

具擴充性的黑悟德互連網路

Scalable Interconnection Networks Based on Heawood Graph

唐景裕

海洋大學資訊科學系

Department of Computer Science
National Taiwan Ocean University
E-mail: b0199@hpwsl.ntou.edu.tw

林銘波

臺灣科技大學電子工程技術系

Department of Electronic Engineering
Technology
National Taiwan University of Science
and Technology

江衍志

海洋大學電機工程研究所

Department of Electrical Engineering
National Taiwan Ocean University
E-mail: m8553009@ind.ntou.edu.tw

摘要

本篇論文將提出兩個嶄新的互連網路(interconnection networks)，皆以黑悟德圖形(Heawood graph)為基圖擴展而成，分別稱之為階層式黑悟德網路(folded-Heawood networks)及擴展式黑悟德網路(recursively expanded Heawood networks)。黑悟德圖形具有下列特性：規則的拓撲性質、具高度擴展性及小的直徑，由於它屬於規則圖形(regular graph)非常適用於平行電腦的互連網路，本論文將針對階層式黑悟德網路及擴展式黑悟德網路提出與階層式彼得森網路(folded-Petersen networks)及擴展式彼得森網路(recursively expanded Petersen networks)類似的傳繞與廣播演算法(routing and broadcast algorithms)，並分析評估這兩種新網路的拓撲性質與效能。

關鍵字：彼得森圖形、階層式彼得森網路、擴展式彼得森網路、黑悟德圖形、階層式黑悟德網路、擴展式黑悟德網路、傳繞演算法、廣播演算法、互連網路。

ABSTRACT

In this paper, we propose two novel interconnection networks based on the Heawood graph. Due to some interesting properties of this work, they seem to have the potential as alternatives for the future interconnection structure of multicomputer systems. Like the folded-Petersen networks and recursively expanded Petersen networks, the folded-Heawood networks and recursively expanded Heawood networks have the following properties: regular topology, high scalability, and small diameter. In the paper, we also demonstrate that the proposed networks have elegant routing and broadcast algorithms as those for the folded-Petersen networks and recursively expanded Petersen networks.

Keywords: Broadcast algorithm, Petersen Graph, folded-

Petersen networks, recursively expanded Petersen networks, Heawood graph, folded-Heawood networks, routing algorithm, spanning tree.

一. 簡介

許多高階電腦系統其內部架構大多採用平行處理，處理器間與處理器和記憶體間的資料藉由互連網路(interconnection networks)來傳繞，例如：Intel iPSC[7]，NCUBE[8]和IBM公司的RS/6000。電腦系統的效能除了取決於處理器的速度快慢之外，也和互連網路的架構具有密切的關連。故要提高電腦整體效能，需先具有優良的互連網路架構。

許多高效能的互連網路相繼被學者提出討論，例如：樹狀(tree)、網狀(mesh)、超立方體(hypercube)、遞迴式網路(recursive networks)[9, 10]、金字塔狀(pyramid)[11]。在這些架構中，之前以超立方體(hypercube)應用較廣泛，因其具有許多優良性質：對稱性(symmetry)、規則性(regularity)、高容錯率(high fault-tolerance)、較小的直徑(diameter)及分支度(degree)、具自我傳繞(self-routing)的特性[1, 4]。改進網路拓撲性質及效能的新架構不斷地被提出，其中較優良的架構例如：彼得森圖形(Petersen graph)其具有10個節點(nodes)、分支度為3、直徑為2，以彼得森圖形為基圖所擴展而成的網路，包含階層式彼得森網路(folded-Petersen networks)[6]、彼得森立方體網路(folded-Petersen cube networks)[6]、擴展式彼得森網路(recursively expanded Petersen networks)[11]和超彼得森網路(hyper-Petersen networks)[2]。

評估互連網路性質，主要依據的參數如下[6]：直徑：指網路中任意兩節點間距離(distance)的最大值；分支度：指連接至節點的連結線(links)數目；連通性(connectivity)：指移去節點(連結線)的數目以造成網路不連結；成本(cost)：指直徑與分支度的相乘積。黑悟德圖形其具有14個節點、分支度為3、直徑為3。本文將探討黑悟德網路的性質與上述的參數，並針對黑悟德網路中的資料傳繞，提出傳繞與

廣播演算法。

本篇論文主要的架構如下:第二章描述黑悟德網路及其定址方法(addressing scheme)。第三章將黑悟德網路擴展成 n 維的階層式黑悟德網路,亦將提出傳繞與廣播演算法。擴展式黑悟德網路將於第四章提出。上述網路的效能評估於第五章。第六章為本篇的結論。

二. 黑悟德圖形及網路

2.1 定義與性質

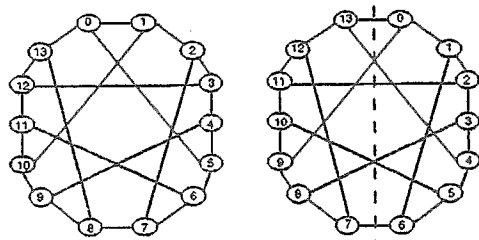
黑悟德圖形具有 14 個節點與 21 條連結線,其數學定義為:

定義 1

$$H = (V_H, E_H), \text{ 其中 } V_H = \{i | 0 \leq i \leq 13\} \quad E_H = \{(i, j) | i - j = 1 \pmod{14} \quad \forall 0 \leq i, j \leq 13\} \cup \{(i, j) | j = (i + 5) \pmod{14}, \forall i \leq 13 \text{ 且 } i \text{ 為奇數}\} \cup \{(i, j) | j = (i - 5) \pmod{14}, \forall i \leq 12 \text{ 且 } i \text{ 為偶數}\}.$$

黑悟德圖形節點的定址方式如下:從網路中任一節點為起點 0 依照順時針或逆時針依序定址,定址後之順時針圖形有兩種同構定址方式如 Fig. 1 所示,由於黑悟德具有對稱性質,逆時針的定址方式與順時針定址方式同構。為了方便起見本論文將採用 Fig. 1(b)的定址方式。

平行電腦中的互連網路以黑悟德圖形為基圖的網路,稱之為黑悟德網路,其網路的架構與基本性質和黑悟德圖形相同。經由定址之後,可以得到下列性質:



(a) 定址方式 1 (b) 定址方式 2
Fig. 1: 黑悟德圖形

- (1) 網路中任意節點 X 具有三個相鄰點,相鄰點位址的計算方式為: $(X+1) \pmod{14}$, $(X-1) \pmod{14}$, 及 $(X+5) \pmod{14}$ (如果 X 為奇數)或 $(X-5) \pmod{14}$ (如果 X 為偶數)。
- (2) 網路中任意兩節點間的訊息傳繞,具有三條不同傳送路徑。其網路的直徑等於 3 與超立方體相同;但黑悟德總節點數較超立方體多。
- (3) 網路中任意兩相異節點間最短迴圈(cycle)為 6,即兩節點間的傳送路徑具有下列性質:
 - (a) 如果最短路徑為 1,則其它兩條路徑的距離皆為 5。
 - (b) 如果最短路徑為 2,則其它兩條路徑的距離皆為 4。

- (c) 如果最短路徑為 3,則其它兩條路徑的距離也皆為 3。

2.2 基本傳繞演算法及廣播演算法

由先前的定義及定址方法,可將黑悟德網路的基本傳繞演算法和廣播演算法表示出來。為了方便起見,將定義下列三個計算相鄰點位址的副程式:

$N_{plus}(S)$:此副程式傳回 S 的相鄰點,其位址為 $(S+1) \pmod{14}$ 。

$N_{minus}(S)$:此副程式傳回 S 的相鄰點,其位址為 $(S-1) \pmod{14}$ 。

$N_{distance}(S)$:此副程式傳回 S 的相鄰點位址,如果 S 是奇數,其位址為 $(S+5) \pmod{14}$; 如果 S 是偶數,其位址為 $(S-5) \pmod{14}$ 。

副程式 $Adjacent(N_1, N_2)$, 表示判別節點 N_1 和節點 N_2 是否相鄰,在此假設網路中任一節點與本身屬於相鄰。黑悟德網路的傳繞演算法將描述如下:

Algorithm Basic_Route(S, D)

Begin

while $D \neq S$ do

if $Adjacent(D, S_{new} = N_{plus}(S))$ then

forwards to S_{new} .

else if $Adjacent(D, S_{new} = N_{minus}(S))$ then

forwards to S_{new} .

else $Adjacent(D, S_{new} = N_{distance}(S))$ then

forwards to S_{new} .

Set $S = S_{new}$

End { of Algorithm Basic_Route }

上述為 Fig. 1(b)之演算法,其中 $N_{plus}(S)$ 、 $N_{minus}(S)$ 、 $N_{distance}(S)$ 及副程式 $Adjacent(N_1, N_2)$ 均可簡易地計算出來,在此省略不提。

在互連網路中除了具備基本的傳繞演算法,還需具備廣播演算法。在黑悟德網路中如何達到廣播功能呢?本篇論文採取簡單的準則,每個根節點(root node)傳送資料至所有相鄰點,接收到資料的節點再將資料複製至另外兩個相鄰點。由於黑悟德網路的直徑為 3,故經過三次傳送,資料便可送達所有節點。在 Fig. 2 中代表從節點 0 傳送資料至所有節點的廣播路徑圖。當第三次傳送時,收到資料的節點會將資料傳送至另外兩個相鄰點,使得節點 3、5、7 和 11 會收到三份複製的資料,故實際上第三次傳送只需傳送一份即可。根據上述特性,廣播演算法可以表示如下:

Algorithm Basic_Broadcast(S, M)

{ S is the source node and M is the message to be broadcasted }

Begin

step 1: Node S forwards M to all its neighbor nodes from all its output ports:

$$N_{pl} = N_{plus}(S), N_{ml} = N_{minus}(S), \text{ and}$$

$$N_{dl} = N_{distance}(S).$$

step 2: Execute the following steps in parallel.

step 2.1: N_{pl} forwards the received M to

$$N_{m2(pl)} = N_{minus}(N_{pl}) \text{ and } N_{d2(pl)} = N_{distance}(N_{pl});$$

step 2.2: N_{ml} forwards the received M to

$$N_{p2(ml)} = N_{plus}(N_{ml}) \text{ and } N_{d2(ml)} = N_{distance}(N_{ml});$$

step 2.3: N_{dl} forwards the received M to

$$N_{m2(dl)} = N_{minus}(N_{dl}) \text{ and } N_{d2(dl)} = N_{plus}(N_{dl});$$

step 3: Execute the following steps in parallel.

step 3.1 $N_{m2(pl)}$ forwards the received M to

$$N_{p3} = N_{plus}(N_{m2(pl)}) \text{ and } N_{d3} = N_{distance}(N_{m2(pl)});$$

step 3.2 $N_{d2(pl)}$ forwards the received M to

$$N_{p3} = N_{plus}(N_{d2(pl)}) \text{ and } N_{m3} = N_{minus}(N_{d2(pl)});$$

End {of Algorithm Basic_Broadcast }

for $0 \leq i, j \leq n-1$.

在 Fig. 3 中為二維的階層式黑悟德圖形。平行電腦中的互連網路採用階層式黑悟德圖形連接而成，則稱之為階層式黑悟德網路。

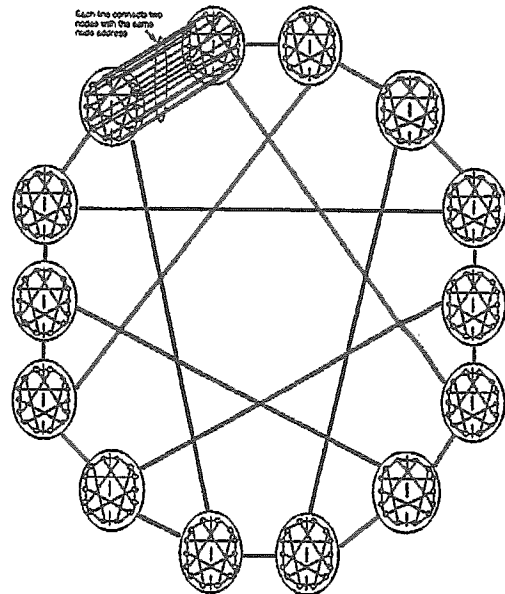


Fig. 3 二維階層式黑悟德網路

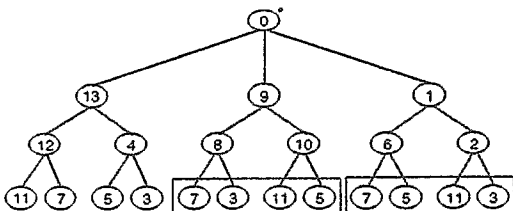


Fig. 2 黑悟德網路最小擴展樹(spanning tree)

上述黑悟德網路基本的資料傳繞及廣播演算法具有簡單且自我傳繞的特性。

三. 階層式黑悟德網路

3.1 定義與性質

本文提出兩種將黑悟德圖形擴展成 n 維網路。首先，介紹階層式黑悟德網路，因為其網路架構具有節點對稱 (node-symmetric) 和邊對稱 (edge-symmetric) 的性質。其擴展方式取自於階層式彼得森網路 [5, 6]，階層式黑悟德網路的數學定義如下：

定義 2

n 維的階層式黑悟德圖形，可定義為 $FH_n = (V_{FH_n}, E_{FH_n})$ ，其中 $V_{FH_n} = \{(V_{n-1}, V_{n-2}, \dots, V_0) | V_i \in V_H\}$ ，且 $E_{FH_n} = \{(U_{n-1}, U_{n-2}, \dots, U_1, \dots, U_0), (V_{n-1}, V_{n-2}, \dots, V_1, \dots, V_0) | U_j = V_j, \forall j \neq i, \text{ 且 } (U_i, V_i) \in E_H\}$

3.2 傳統與廣播演算法

如何將黑悟德網路傳統和廣播演算法擴展至階層式黑悟德網路？首先針對資料傳統的問題，利用基本的傳統演算法擴展而成，其演算法如下：

Algorithm $FH_n_Route(S, D)$

{The source and destination node addresses represented as $S(S_{n-1}, S_{n-2}, \dots, S_0)$ and $D(D_{n-1}, D_{n-2}, \dots, D_0)$ respectively. }

Begin

for $i = n-1$ to 0 step -1 do

Basic-Route(S, D_i)

End { of for }

End { of Algorithm FH_n_Route }

上述演算法中的 Basic-Route(S, D) 函數，其詳細演算法如下：

Algorithm Basic-Route(S, D)

Begin

While $D_i \neq S_i$ do

if Adjacent($D_i, S_{new} = N_{plus}(S_i)$) then

forwards to S_{new}

else if Adjacent($D_i, S_{new} = N_{minus}(S_i)$) then

```

forwards to  $S_{new}$ 
else Adjacent( $D_i, S_{new} = N_{distance}(S_i)$ ) then
    forwards to  $S_{new}$ 
Set  $S_i = S_{new}$ .
End { of Algorithm Basic_Route }
    
```

上述演算法的時間複雜度(time complexity)為 $O(n)$ ，其中 n 為網路的維度。

網路中任一節點廣播資料至所有節點。其演算法可由黑悟德網路廣播演算法擴展而成，其演算法如下所示：

```

Algorithm  $FH_n$ _Broadcast( $S$ )
Begin
    for  $i = n-1$  to 0 step -1 do
        Basic_Level_Broadcast( $S_i, M$ )
    End {of for}
End {of Algorithm  $FH_n$ _Broadcast}
    
```

副程式 Basic_Level_Broadcast(S_i, M) 只需將 Basic_Broadcast(S, M) 中的 step1 做部份修改即可。更改內容如下所示：

```

Algorithm Basic_Level_Broadcast( $S, M$ )
{  $S_i$  is the source node and  $M$  is the message to be broadcasted }
Begin
    step 1: Node  $S$  forwards  $M$  to its  $i$  th level neighbor nodes from its output ports.
     $N_{pl} = N_{plus}(S_i)$ ,  $N_{mi} = N_{minus}(S_i)$ , and
     $N_{di} = N_{distance}(S_i)$ .
    
```

由於副程式 Basic_Level_Broadcast(S, M) 執行時間為常數，則階層式黑悟德網路廣播演算法之時間複雜度為 $O(n)$ 。

四. 擴展式黑悟德網路

在 Shen 的論文[11]中提到利用遞迴擴展方式(recursive expansion(RE) method)將彼得森網路推演至擴展式彼得森網路，總節點數可達 10^{n+1} ($0 \leq n \leq 10$) 其分支度為 6、直徑為 $4n+2$ ，在此 10^{11} 足敷實際使用。同理可將黑悟德網路推演至擴展式黑悟德網路，可使網路總節點數增加至 14^{n+1} ($0 \leq n \leq 14$)。

4.1 擴展式黑悟德網路

使用遞迴擴展方式將黑悟德網路推演成多階

層網路，其推演後仍保有規則性質及有效率的訊息傳繞特性。首先，令黑悟德網路為 H ，令擴展後的網路為 H_n^z ， n 代表階層數， $0 \leq n \leq 14$ 。遞迴擴展方式如下：

1. 令 $H_0^z = H$
2. 將 14 個 $H_{n,i}^z$ 其位址為 $(n-1)$ 的節點互連成黑悟德網路，而形成 H_n^z 網路。

在 H_n^z 中其位址由 0 到 i 的節點各有 6 條連結線，其中 3 條連結線屬於階層 0 另外 3 條屬於階層 i ，其餘的節點(即位址由 $i+1$ 到 n)各有三條連結線連接到同一個 H_i^z 的相鄰點。根據上述的擴展方式，在 Fig. 4 中為 H_i^z 網路。

4.2 傳繞與廣播演算法

擴展式黑悟德網路的傳繞方式可從黑悟德網路的基本傳繞方式擴展而成，擴展後仍具有簡易的傳繞方式且為遞迴式演算法。以下為擴展式黑悟德傳繞演算法：

```

Algorithm  $H_n^z$ _Route( $S, D$ )
{The source and destination node addresses are represented as  $S(S_n, S_{n-1}, \dots, S_0)$  and  $D(D_n, D_{n-1}, \dots, D_0)$ , respectively. } ( $0 \leq n \leq 14$ )
Begin
    for  $i = n$  to 1 step -1 do
        Basic_Level_Route( $S, D_i$ )
    End { of Begin }
End { of Algorithm  $H_n^z$ _Route }
    
```

副程式 Basic_Level_Route(S_i, D_i) 之演算法如下：

```

Algorithm Basic_Level_Route( $S, D_i$ )
Begin
    While  $D_i \neq S_i$  do
        Basic-Route( $S_i, N_o(i)$ )
        [  $N_o(i)$  : the node at level 0 with address  $i$  ]
        if Adjacent( $D_i, S_{new} = N_{plus}(S_i)$ ) then
            forwards to  $S_{new}$ 
        else if Adjacent( $D_i, S_{new} = N_{minus}(S_i)$ ) then
            forwards to  $S_{new}$ 
        else if Adjacent( $D_i, S_{new} = N_{distance}(S_i)$ ) then
            forwards to  $S_{new}$ 
        Set  $S_i = S_{new}$ .
    End { of Algorithm Basic_Level_Route }
    
```

由於副程式 Algorithm Basic_Level_Route

(S, D) 執行時間為常數，則 Algorithm H_n^Z _Route(S, D)廣播演算法之時間複雜度為 $O(n)$ ，其中 n 為網路的維度。

擴展式網路中任一節點廣播資料至所有節點的廣播演算法如下所示：

```

Algorithm  $H_n^Z$  _Broadcast( $S$ )
 $k = 6n + 3$ 
Begin
  Basic_Broadcast_All( $S, M, k$ )
  Parallel do
     $H_n^Z$  _Broadcast(any neighbor node of  $S_i$ )
     $H_n^Z$  _Broadcast(the other neighbor node of  $S_i$ )
    .
    .
    .
     $H_n^Z$  _Broadcast(the last neighbor node of  $S_i$ )
  Parallel End
End {of Begin}
End {of Algorithm  $H_n^Z$  _Broadcast}
    
```

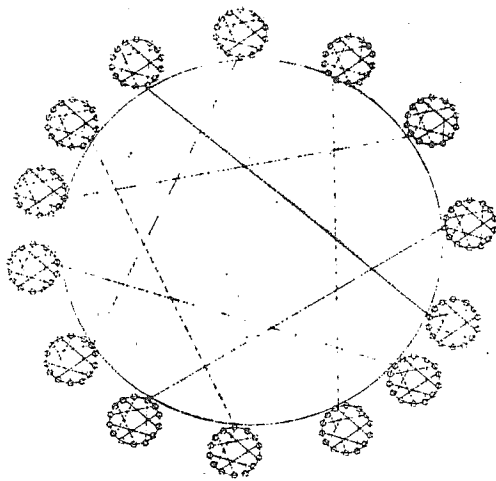


Fig. 4 H_n^Z 網路

上述演算法中的參數 flag 為 0 時代表節點未接收資料 M ，若 flag 為 1 代表節點已收到資料 M 。參數 k 表示程式的遞迴次數，由於擴展式黑悟德網路的直徑為 $6n + 3$ ，故程式的遞迴次數只需 $6n + 3$ 便可將資料傳至所有節點。其中副程式 Basic_Broadcast_All(S, M, k)如下所示：

```

Algorithm Basic_Broadcast_All( $S, M, k$ )
if ( $k = 0$ ) then Stop
else
  Begin
    node  $S$  forwards  $M$  to all its neighbor nodes which
    flag is not equal to address of  $S$  and then setup their
    flag with the address of  $S$ .
    
```

```

Return( $k = k - 1$ )
End {of Begin}
End {of Algorithm Basic_Broadcast_All}
    
```

五. 效能評估

評估互連網路效能的主要因素為成本、分支度、直徑等，上述因素與網路的拓撲性質、規則性、擴展性息息相關，具有優良拓撲性質的網路其擴展非常簡易，擴展後的網路仍具有優良的網路性質。階層式黑悟德網路的總節點數為 14^{n-1} 、直徑為 $3n + 3$ 、分支度為 $3n + 3$ 其網路成本為 $9n^2 + 18n + 9$ 。擴展式黑悟德網路效能評估，依據 Shen 論文中的計算方式，可計算出總節點數節點為 14^{n-1} 其直徑為 $6n + 3$ 、分支度為 6 和成本為 $36n + 18$ 。吾人將超立方體、CCC(cube-connected-cycles)、階層式彼得森網路、擴展式彼得森網路、階層式黑悟德網路與擴展式彼得森網路做效能比較，如 Table. 1 所示。本文假設總節點數為 N ，在橫座標軸上以 2^k 為單位，階層數 k 從 0, 1, 2, ..., 100 依序增加，其網路成本的變化情形如 Fig. 5 所示：，由 Fig. 5 中得知擴展式彼得森網路及擴展式黑悟德網路的成本複雜度與 CCC 相同，而只有超立方體成本的 $1/n$ 。 k 增加時網路的直徑變化情形如 Fig. 6 所示，在此圖中得知，階層式黑悟德網路的直徑略高於階層式彼得森網路而小於超立方體與 CCC。

六. 結論

本文提出兩個適用於高階多處理器系統的互連網路，稱之為階層式黑悟德網路及擴展式黑悟德網路。它們與階層式彼得森網路及擴展式彼得森網路皆具有下列性質：規則的拓撲性質、高度擴展性、較小的直徑，且證明其傳繞與廣播演算法的時間複雜度與階層式彼得森一樣皆為 $O(n)$ 。本文亦分析擴展式黑悟德網路及階層式黑悟德網路的成本遠優於超立方體網路，階層式黑悟德網路的直徑略高於階層式彼得森網路而小於超立方體與 CCC。未來將針對階層式黑悟德網路和擴展式黑悟德網路提出排列(permutation)傳繞演算法、一對多傳繞演算法、多對多傳繞演算法，並將提出一個新的架構其成本將低於上述六種架構。

參考資料

- [1] Kai Hwang, *Advanced Computer Architecture: Parallelism, Scalability, Programmability*, New York: McGraw-Hill, 1993.
- [2] Das, S. Ohring, and A. K. Banerjee, "Embeddings

into hyper Petersen networks: Yet another hypercube-like interconnection topology,” *Journal of VLSI*, special issue on interconnection networks, Vol. 2, No. 4, pp. 335-351, 1995.

- [3] James A. McHugh, *Algorithmic Graph Theory*, Englewood Cliffs, New Jersey: Prentice-Hall, 1990.
- [4] Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis, “*Introduction to Parallel Computing: Design and analysis of Algorithms*,” Redwood City, California: The Benjamin/Cummings Publishing Company, Inc., 1994.
- [5] Sabine Ohring and Sajal K. Das, “Folded Petersen cube networks: new competitors for the hypercubes,” *IEEE Trans. On Parallel and Distributed Systems*, Vol. 7, No. 2, pp. 151-168, February 1996.
- [6] Sabine Ohring and sajal K. Das, “The folded Petersen network: A new communication-efficient multiprocessor topology,” *1993 International Conference on Parallel Processing*, Vol. 1, pp. 311-314, Aug. 1993.
- [7] C. Seitz, “The Cosmic Cube,” *Commun. ACM*, 28(1): 22-33, Jan. 1985.
- [8] J. P. Hayes, T. N. Mudge, Q. F. Stout, S. Colley, and J. Palmeer, “Architecture of a Hypercube Supercomputer,” *Proc. Int. Conf. Parallel Process*, C-36(12): 1450-1467, Dec. 1987.
- [9] S. K. Das, “Designing Recursive Networks Combinatorially,” *Proc. 22nd Int. Conf. Graph Theory, Combinatorics, and Computing*, Batopn Rouge, LA, Feb. 1991.
- [10] S. K. Das and A. Mao, “Embeddings in Recursive Combinatorial Networks,” *Proc. Int. Workshop Graph-Theoretic Concepts in Computer Science*, Wiesbaden, Germany, June 18-20, 1992.
- [11] Hong Shen, “A high performance interconnection network for multiprocessor,” *Parallel Computing*, pp. 993-1001, Sep. 1992.

	總節點數 (nodes)	分支度 (degree)	直徑 (diameter)	成本 (cost)
超立方體	2^n	n	n	n^2
C C C	$n2^n$	3	$(2n-1) + \lfloor n/2 \rfloor$	$\cong 7.5n$
階層式彼得森 (F-P)	10^{n+1}	$3n+3$	$2n+2$	$\cong 6n^2$
擴展式彼得森 (RE-P)	10^{n+1}	6	$4n+2$	$\cong 24n$
階層式黑悟德 (F-H)	14^{n+1}	$3n+3$	$3n+3$	$\cong 9n^2$
擴展式黑悟德 (RE-H)	14^{n+1}	6	$6n+3$	$\cong 36n$

其中 n 代表階層數。

Table. 1 效能評估表

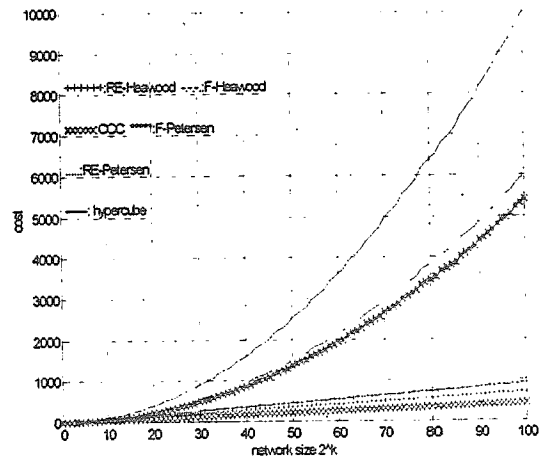


Fig. 5 網路成本比較圖

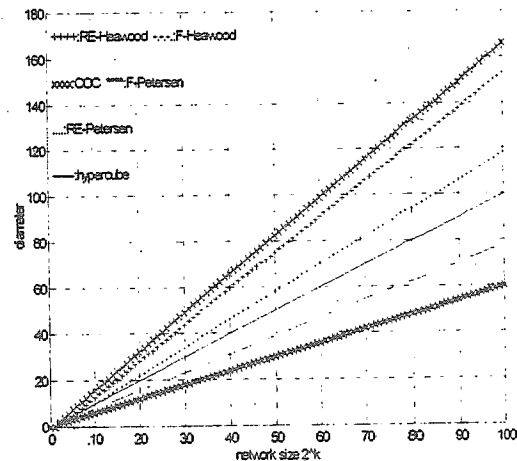


Fig. 6 網路直徑比較圖