高並列計算機に於ける相互結合網の構成

5N-5

═ 多段結合網の多ルート化 ═

坂井　修一　　田中　英彦　　元岡　達

（東京大学　工学部）

1．はじめに

高並列計算機（数十～数千台）に於ける通信系の役割は重要である。当研究室では既に回線交換方式の多段クロスバスイッチ網の検討［１］や、蓄積交換方式の汎用スイッチング・ユニットの試作と評価［２］、　並列データベースマシンの結合網の検討［３］等を行ってきた。

本稿では、最近研究の活発な$\log_k N$段の多段結合網の多ルート化による性能向上の試みの一つに関して報告する。スイッチング・ユニット（ＳＵ）の構成法やＳＵの追加によるスループットの向上をシミュレーションによって評価した結果を報告する。

2．$\log_k N$段の多段結合網

$\log_k N$段の網は、多段スイッチ網としては最も段数の低い網である。ハードウェア量が少なくルーティングが容易（いわゆるセルフ・ルーティングが可能）で遅延を低く抑えられる利点がある反面、閉塞による性能低下が大きく、経路がユニークなため故障に弱い欠点を持つ。図１にその一例として間接キューブ網を示した。

図１では、ＳＵとして２×２のクロスバスイッチを用いているが、４×４あるいはそれ以上の大きさのスイッチにより閉塞率をより低く抑えることが考えられる。ただし１ＳＵを必要以上に大きくすると、全体のハードウェア量の増大、遅延の増大が生じる。

また、先に述べた利点を活かし欠点を補うために、ＳＵ数を増し迂回路を設ける方式を検討した。今回は、

ｉ）$\log_k N$段の網を複数並置する方式

ii）１段分のＳＵを追加する方式

の二つを考察の対象とした。定められたポート間には、ｉ）では網の数だけの、ii）では１ＳＵのポート数だけの経路が存在する。

3．シミュレーションによる評価

２で述べた多段結合網のスループットを以下の環境を仮定して、シミュレーションによって評価する。

（１）ソースモジュールは単位時間あたり m（m ≦１）の確率で各々独立に固定長のデータを生成する。行先番地はランダムに与えられる。

（２）ソースモジュール内にはキュー長ＣのＦＩＦＯキューが存在する。

（３）データ転送は単位時間に１回のみ行われる。

（４）行先モジュールは必ず到着した転送要求を受理する。ただし複数ポートから同時にデータを受けることはしない。

（５）網は回線交換方式とする。

網としては、

（１）$\log_2 N$段網（２×２ＳＵ、１～４枚並置）

（２）$\log_4 N$段網（４×４ＳＵ、１～３枚並置）

（３）$\log_4 N+1$段網（４×４ＳＵ、１～２枚並置）

を対象とした。m ＝１とした時のシミュレーション結果を図２、３、４に示す。グラフの縦軸は１ポートあたりのスループット、横軸はモジュールの台数をそれぞれ表わす。

グラフより以下のことが得られる。

◎$\log_2 N$段網１枚を用いた場合の閉塞による性能低下は非常に大きく、256× 256の網の場合、理想的な時の約２８％のスループットしか得られない。

◎$\log_2 N$段網を４枚並置して、Ｎ×Ｎクロスバスイッチと同程度の結合能力が得られる。$\log_4 N$段網では３枚、$\log_4 N+1$段網では２枚で同様のことが言える。

◎ 256× 256の網を１枚用いた場合、$\log_2 N$段網より$\log_4 N$段網の方が約22％スループットが高い。また$\log_4 N+1$段網にすれば、さらに約54％のスループット向上が得られる。前者はＳＵサイズを大きくしたこと、後者は迂回路を設けたことにより閉塞が抑えられたためである。

このように付加ＳＵにより結合能力が高められるが、同時に迂回路ができたことで網の信頼性が向上する。

4．検討・考察

ＳＵは、バス幅を１バイトとすれば、ハードウェア量の点、ＬＳＩ化にあたってのピン数の制限の点等から、４×４クロスバスイッチを用いるのが妥当

と思われる。また網の枚数を増す方式は性能と信頼性を大きく向上させる反面、ハードウェア量が著しく増す難点がある。（SUの数や線数の増加のみならず、インタフェース部にマルティプレクサ等が必要になる。）網の段数を若干増す方式は、追加ハードウェアも少く高スループットが得られるが、迂回ルーティングのための制御が必要となる。

5．おわりに

$\log_k N$段結合網の改良に関して一つの見解を示した。今後は制御方式を考慮した、より精密な検討を行うことが課題である。

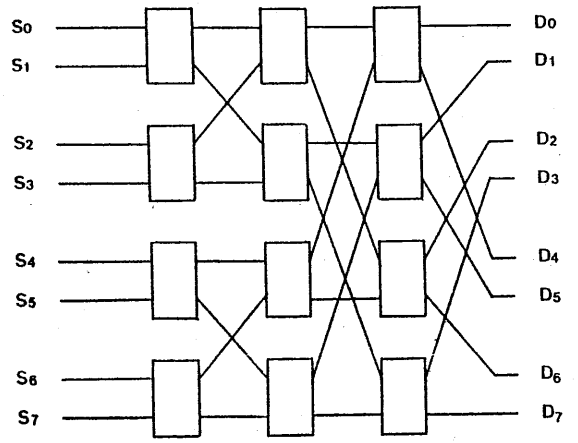

Fig.1 indirect binary 3-cube network

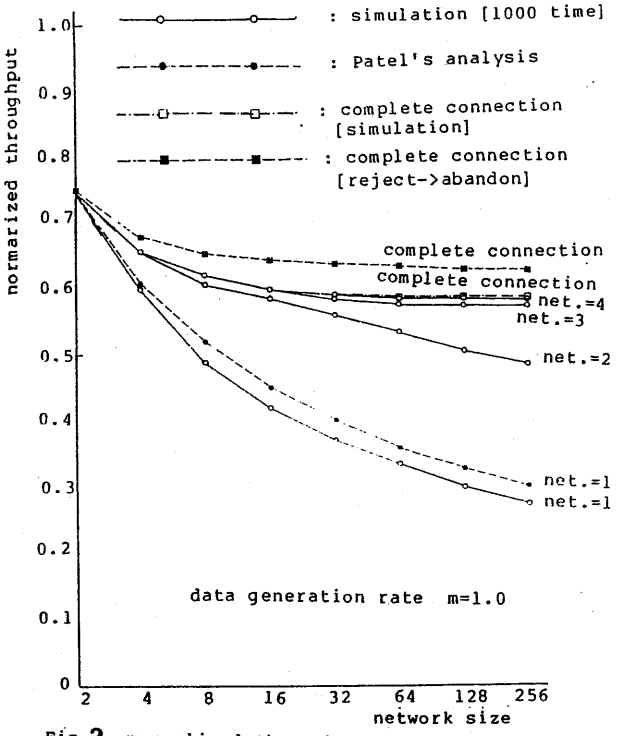

Fig.2 normalized throughput v.s. network size (indirect binary n-cube) circuit switching (without buffer)

＜参考文献＞

［1］菅原他、「超多重プロセッサ間接続機構」第23回情処全大 1981

［2］南、服部他、「プロセッサ間結合網に於けるスイッチング・ユニットの試作と評価」第27回情処全大 1983

［3］坂井他、「データベースマシンGRACEに於けるモジュール間結合網」信学技報 EC 83-14

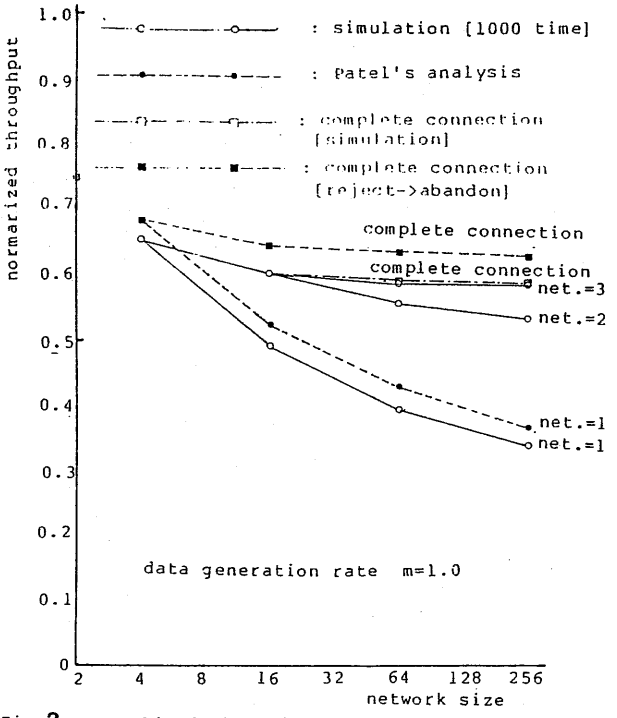

Fig.3 normalized throughput v s. network size (indirect 4-ary n-cube)

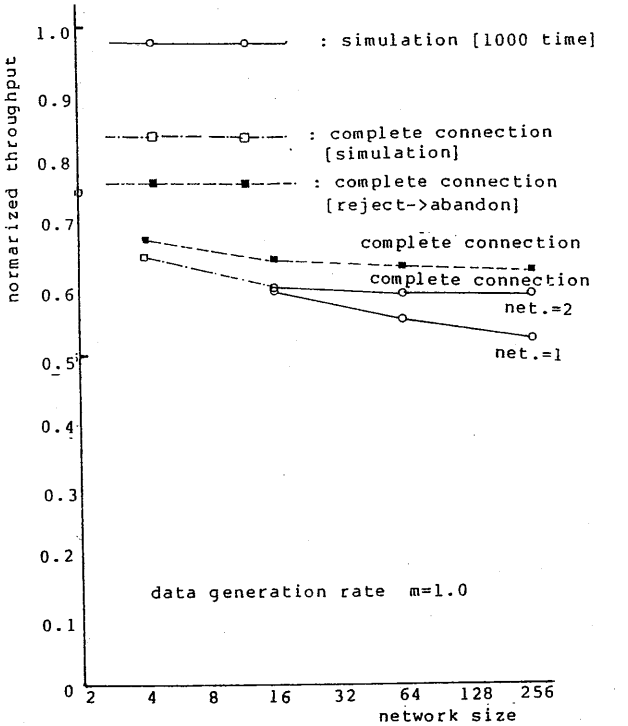

Fig.4 normalized throughput v s. network size ((log$_k$N+1) stage network)