RC-005

Turn モデルに基づく二次元トーラス網の適応型ルーティング

Adaptive routing of the 2-D torus network

的山　和也†　　三浦　康之†　　渡辺 重佳†
Kazuya Matoyama　Yasuyuki Miura　Shigeyoshi Watanabe

１．はじめに

並列処理の分野において，相互結合網に関する研究は，重要なトピックの一つに位置づけられている．かつては，*k*-ary *n*-cube 等の直接網により PE 同士が結合された並列計算機が数多く開発され，商用に提供されていた．近年においては，汎用の並列計算機システムの多くは，PC クラスタなどの PE をスイッチ結合した構成のものが主流となっているが，その一方で「オンチップマルチプロセッサ」の分野においては，PE 間を結合する相互結合網の役割はますます大きなものとなっている．

そのような背景から，これまでに様々な並列計算機向け相互結合網が提案され，実システムに搭載されている．中でも 2-D トーラス網は一般的な相互結合網の一種であり，他の階層型相互結合網の一部として用いられるなど、様々な所で用いられている[1][5].

相互結合網のルーティングには，経路が固定される固定ルーティングと，途中経路の故障や混雑に応じて経路を適応的に変化させる適応型ルーティングの，大きく分けて二つの種類に分けられる．後者は前者に比べて，耐故障性が優れ，局所的な混雑に対する耐性が高いことから，さまざまな研究がなされている．2-D トーラスにおいては，固定ルーティングとして，*x* 方向、*y* 方向、またはその逆の順序で座標を合わせる次元順ルーティングが用いられる．適応型ルーティングにおいては，次元逆転ルーティング[2]，構造化バッファ[3]等，様々なものが提案されているが，これらの手法は追加の仮想チャネルが必要となることから，実装に伴い多大なハードウェアコストを要することから，ハードウェア量を大きく増やせないようなケースでは，追加の仮想チャネルを必要としない手法が求められる．メッシュ網向けの適応型ルーティングアルゴリズムとして，追加の仮想チャネルを必要としない Turn モデルによる方法があり，Turn モデルに基づいたいくつかの手法が提案されている[4]．これらの手法の多くは，メッシュ網向けの手法であり，二次元トーラスにそのまま適用することができないが，Turn モデルの一部の変更によりトーラス網に適用可能な適応型ルーティングがあれば，仮想チャネルの追加によるハードウェアコストの増大を招くことなく適応型ルーティングを実現することが可能となる．そこで，本稿では，二次元トーラスに適用可能な方法として，Turn モデルによる手法の一つである North First 法を応用した North First+1 を提案する．また，シミュレーションプログラムを用いて性能評価を行う．

2.章では 2-D トーラスの構造，固定ルーティングについて記述し，3 章では North Farst+1 のルーティングアルゴリムズ，チャネル切り替え，デッドロックフリーの証明について述べる．4 章においてはランダム通信，ホットスポット通信，マトリクストランスポーズの 3 つの転送方法を用いた実験結果，考察を記述する．5 章は本稿のまとめとなっている．

2．2-D トーラスネットワーク

2.1　構造

図 1 に 2-D トーラスを示す．2-D トーラスは m×m の二次元構造をしており，上下左右の端にある PE がそれぞれ wrap-around チャネルで繋がれている．メッシュに比べて 2 倍の分割帯域幅を持ち，平均ポップ数においても有利であることや，さまざまな並列アルゴリズムとの親和性が高いことなどから過去にさまざまな並列計算機に用いられ，RDT[5]や HTN[1]のように，2-D トーラスを内包した結合網も提案されている．

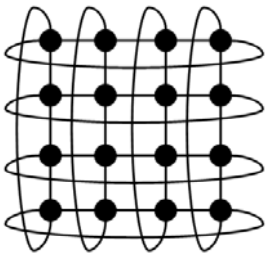

図 1 2-D トーラス

2.2　2-D トーラスの固定ルーティング

2-D トーラスの固定ルーティングアルゴリズムとしては次元順ルーティングが広く用いられている．次元順ルーティングは送信元 PE から， x 軸方向のチャネルのみを使って移動を行い x 軸方向の座標を合わせた跡に， y 軸方向のチャネルを使って目的地の PE に移動する．2-D トーラスにおいて次元順ルーティングを用いる場合にはデッドロックを回避するために 2 本の仮想チャネルを必要とする．

3．k-ary n-cube の適応ルーティング

3.1　Turn モデル

Turn モデルはパケットがルーティング中に進路変更(Turn)するパターンに制限を加え，循環させないようにするものである．考えられる進路変更(Turn)は全部で 8 通りある．そのため，8 通りの進路変更パターンの中から制限を加えなければならない．このモデルは論理的な循環構造に着目し，結合網に依存しないのが特徴である．これにより，故障地点や，混雑地点を迂回する適応ルーティングが可能となる． Turn モデルに基づく適応型ルーティング法として，いくつかの手法が提案されている． 2 次元メッシュ網においては，これらの手法はいずれも、8 通りの進路変更パターンのうち 2 箇所を制限するというものであり，本質的に大きな差異のあるものではないことから，本稿で

は主要な手法の一つである North First(NF)法を取り上げ，2-D トーラスへの応用法を検討する．

3.2 North First(NF)法

図 2 に，次元順ルーティングの Turn モデルを，図 3 に，二次元メッシュによる NF 法の Turn モデルを示す．次元順ルーティングは 8 通りの進路変更パターンのうち 4 箇所を制限しているのに対して，NF 法では，左方向に移動した後上方向に移動するというターンと，右方向に移動した後上方向に移動するというターンの 2 箇所のみに制限を加えたものとなっている．NF 法では制限されるターンの種類が少ないため，経路選択の自由度が高くなっている．

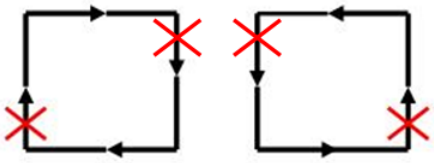

図 2 次元順ルーティングの turn モデル

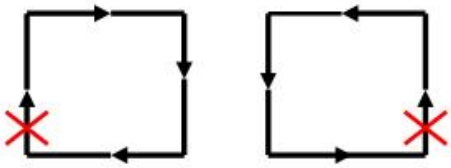

図 3 North First(NF)法の turn モデル

4．適応型ルーティングアルゴリズム NF+1(North First+1)

4.1 ルーティングアルゴリズム

NF 法の Trun モデルを 2-D トーラスにそのまま使用すると，ラップアラウンドチャネルを通過するパケットにより図 4 のような循環が起こってしまうので，NF 法に追加で制限を加える必要がある．そこで，図 5 のように NF 法に更に 1 つ制限を加えることにより循環を回避する．

Turn モデルより，8 つの Turn のうち 3 つに制限が加えられる．具体的には，右方向に移動した後上方向に移動する Turn，左方向に移動した後上方向に移動する Turn，右方向に移動した後下方向に移動する Turn の 3 つに制限を加える．

通常は y 軸方向のチャネルを使って移動を行い，その後に x 軸方向のチャネルを使って目的地の PE に移動する．

横方向のラップアラウンドチャネルを通過後は固定ルーティングを行う．

まず y 軸の座標の差を求め，上下どちらに移動するかを判定をし，上方向の移動が必要な時は上方向の移動を行う．下方向に移動が必要な時，左方向のリンクと下方向のリンクが開いているなら下方向の移動を行う．この時，下方向のリンクが開いていなければ左に移動を行う．上下方向の移動が終わった後，横方向の座標の差を元に左右の移動を行う．ただし，上記で述べたように横方向のラップアラウンドチャネル通過後は固定ルーティングを行う．

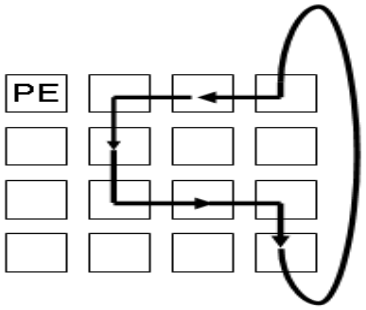

図 4 North First 法をトーラスに適応させた時に発生する循環依存

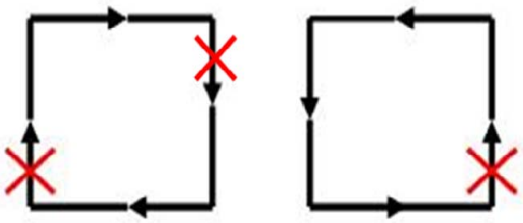

図 5 North First+1(NF+1)

次元順ルーティング、および NF+1 のルーティングアルゴリズムを、それぞれ図 7,図 8 に示す.これらの図のルーティングアルゴリズムは,それぞれ送り元 PE アドレス sx, sy、および送信先 PE アドレス dx, dy を入力とし,途中経路となる PE アドレスを出力するものである.図中の send_packet() 関数は、cx, cy の値を変更し、変更後の値（パケットの現在地アドレス）を出力する関数となる.図 7 の次元順ルーティングでは,Y 方向→X 方向の順に、ルーティングを行う.図 8 の NF+1 では、Y－方向（Y 方向で、PE アドレスが降順になる方向）に向かうパケットで、かつ X－方向（X 方向で、PE アドレスが降順になる方向）へのルーティングが必要になるパケットのみを適応ルーティングの対象とする.それ以外のパケットを適応ルーティングの対象にすると,NF+1 で禁止されているターンのいずれかを含むため,適応ルーティングの対象に加えることはできない.適応ルーティングの対象となるパケットでは、adaptive_send()関数において,Y－方向と X－方向の状態を確認し,Y－方向が混雑していて、かつ X＋方向のチャネルがラップアラウンドチャネルでない場合にのみ X－方向を選択する.上記以外の場合は Y－方向を選択する.

上記のようなルーティングを行うことで,図 5 の Turn モデルに従ったルーティングを最短経路により実現できる.

† 湘南工科大学大学院
Shonan Institute of Technology

```
// Routing Algorithm for Dimension-Order Routing
Routing(sx, sy, dx, dy)
sx, sy;  // source       0 ≦ sx, sy ≦ n－1
dx, dy;  // destination 0 ≦ dx, dy ≦ n－1
{
cx, cy;  //current PE   0 ≦ cx, cy ≦ n－1
   cx =sx; cy=sy;

// dimension Y
   if((dy－sy＋n) mod n ≦ n/2)
    while(cy ≠ dy) send_packet(y+);
   else while(cy ≠ dy) send_packet(y－);

// dimension X
   if((dx－sx＋n) mod n ≦ n/2)
    while(cx ≠ dx) send_packet(x+);
   else while(cx ≠ dx) send_packet(x－);

}
```

図 6　次元順ルーティングのルーティングアルゴリズム

```
// Routing Algorithm for North First +1
Routing(sx, sy, dx, dy)
sx, sy;  // source       0 ≦ sx, sy ≦ n－1
dx, dy;  // destination 0 ≦ dx, dy ≦ n－1
{
cx, cy;  //current PE   0 ≦ cx, cy ≦ n－1
   cx =sx; cy=sy;

// dimension Y
   if((dy－sy＋n) mod n ≦ n/2)
 while(cy ≠ dy) send_packet(y+);
   else{
 if((dx－sx＋n) mod n ≦ n/2)
       while(cy ≠ dy) send_packet(y－);
 else while(cy ≠ dy) adaptive_send(cx, cy, dx, dy);
   }

// dimension X
   if((dx－sx＋n) mod n ≦ n/2)
 while(cx ≠ dx) send_packet(x+);
   else while(cx ≠ dx) send_packet(x－);

}

adaptive_send(cx, cy, dx, dy){
if(chanel_is_not_full(y-)) send_packet(y－);
else if(chanel_is_wa(x-)) send_packet(y－);
   else send_packet(x－);
}
```

図 7　NF+1 のルーティングアルゴリズム

4.2　チャネルの切り替え

本手法において必要な仮想チャネル数は 2 本である．本節では，これら 2 本の仮想チャネルの切り替えルールについて説明する．NF+1 は、Y－方向に向かうパケットで、かつ X－方向へのルーティングが必要になるパケットのみが適応ルーティングの対象となり、それ以外のパケットは次元順ルーティングと同じ振る舞いをするので、詳細は割愛する。なお，説明にある(i→j)はチャネル i からチャネル j に切り替わることを示す.

（状態１）　最初はチャネル 0 を使い，適応ルーティングを行う．(0→0)

（状態２）　（状態１）のルーティング中、縦方向のいずれかのラップアラウンドチャネルに達した場合，チャネル 1 に移動して適応ルーティングを続ける．(0→1)

（状態３）　（状態２）によってチャネル 1 移動後に縦→横のルーティングが行われたとき，チャネル 0 を選ぶ (1→0)

（状態４）　（状態１）～（状態３）において、横方向のラップアラウンドチャネルに達することがある．その際，「横のラップアラウンドチャネル」通過後に縦のチャネルを通過する可能性があるときは，その「横のラップアラウンドチャネル」を選択できない．必ずディスティネーション PE 真横の「横のラップアラウンドチャネル」を選択する．それまでは，縦のチャネルを選択し続ける．そして，「横のラップアラウンドチャネル」通過後，チャネル 1 に切り替える．(0 or 1 → 1)

（状態４）において「横のラップアラウンドチャネル」通過後は固定ルーティングとなる．つまり，縦のチャネルは選べない．チャネルは常に 1 を選ぶ.

4.3　デッドロックフリーの証明

4.1 および 4.2 で示されたルーティングアルゴリズムがデッドロック・フリーであることを証明するために、各チャネルに番号をつけて、パケットがチャネル上を通過するときに通過する順番にチャネル番号が必ず昇順になることが証明される必要がある。そこで、各 PE から出ている、2 本×4 方向＝8 本のチャネルに対して 3 桁の番号を付ける。上位の桁から順に a, b, c の値を割り当て、(a, b, c)の形式で表現するものとすると、チャネルに割り当てられる番号は下記のようになる。

チャネル 0 の各チャネルのチャネル番号は

$$(a,b,c)=\begin{cases}(0,0,y) & \text{Y＋チャネル}\\ (0,2(n-x)-1,n-1-y) & \text{Y－チャネル}\\ (1,0,x) & \text{X＋チャネル}\\ (0,2(n-x),0) & \text{X－チャネル}\end{cases}$$

チャネル 1 の各チャネルのチャネル番号は

$$(a,b,c)=\begin{cases}(0,1,y) & \text{Y＋チャネル}\\ (0,2(n-x)-1,2n-1-y) & \text{Y－チャネル}\\ (1,1,x) & \text{X＋チャネル}\\ (0,1,n-1-x) & \text{X－チャネル}\end{cases}$$

となる。ここで、x、y はそれぞれ PE アドレスの x 成分と y 成分、n は 2-D トーラスの 1 辺の PE 数である。

4×4 のトーラスを例に取ったデッドロック・フリーの証明を図 8, 図 9 に示す.Y+チャネルおよび X+チャネルに関しては、a,が X、Y 方向の区別、b がチャネル 0、1 の区別、c が X、Y それぞれの方向における座標を示しており、これらのチャネルは次元順ルーティングのみを行うものであ

ることから、ルーティングに従ってチャネル番号が昇順になることが分かる。また、本手法で可能な Turn である Y－→X－、X－→Y－、Y＋→X－、Y－→X＋のいずれの場合においてもチャネル番号が昇順になることから、本手法がデッドロック・フリーであることが分かる。

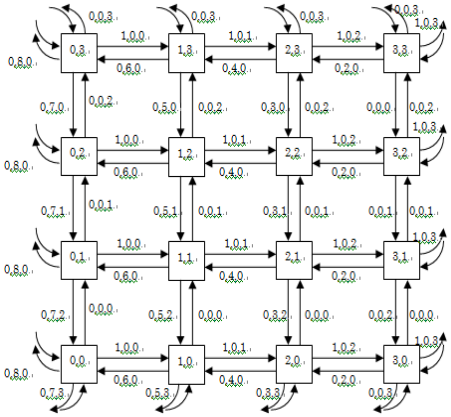

図 8　4×4 トーラス(チャネル 0)

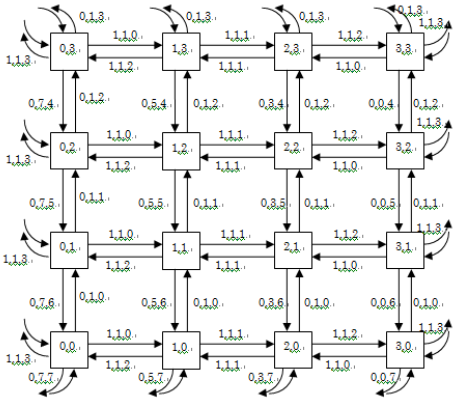

図 9　4×4 トーフス(チャネル 1)

5. 実験

5.1　実験内容

PE 数が 256 である、16×16 トーラス網の NF+1 適応型ルーティングと固定ルーティングでパケットの通信実験を行い，性能を比較する．パケットの通信方法としてはランダム通信，ホットスポット通信，マトリクストランスポーズの 3 種類の通信方式で固定ルーティングとの性能比較を行う．ランダム通信は送信元 PE と目的地 PE をランダムで決め，すべての PE から同時にそれぞれの目的地の PE に向けてパケットが送られる．ホットスポット通信は 10%の確率で左下端の PE である PE(0)にパケットを送り，90%の確率でランダム通信を行う手法である．マトリクストランスポーズは行列の転置を基にした通信方法であり，対角線をまたいで折り返すような転送方法である．マトリクストランスポーズの通信方法を 4×4 のメッシュを例に，図 10 に示す．図 10 のように番号 1,2,3,4 をまたいで折り返す通信を行うのがマトリクストランスポーズである．

性能評価の基準として，平均転送時間，およびスループット用いている．平均転送時間は，パケットの先頭がソース PE を出発してからパケットの最後尾がディスティネーション PE に到達するまでの時間の平均値，スループットは，単位時間，1 PE あたりに到着したフリットの数の平均値である．

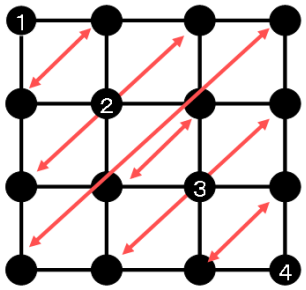

図 10 マトリクストランスポーズ

5.1　シミュレーションプログラムの構成

本研究で用いた PE の構成は PE プロセッサ，送信用ネットワークインターフェイス，受信用ネットワークインターフェイス，そしてルータとなっている．ルータの中身は FIFO，制御回路，デマルチプレクサ，マルチプレクサ，クロスバスイッチとなっている．これら全部を 1 つの PE として扱う．図 11 は PE の構成図となっている．

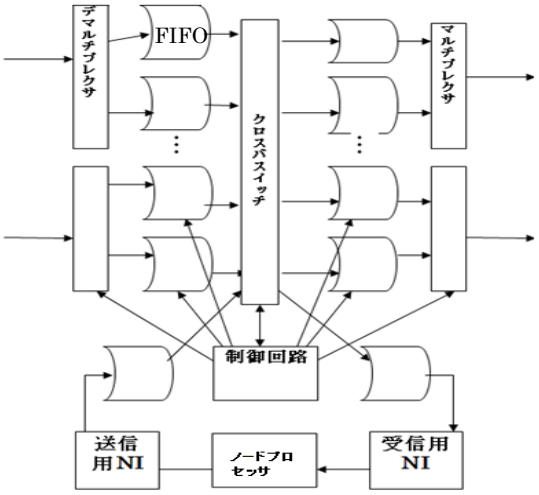

図 11 PE 構成図

5.3　シミュレーションプログラム環境設定

本研究では PE 数が 16×16=256，1PE あたりのリンク数を 4 とし，送信バッファの容量は 2，クロスバスイッチの最大数は 1PE あたりのリンク数+1 なので 4+1=5 となっている．

5.4　ランダム通信

ランダム通信による結果を図 12 に示す．ランダム通信は送信元 PE と目的地 PE をランダムで決め，すべての PE から同時にそれぞれの目的地の PE に向けてパケットが送られる．ランダム通信のように結合網が全体的に混雑する

ような通信パターンでは，適応ルーティングにより混雑を避ける効果が見込めず，適応ルーティングの効果は限定的なものとなる．逆に適応ルーティングの効果により，結合網中に滞留するパケットが増加して混雑が助長されるため，逆にスループットが低下する場合がある．

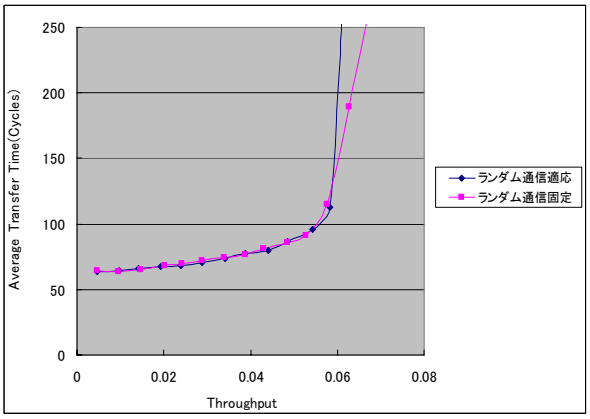

図 12 ランダム通信結果

5.5　ホットスポット通信

ホットスポット通信による結果を図 13 に示す．ホットスポット通信は 10%の確率で PE(0)に通信を行い，残り 90%の確率でランダム通信を行う通信パターンである．

ホットスポット通信で性能比較した場合，NF+1 が固定ルーティングよりもわずかに性能が良くなることが分かる．ホットスポット通信においても，全体的に混んでいる場合は混雑を避ける効果が見込めないが，10%の確率で PE(0)に通信を行っているため，PE(0)周辺だけが混雑する場合ある．その場合，他の PE 同士での通信においては効果的な通信が行える条件が増えるので，わずかに固定ルーティングよりも良くなっていると考えられる．

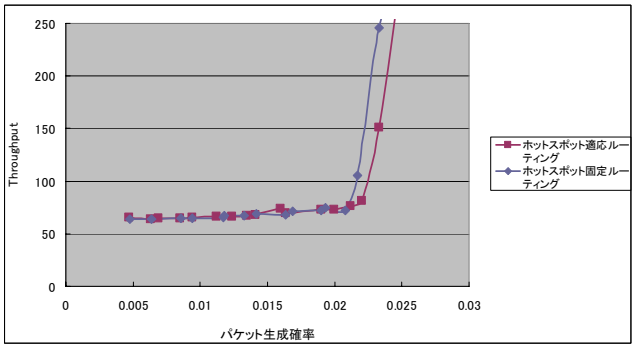

図 13 ホットスポット通信結果

5.6　マトリクストランスポーズ

マトリクストランスポーズは転置行列に基づいた通信方法であり，対角線をまたいで折り返す転送方法である．この通信方法を用いた実験の結果を図 14 と図 15 に示す．

図 14 は縦軸を平均転送時間，横軸をスループットとした実験結果である．図 14 から，スループットが 0.05 以上の部分から固定ルーティングの平均転送時間が適応ルーティングに比べて悪化していることが分かる．

固定・適応双方とも，ある地点をピークにして平均転送時間が低下している．これは，トラフィックが混雑するに従って，通信距離の短いパケットがディスティネーション PE に到達する頻度が増加するためである．

図 15 に，縦軸をスループット，横軸をパケット生成要求を出す確率とした評価結果を示す．図 15 は，パケット生成要求確率に対して，実際にパケットの転送が完了した比率を示したものである．縦軸と横軸の値が接近し縦軸が大きな値を取るほど，ネットワークの処理能力が高いことを意味する．図 15 に示すように，適応ルーティングの方がスループットにおいて高い値を示している．以上の結果により，マトリクストランスポーズで性能比較した場合，グラフより，適応ルーティングの方がグラフの右側にあることから，固定ルーティングよりも高い処理能力があると分かる．

上記のホットスポット通信とマトリクストランスポーズの結果より，限定的な通信手段を用いると NF+1 の方が性能が良くなることがわかった．

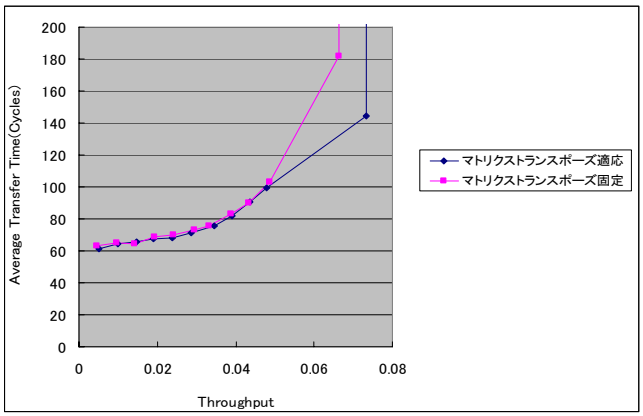

図 14 マトリクストランスポーズ通信結果 1

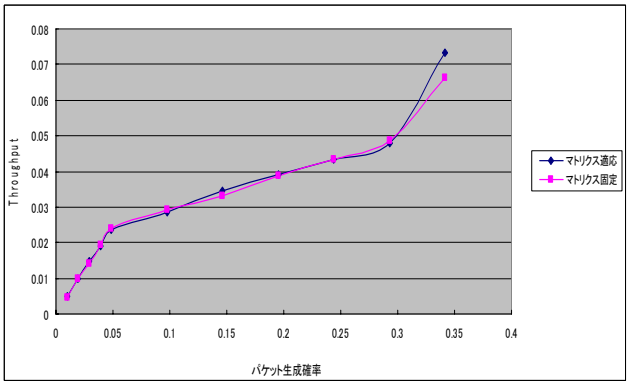

図 15 マトリクストランスポーズ通信結果 2

6. 終わりに

本校では 2-D トーラス網向けの適応ルーティングの一種である NF+1 の提案を行った．ホットスポット通信やマトリクストランスポーズによる実験を行った場合，固定ルーティングよりも性能が良いことが明らかになった．限定的な通信手段を用いた場合は NF+1 の方が性能が良いと分かり，ランダム通信においてもわずかに性能が劣っているが耐故障性という点で NF+1 の方が優れていると思われる．

今後，他の 2-D トーラス網向けのルーティングアルゴリズムを用いた時と性能比較等も行っていきたい．特に，仮想チャネルを必要としない他の手法の比較を行い，本手法の有用性を検証する必要がある．また，デッドロックフリーに関する一般的な証明が今後の課題となる．

7. 参考文献

[1] M.M. Hafizur Rahman and Susumu Horiguchi, HTN: A New Hierarchical Interconnection Networks for Massively Parallel Computers, IEICE TRANSACTIONS on Information and Systems, Vol.E86-D No.9, pp.1479-1486, 2003

[2]William J. Dally and Hiromichi Aoki, Deadlock-Free Adaptive Routing in Multicomputer Networks Using Virtual Channels, IEEE Trans. On Parallel and Distributed Systems, Vol,4, pp. 466-475, 1993

[3]M.P.Merlin and J.P.Schweitzer, Deadlock Avoidance in Store-and-Forward Networks-1: Store and Forward Deadlock, IEEE Trans. On Comm., Vol.COM-28, No.3, pp.345-354, 1980

[4] C.J.Glass and L.M.Ni, Maximally Fully Adaptive Routing in 2D Meshes, ISCA92, pp.278-287, 1992

[5]Akira Funahashi and Akiya Jouraku and Hideharu Amano, Adaptive Routing for Recursive Diagonal Torus, The transactions of the Institute of Electronics, Information and Communication Engineers. D-I J83-D-I(11) pp.1143-1153