabu Search are done in Visual C++ (version 6.0). All 72 results are reported in the following tables.

Table 1 presents the computational results of Salhi & Nagy's problems. This dataset is one of the most popular benchmarks among VRPSPD instances. Although there are no new best known solutions found by our method, we can observe that Customized Tabu Search is competitive with Zachariadis[11] and better than both Montane & Galvao[7] and Ai & Kachitvichyanukul [12] on average.

**Table 1** Computational Results of Salhi & Nagy's Problem

Problem	No. of Cust.	Montane & Galvao (2006)			Zachariadis et al. (2009)			Ai & Kachitvichyanukul (2009)			Customized Tabu Search		
		No. of Vhcls	Total Dist.	CPU Time <sup>a</sup>	No. of Vhcls	Total Dist.	CPU time <sup>b</sup>	No. of Vhcls	Total Dist.	CPU time <sup>c</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>d</sup>
CMT1X	50	3	472	3.73	3	470.48	4.06	3	<b>467</b>	-	3	470.48	2.08
CMT1Y	50	3	470	4.37	3	470.48	3.21	3	<b>467</b>	-	3	470.48	86.13
CMT2X	75	7	695	6.91	6	<b>682.39</b>	6.53	6	710	-	6	690.31	144.52
CMT2Y	75	7	700	7.61	6	<b>682.39</b>	7.94	6	710	-	6	686.86	42.63
CMT3X	100	5	721	11.04	5	<b>719.06</b>	10.52	5	738	-	5	719.90	1229.31
CMT3Y	100	5	<b>719</b>	12.01	5	719.06	13.25	5	740	-	5	723.67	771.66
CMT12X	100	6	675	12.23	5	<b>658.83</b>	12.04	5	691	-	5	659.82	1423.34
CMT12Y	100	6	689	12.80	5	<b>660.47</b>	10.43	5	697	-	5	667.12	192.47
CMT11X	120	4	900	18.17	4	<b>831.09</b>	16.82	4	895	-	4	846.94	1.67
CMT11Y	120	5	910	18.04	4	<b>829.85</b>	15.26	4	900	-	4	849.89	563.75
CMT4X	150	7	880	24.60	7	<b>854.21</b>	22.98	7	912	-	7	859.78	1755.11
CMT4Y	150	7	878	29.07	7	<b>852.46</b>	28.65	7	913	-	7	860.68	981.05
CMT5X	199	11	1098	51.50	10	<b>1030.56</b>	57.62	10	1167	-	10	1039.77	5277.00
CMT5Y	199	10	1083	56.21	10	<b>1031.69</b>	53.80	10	1142	-	10	1038.11	2353.97
<b>Average</b>			777.86			749.50			796.36			755.99	

<sup>a</sup>CPU Seconds in Athlon 2.0 GHz and 256 MB of RAM.

<sup>b</sup>CPU Seconds in Pentium IV 2.4 GHz and 1 GB of RAM.

<sup>c</sup>No available data

<sup>d</sup>CPU Seconds in Intel Centrino Duo 1.66 GHz (Laptop) and 1 GB of RAM.

Results shown in Table 2 indicate that Customized Tabu Search can find the best known solution of as many as 32 problems from 40 of Dethloff's problems. Among these, there are five answers that are the new best solutions; CON3-4, CON3-6, CON3-9, CON8-2 and CON8-5. The CPU

time is difficult to compare because of the difference between computer specifications and stopping criteria. However, our proposed algorithm can find better or equivalent results with less time used in many problems.

**Table 2** Computational Results of Dethloff’s Problem

Problem	No.of Cust.	Dethloff (2001)			Montane & Galvao (2006)			Zachariadis et al. (2009)			Customized Tabu Search		
		No. of Vhcls	Total Dist.	CPU Time <sup>a</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>b</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>c</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>d</sup>
SCA3-0	50	-	689.00	-	4	640.55	3.37	4	<b>636.06</b>	2.83	4	<b>636.06</b>	7.27
SCA3-1	50	-	765.60	-	4	<b>697.84</b>	3.25	4	<b>697.84</b>	2.12	4	<b>697.84</b>	0.48
SCA3-2	50	-	742.80	-	4	<b>659.34</b>	3.52	4	<b>659.34</b>	2.58	4	<b>659.34</b>	1.81
SCA3-3	50	-	737.20	-	4	<b>680.04</b>	3.31	4	<b>680.04</b>	3.13	4	<b>680.04</b>	18.09
SCA3-4	50	-	747.10	-	4	<b>690.50</b>	3.43	4	<b>690.50</b>	2.68	4	<b>690.50</b>	0.86
SCA3-5	50	-	784.40	-	4	<b>659.90</b>	3.67	4	<b>659.90</b>	2.56	4	<b>659.90</b>	0.55
SCA3-6	50	-	720.40	-	4	653.81	3.35	4	<b>651.09</b>	4.40	4	<b>651.09</b>	4.45
SCA3-7	50	-	707.90	-	4	<b>659.17</b>	3.33	4	<b>659.17</b>	2.98	4	666.60	34.09
SCA3-8	50	-	807.20	-	4	<b>719.47</b>	3.40	4	<b>719.47</b>	3.38	4	<b>719.47</b>	0.11
SCA3-9	50	-	764.10	-	4	<b>681.00</b>	3.41	4	<b>681.00</b>	3.86	4	<b>681.00</b>	4.50
SCA8-0	50	-	1132.90	-	9	981.47	4.14	9	<b>961.50</b>	3.21	9	<b>961.50</b>	13.55
SCA8-1	50	-	1150.90	-	9	1077.44	4.27	9	<b>1050.20</b>	3.55	9	1051.56	36.56
SCA8-2	50	-	1100.80	-	10	1050.98	4.20	9	<b>1039.64</b>	4.67	9	<b>1039.64</b>	82.95
SCA8-3	50	-	1115.60	-	9	<b>983.34</b>	4.17	9	<b>983.34</b>	3.29	9	<b>983.34</b>	0.14
SCA8-4	50	-	1235.40	-	9	1073.46	4.13	9	<b>1065.49</b>	2.68	9	<b>1065.49</b>	4.27
SCA8-5	50	-	1231.60	-	9	1047.24	4.02	9	<b>1027.08</b>	4.50	9	1027.12	36.86
SCA8-6	50	-	1062.50	-	9	995.59	3.85	9	<b>971.82</b>	2.67	9	<b>971.82</b>	22.84
SCA8-7	50	-	1217.40	-	10	1068.56	4.22	9	<b>1052.17</b>	4.32	9	1062.30	37.47
SCA8-8	50	-	1231.60	-	9	1080.58	3.85	9	<b>1071.18</b>	3.43	9	1082.12	4.25
SCA8-9	50	-	1185.60	-	9	1084.80	4.20	9	<b>1060.50</b>	4.12	9	<b>1060.50</b>	117.61
CON3-0	50	-	672.40	-	4	631.39	3.64	4	<b>616.52</b>	3.89	4	<b>616.52</b>	5.95
CON3-1	50	-	570.60	-	4	<b>554.47</b>	3.31	4	<b>554.47</b>	2.97	4	<b>554.47</b>	12.88
CON3-2	50	-	534.80	-	4	522.86	3.45	4	<b>519.26</b>	3.32	4	523.16	141.34
CON3-3	50	-	656.90	-	4	<b>591.19</b>	3.28	4	<b>591.19</b>	2.78	4	<b>591.19</b>	6.89
CON3-4	50	-	640.20	-	4	591.12	3.47	4	589.32	3.12	4	<b>588.79</b>	59.92
CON3-5	50	-	604.70	-	4	<b>563.70</b>	3.38	4	<b>563.70</b>	3.45	4	<b>563.70</b>	1.97
CON3-6	50	-	521.30	-	4	506.19	3.32	4	500.80	2.98	4	<b>499.05</b>	19.33
CON3-7	50	-	602.80	-	4	577.68	3.51	4	<b>576.48</b>	2.40	4	<b>576.48</b>	45.83
CON3-8	50	-	556.20	-	4	<b>523.05</b>	3.66	4	<b>523.05</b>	5.02	4	<b>523.05</b>	3.80
CON3-9	50	-	612.80	-	4	580.05	3.36	4	580.05	3.14	4	<b>578.25</b>	33.08
CON8-0	50	-	967.30	-	9	860.48	4.19	9	<b>857.17</b>	3.40	9	857.40	13.39
CON8-1	50	-	828.70	-	9	<b>740.85</b>	3.89	9	<b>740.85</b>	3.73	9	<b>740.85</b>	34.45
CON8-2	50	-	770.20	-	9	723.32	3.76	9	713.44	2.87	9	<b>712.89</b>	1.09
CON8-3	50	-	906.70	-	10	811.23	4.12	10	<b>811.07</b>	3.82	10	<b>811.07</b>	2.20
CON8-4	50	-	876.80	-	9	<b>772.25</b>	3.75	9	<b>772.25</b>	2.98	9	<b>772.25</b>	7.28
CON8-5	50	-	866.90	-	9	756.91	3.99	9	756.91	5.76	9	<b>754.88</b>	1.77
CON8-6	50	-	749.10	-	9	<b>678.92</b>	4.04	9	<b>678.92</b>	4.00	9	<b>678.92</b>	9.34
CON8-7	50	-	929.80	-	9	814.50	4.00	9	<b>811.96</b>	2.46	9	812.70	9.83
CON8-8	50	-	833.10	-	9	775.59	3.74	9	<b>767.53</b>	4.21	9	<b>767.53</b>	3.34
CON8-9	50	-	877.30	-	9	809.00	4.13	9	<b>809.00</b>	3.87	9	<b>809.00</b>	0.53
<b>Average</b>			842.72			764.25			758.78			759.48	

<sup>a</sup> No available data

<sup>b</sup> CPU Seconds in Athlon 2.0 GHz and 256 MB of RAM.

<sup>c</sup> CPU Seconds in Pentium IV 2.4 GHz and 1 GB of RAM.

<sup>d</sup> CPU Seconds in Intel Centrino Duo 1.66 GHz (Laptop) and 1 GB of RAM.

**Table 3** Computational Results of Montane & Galvao's Problem

Problem	No. of Cust.	Montane and Galvao (2006)			Zachariadis et al. (2009)			Customized Tabu Search		
		No. of Vhcls	Total Dist.	CPU Time <sup>a</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>b</sup>	No. of Vhcls	Total Dist.	CPU Time <sup>c</sup>
r101	100	12	1042.62	13.20	12	1019.48	10.50	12	<b>1018.17</b>	81.78
r201	100	3	671.03	12.02	3	<b>666.2</b>	8.70	3	666.67	1059.47
c101	100	17	1259.79	12.07	16	<b>1220.99</b>	10.20	16	1235.26	22.63
c201	100	5	666.01	12.40	5	<b>662.07</b>	5.70	5	<b>662.07</b>	1.39
rc101	100	11	1094.15	12.30	10	<b>1059.32</b>	12.90	10	1074.50	126.80
rc201	100	3	674.46	12.07	3	<b>672.92</b>	10.50	3	<b>672.92</b>	669.72
r1_2_1	200	23	3447.2	55.56	23	<b>3393.31</b>	61.80	23	3408.82	2049.30
r2_2_1	200	5	1690.67	50.95	5	<b>1673.65</b>	47.40	5	1685.40	6337.34
c1_2_1	200	29	3792.62	52.21	28	<b>3652.76</b>	66.30	28	3663.41	1542.97
c2_2_1	200	9	1767.58	65.79	9	<b>1735.68</b>	60.90	9	1762.87	2442.53
rc1_2_1	200	24	3427.19	58.39	23	<b>3341.25</b>	45.30	23	3385.39	676.19
rc2_2_1	200	5	1645.94	52.93	5	<b>1562.34</b>	62.40	5	1576.05	5307.84
r1_4_1	400	54	10027.81	330.42	54	<b>9758.77</b>	315.30	54	9793.60	15979.30
r2_4_1	400	10	3695.26	324.44	10	<b>3606.72</b>	273.60	10	3635.98	3749.77
c1_4_1	400	65	11676.27	287.12	63	11207.37	283.50	63	<b>11194.30</b>	32551.10
c2_4_1	400	15	3732	330.20	15	3630.72	336.00	15	<b>3626.57</b>	22508.00
rc1_4_1	400	52	9883.31	286.66	52	9697.65	145.80	52	<b>9682.67</b>	18220.30
rc2_4_1	400	11	3603.53	328.16	11	<b>3498.3</b>	345.00	11	3578.05	21752.80
<b>Average</b>			3544.30			3447.75			3462.37	

<sup>a</sup>CPU Seconds in Athlon 2.0 GHz and 256 MB of RAM.

<sup>b</sup>CPU Seconds in Pentium IV 2.4 GHz and 1 GB of RAM.

<sup>c</sup>CPU Seconds in Intel Centrino Duo 1.66 GHz (Laptop) and 1 GB of RAM.

Table 3 shows the results of Montane & Galvao's problems. Six of best known solutions are found by using of our Customized Tabu Search and four of them are the new best results; r101, c1\_4\_1, c2\_4\_1 and rc1\_4\_1. The higher computational time may come from the stopping conditions, the compiler program, and our unoptimized programming.

## 8. Conclusion

The Vehicle Routing Problem with Simultaneous Pick-up and Delivery or VRPSPD can be implemented to solve many situations in real life. The special aspect of VRPSPD is that there are both pick-up and delivery demands at each customer. It is nearly impossible to find the optimal solution for the medium and large-sized problems within a reasonable time. Heuristics and metaheuristics are developed

to handle this type of problem instead of an exact method.

In this article, we proposed our modified metaheuristic, namely Customized Tabu Search to solve VRPSPD. Firstly, the initial solutions are constructed by the Savings method. Then four types of movement; Two-opt, Crossover, Insertion and Reverse moves are used to search neighborhood space of the beginning solution. The Customized Tabu Search differs from the standard Tabu Search in the sense of Tenure, Penalty, and Allowance values, that will be set to fit the problem size. The solutions are changed from iteration to iteration but the intensification and diversification are controlled by proper parameters during the search.

After computational results of testing 72 problem instances it can be concluded that the proposed algorithm is

competitive with existing benchmarks. Many of the new best known solutions are found by using this method. Moreover, some of them are obtained with less CPU time used.

Further study should be extended to consider more constraints on maximum distance or time window requirements. Other heuristics or metaheuristics can also be investigated with VRPSPD.

## 9. Acknowledgments

This research is supported by the Commission on Higher Education, Ministry of Education, Thailand. In addition, the authors wish to say thanks to Professor Fermin Alfredo Tang Montané for his inspiring paper and his kindness in providing us with datasets used in this paper.

## 10. References

- [1] Min, H., The Multiple Vehicle Routing Problem with Simultaneous Delivery and Pickup Points, *Transportation Research*, Vol. 23A, pp. 377- 386, 1989.
- [2] Halse, K., Modeling and Solving Complex Vehicle Routing Problems, Ph.D. Thesis. Denmark: Technical University of Denmark, 1992.
- [3] Dethloff, J., Vehicle Routing and Reverse Logistics: The Vehicle Routing Problem with Simultaneous Delivery and Pick-up, *OR Spektrum*, Vol. 23, pp.79-96, 2001.
- [4] Salhi, S. and Nagy, G., A Cluster Insertion Heuristic for Single and Multiple Depot Vehicle Routing Problems with Backhauling, *Journal of the Operational Research Society*, Vol. 50, pp. 1034-1042, 1999.
- [5] Salhi, S and Nagy, G., Heuristic Algorithms for Single and Multiple Depot Vehicle Routing Problems with Pickups and Deliveries, *European Journal of Operational Research*, Vol. 162, pp. 126-141, 2005.
- [6] Wassan, N. and Nagy, G., A Tabu Search Algorithm for the Vehicle Routing Problem with Deliveries and Pickups, in *Proceedings of the International Conference on Software Knowledge Information Management and Application*. Chiangmai: SKI MA, 2006.
- [7] Montane, F.A. and Galvao, R. D., A Tabu Search Algorithm for the Vehicle Routing Problem with Simultaneous Pick-up and Delivery Service, *Computer & Operations Research*, Vol. 33, pp. 595-619, 2006.
- [8] Chen, J. F. and Wu, T. H., Vehicle Routing Problem with Simultaneous Deliveries and Pickups, *Journal of the Operational Research Society*, Vol. 57, pp. 579-587, 2006.
- [9] Chen, J. F. and Chia, F., Approaches for the Vehicle Routing Problem with Simultaneous Deliveries and Pickups, *Journal of the Chinese Institute of Industrial Engineers*, Vol. 23, pp. 141-150, 2006.
- [10] Bianchessi, N. and Righini, G., Heuristic Algorithms for the Vehicle Routing Problem with Simultaneous Pick-up and Delivery, *Computer & Operations Research*, Vol. 34, pp. 578-594, 2007.
- [11] Zachariadis, E. E. and et al., A Hybrid Metaheuristic Algorithm for the Vehicle Routing Problem with Simultaneous Delivery and Pick-up Service, *Expert Systems with Applications*, Vol. 36, pp. 1070-1081, 2009.
- [12] Ai, T. J. and Kachitvichyanukul, V., A Particle Swarm Optimization for the Vehicle Routing Problem with Simultaneous Pickup and Delivery, *Computer & Operations Research*, Vol. 36, pp. 1693-1702, 2009.

[13] Gendreau, M., An Introduction to Tabu Search, Universitetet I Oslo. Available on. <http://www.ifi.uio.no/>

[infheur/Bakgrunn/Intro\\_to\\_TS\\_Gendreau.htm](http://www.ifi.uio.no/infheur/Bakgrunn/Intro_to_TS_Gendreau.htm), October 17, 2008.