計算機アーキテクチャ 132－3
システムソフトウェアとオペレーティング・システム 80－3
ハイパフォーマンスコンピューティング 75－3
（1999．3．4）

超並列計算機向き相互結合網 SRTにおける適応型ルーティング

川井 雅之 井口 寧* 堀口 進
北陸先端科学技術大学院大学
情報科学研究科, 情報科学センター*
〒923-1292 石川県能美郡辰口町旭台 1-1
e-mail:{ms-kawai,inoguchi,hori}@jaist.ac.jp

超並列システムに適合する結合網には，科学技術計算によく用いられる2次元格子結合を含み，ノード当りのリンク数が少数であるなどの実装性，耐故障性などの要件が求められている．SRT (Shifted Recursive Torus) はグリッドの大きさが異なるトーラス結合を再帰的にシフトして構成された，超並列計算機に適した結合網である．SRTは，トーラス結合網に遠距離ノード間通信のためのバイパスリンクを付加しノード当りのリンク数を固定した階層構造を有する結合網であり，従来の相互結合網に比べて遜色ない次数や直径を有している．

SRTにおけるルーティング（再帰ルーティング）は直径や平均距離などの点で十分に高い性能を有しているが、転送経路が固定であるため混雑や故障に対応できない．そこで、本論文では，SRTのデッドロックフリーな適応型ルーティング手法を提案する．また，シミュレーションにより適応型ルーティングの性能評価を行ない従来手法と比較検討を行った．更に，提案する適応型ルーティングは仮想チャネルを増設する必要がなく、また、デッドロックフリーなルーティングアルゴリズムに比べ非常に高い転送能力を有していることを示す．

An Adaptive Routing for Shifted Recursive Torus Networks

Masayuki KAWAI, Yasushi INOGUCHI*, Susumu HORIGUCHI
School of Information Science,Center for Information Science*
Japan Advanced Institute of Science and Technology

Tatsunokuchi, Ishikawa 932-1292. Japan.
e-mail:{ms-kawai,inoguchi,hori}@jaist.ac.jp

A massively parallel computer requires interconnection networks with excellent features of a small diameter, a small number of links, expendability and fault-tolerance. Shifted Recursive Torus (SRT) consists of torus networks which are shifted recursively. SRT has the advantage of that the number of links a node is fixed and the diameter is relatively small. We have proposed a deadlock-free routing of SRT and proved the recursive routing is a near-optimal static routing. However,the proposed recursive routing does not have adaptability and fault-tolerance.

In this paper, we propose a deadlock-free algorithm for adaptive routing of SRT without additional virtual channels. This algorithm allows a detour routing on the same dimension. The adaptive routing algorithm has been proved as a deadlock-free adaptive routing and performances are evaluated by computer simulation. It's seen that the proposed adaptive routing achieves much better dynamic communication performance than a statistic recursive routing.

1 はじめに

自然科学におけるシミュレーションやVLSI設計など，先端科学技術分野における大規模科学技術計算の需要は増大しており，多数のマイクロプロセッサ (MPU) を用いた超並列計算機による高速化が求められている．超並列計算機では，プロセッサ要素 (PE) 間の通信性能が並列処理の効率に大きな影響を与えるために，様々な視点から数多くの相互結合網が提案されている．

科学技術計算の多くは2次元または3次元構造のデータを対象とするため，格子型の結合網と適合しやすい．しかし，格子結合網は，システムの規模が大きくなると通信性能が急速に低下してしまう．

井口ら [1, 2] により提案された Shifted Recursive Torus (SRT) はトーラス結合網を基に，ノード間距離の異なるバイパスリンクを再帰的に付加した結合網であり，少ない次数で RDT[3] や PEC[4] と同程度の直径を実現している．しかしながら，SRT における再帰ルーティングはネットワークの特性を十分に考慮した手法である．しかし，再帰ルーティングは固定型であるため，混雑や故障を回避できず，超並列計算機において重要な適応性や耐故障性を有していない．相互結合網における適応化手法としては Duato の手法 [6, 7] や Turn Model[8] などが代表的であるが，いずれも SRT に適用するにはいくつかの問題がある．そこで本稿では SRT の構造的特徴を考慮した適応型ルーティングを提案する．また，適応型ルーティングがデッドロックフリーであることを証明する．更に，シミュレーションにより動的通信性能について性能評価を行い，適応型ルーティングが高い通信性能を有することを示す．

2 SRT の構成

2.1 1次元 SRT の構成

1次元SRT(1D-SRT) は，ノード数 $N(N = 2^n)$ から成る環状網を基に構成される．環状網のあるノードを番号0とし，ノードを昇順に番号付ける．任意のノード x は，隣接する左右のノード $(x \pm 1)\ mod\ N$ と次式を満たすノードと結合される．

$$x_l\ mod\ 2^l \quad = \quad min(2^{l-1}, 2^T) \qquad (1)$$

ここで，l_{max} は基本型 1D-SRT の最大レベルで，$l_{max} = \log_2 N = n$ となる．また，T は 1D-SRT のタイプを表す．1D-SRT では T の採り方によりレベル $0, l_{max}, l_{max-1}$ ノードの結合方法が若干異なる．$T = n$ の場合を基本型と呼ぶ．基本型ではレベル $0, l_{max}$ ノードで上位リンクを持つことができないが，$T = n-2, n-3$ とすることで全てのノードが上位リンクを持つことが出来る．またを $T = n-2$ を Long-Span 型，$T = n-3$ を Short-Span 型と呼ぶ．図1に 32 ノードから成る基本型 1D-SRT のリンク結合の様子をに示す．

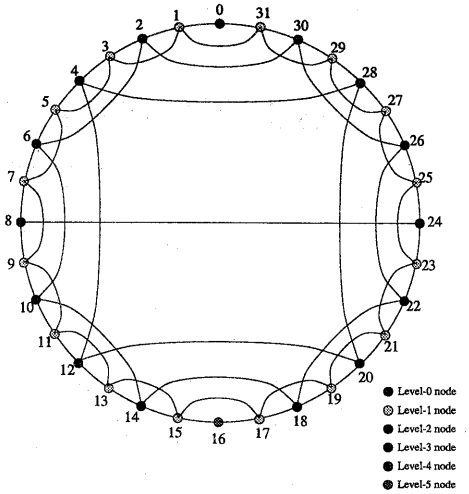

図 1: 32 ノードから成る基本型 1D-SRT.

2.2 2次元 SRT の構成

本節では，1D-SRT を2次元に拡張した 2D-SRT について述べる．$N \times N(N = 2^n)$ のノードから成る 2D-SRT は，各方向に ± 1 離れたノードと接続する基本トーラスと $\pm 2^l$ 離れたノードと接続する上位トーラスにより構成される．長さ 2^l のバイパスリンクを持つノードをレベル l ノードと呼ぶ．レベル l ノード (x_l, y_l) は，次式を満すノードである．

$$(x_l + s_x \cdot y_l)\ mod\ min(2^l, 2^T) \quad = \quad 2^{l-1} \qquad (2)$$

ここで s_x は x 方向のシフト幅である．

2.3 再帰ルーティング

SRT におけるルーティングは，まず，ルーティングに使用するリンクの最大レベルを求め，始点ノードに最も近いそのレベルのノードを探す．次に，始点ノードと得られたノード間に存在する最大レベルを求める．この手続きを始点ノードと接続するノードが見つかるまで，再帰的に繰り返す．同様な手続きを目的ノード側でも行ない，経路を算出する．この方針に基づくルーティングを再帰ルーティングと

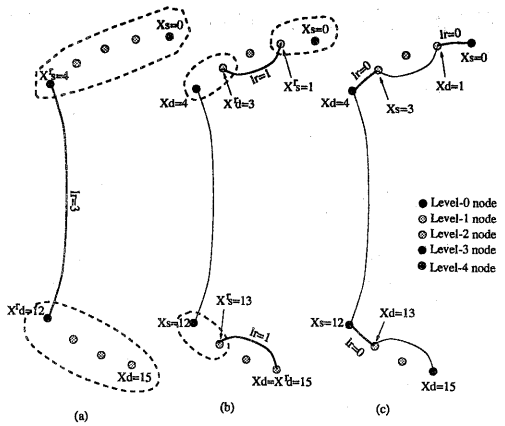

図 2: 1 次元 SRT のルーティングの概念図

呼ぶ.

32 ノードから成る 1D-SRT における，ノード 0 ($x_s = 0$) からノード 15 ($x_d = 15$) までの例を，図 2 に示す．SRT では再帰ルーティングを行なうことにより，最適ではないものの，実用上十分な性能を得ることができる．従来の再帰ルーティングはデッドロックフリーが保証していないが SRT の基本結合がトーラス結合であるため，代表的なデッドロック回避の手法である dimension-order routing が適用できる [9]．しかしながら，再帰ルーティングは固定型ルーティングであるため，適応性や耐故障性を有していない．

3 1D-SRT の適応型ルーティング

1D-SRT のデッドロックフリー・ルーティングは，monotonic order routing を適用することで簡単に得ることができた [9]．しかし，1D-SRT における適応型ルーティングを考える場合，従来の代表的な適応化の手法である Duato の手法 [6, 7] や Turn Model[8] を適用するにはいくつかの問題がある．

まず，Duato の手法であるが，一次元の場合，チャネル選択の自由度は上がるものの，経路選択の自由度といった観点からは適応性が得られない．次に Turn Model であるが， モデルを作成する際のいくつかの禁止事項に 180 度のターンの禁止がある．方向転換が 180 度の方向の一つしかない一次元網ではモデルを作成することは実質的に不可能である．そこで本稿では Turn Model を用いずに，1D-SRT でターンを可能にする手法を提案する．

3.1 諸定義

本稿で用いる用語，諸定義について述べる．

定義 1 ノード番号が大きく（小さく）なる方向を正（負）方向と呼ぶ．

定義 2 パケットがある次元で一方向のみ (monotonic order [5]) に転送されるルーティングを *monotonic order routing* と呼び，*monotonic order routing* を次元順で用いたルーティングを *dimension order routing* と呼ぶ．

定義 3 あるチャネル番号を n 次元ベクトル $C = (c_{n-1}, c_{n-2}, \cdots, c_1, c_0)$ で表した時，任意の 2 つのチャネル番号 $C_1 = (c_{1_{n-1}}, c_{1_{n-2}}, \cdots, c_{1_1}, c_{1_0})$, $C_2 = (c_{2_{n-1}}, c_{2_{n-2}}, \cdots, c_{2_1}, c_{2_0})$ には大小関係が定義され，次の条件を満たす時，C_1 が C_2 より大きいものとし，$C_1 > C_2$ と書く．

（条件） $\exists i, \nexists j$

s.t. $c_{1_i} - c_{2_i} > 0$ *and* $c_{1_j} - c_{2_j} < 0$ *and* $j > i$

3.2 チャネル割当

本稿では，Turn Model を用いずにチャネル番号が昇順となる領域（ノードの組み合わせ）を導き出し，Turn Model では禁止されていた 180 度のターンが可能なルーティングアルゴリズムを提案する．Turn Model に限らず，従来のルーティングアルゴリズムは，アルゴリズムの提案が先に行なわれ，デッドロックフリーは，そのあとでチャネルに適当な番号を与え，循環が生じないことを示すことで保証してきた．

ここで提案する手法はこれまでの手法とは全く逆のアプローチを採る．つまり，まず先に，各チャネルに対して番号を割り当てる．そしてその範囲内で可能なルーティングを順次明確にしていく．そのため提案手法では先にチャネル番号を与える必要がある．そこで 1D-SRT の任意のチャネルに対し次にのように番号を割り当てる．

定義 4 各ノードのチャネルに対し，次のようなチャネル番号 (v, n, l) を割り当てる．

v : 仮想チャネル番号 (*0 or 1*)

n : - チャネルの方向が正:ノード番号 n

- チャネルの方向が負:サイズに対する補数 $(N-1-n)$

l : チャネルのレベル □

3.3 適応型ルーティングの導出

本節では定義4に基づきチャネル番号が割り当てられたSRTにおいて可能となるルーティング手法を導出する．まず，monotonic order routingが可能であることを示し，次に180度のターンが可能であることを示す．

定理 1 定義4に従いチャネル番号を割当てた*1D-SRT*では，*monotonic order routing*が可能である．

証明 パケットのヘッダが存在するノードをn_{cur}，目的ノードをn_{dst}とする．

(i) $n_{cur} < n_{dst}$のとき

monotonic order routingではメッセージは一方向にのみ進むことが許される．そのため，この場合，チャネルは方向は正のみ使用されチャネル番号は常に

$$(v, n_c, l_{cur}) < (v, n_c + d_{l_{cur}}, l_{next})$$

となり，一様に昇順である．また，ラウンドトリップループを通過する場合は，最初にclass 0の仮想チャネルを使用し，ラウンドトリップループを通過したときにclass 1の仮想チャネルに切替えるることで，チャネル番号が常に昇順となることが保証される．

ここでd_lはレベルlのリンクの大きさを示し，l_{cur}は現在のパケットが選択しているチャネルのレベルを，l_{next}は次にパケットが選択するチャネルのレベルを示す．

(ii) $n_{cur} > n_{dst}$のとき

この場合も，(i)の場合と同様な議論ができ，常に，$(v, N-1-n_{cur}, l_{cur}) < (v, N-1-n_{cur}+d_{l_{cur}}, l_{next})$が成り立つ．

(i)，(ii)より常にチャネル番号は昇順となる．つまりデッドロックフリーである． □

次に180度のターンが可能な領域を示す．この領域では，パケットは目的ノードをバイパスリンクを使用して飛び越え，その後で目的ノードへ引き返すことが可能である．つまり，この領域では行き過ぎと逆方向へのルーティングが可能となる．なお，本稿では，180度のターンが可能なルーティングを同次元迂回ルーティングと呼ぶこととする．

定理 2 定義4に従いチャネル番号を割り当てた*1D-SRT*は，カレントn_{cur}と目的ノードn_{dst}が次の条件を満たす時，同次元迂回ルーティングが可能である．

(条件)

$$\begin{aligned} & n_{cur} < n_{dst} \quad , \quad n_{cur} < \frac{N}{2} - \frac{l_{cur}}{2} \\ & \qquad\qquad or \\ & n_{cur} > n_{dst} \quad , \quad n_{cur} > \frac{N}{2} + \frac{l_{cur}}{2} \end{aligned} \tag{3}$$

証明 (i) $n_{cur} < n_{dst}$のとき

このとき，始め，パケットは正の方向に進んでいる．そのパケットがあるノードで一旦，バイパスリンクにより目的ノードを飛び越し，その後，目的ノード側へ進む，つまり後戻りする状況を考えると，カレントn_{cur}と目的ノードn_{dst}，そしてパケットの次の転送先n_{next}の間には次の関係が成り立つ．

$$n_{cur} < n_{dst} < n_{next}$$

また，パケットの通過するチャネル番号は

$$\begin{aligned} & (v, n_{cur}, l_{cur}) \\ & \quad \rightarrow \quad (v, N-1-n_{next}, l_{next}) \\ & \quad \rightarrow \quad (v, N-1-(n_{next}-d_{l_{next}}), l_{next}) \end{aligned}$$

と書ける．したがって，デッドロックフリーであることを示すためには上記のチャネル番号に次のような大小関係が成り立てば良い．

$$\begin{aligned} & (v, n_{cur}, l_{cur}) \\ & \quad < \quad (v, N-1-n_{next}, l_{next}) \\ & \quad < \quad (v, N-1-(n_{next}-d_{l_{next}}), l_{next}) \end{aligned}$$

第2式の大小関係が真であるのは自明である．ここでは第1式の大小関係が真となる条件を導く．

$$(v, n_{cur}, l_{cur}) < (v, N-1-n_{next}, l_{next})$$

が真であるための条件は，仮想チャネルの切替えは考えないので，

$$n_{cur} < N-1-n_{next}$$

となり，$n_{next} = n_{cur} + d_{l_{cur}}$であるので，

$$n_{cur} < \frac{N-1}{2} - \frac{d_{l_{cur}}}{2}$$

が成立する．

(ii) $n_{cur} > n_{dst}$のとき

この場合も，(i)の場合と同様な議論ができ，パケットの通過するチャネル番号は順次

$$\begin{aligned} & (v, N-1-n_{cur}, l_{cur}) \\ & \quad \rightarrow \quad (v, n_{next}, l_{next}) \\ & \quad \rightarrow \quad (v, (n_{next}+d_{l_{next}}), l_{next}) \end{aligned}$$

となり，デッドロックフリーであるための条件は

$$N-1-n_{cur} < n_{next}$$

となる．$n_{next} = n_{cur} - d_{l_{cur}}$ なので，

$$n_{cur} > \frac{N-1}{2} + \frac{d_{l_{cur}}}{2}$$

が成り立つ．

以上より条件を満たす領域では，デッドロックフリーな同次元迂回ルーティングが可能である． □

3.4 1D-SRT の適応型ルーティング

定理 1，2 を用いると，monotonic order routing による通常の再帰ルーティングと同次元迂回ルーティングとを選択的に使用する適応ルーティングが可能である．1D-SRT における適応型ルーティングは次のようなアルゴリズムとしてまとめられる．

```
AdaptiveRouting(n_cur, n_dst){
  if( n_cur < n_dst ) { dir = +1 }
  else { dir = -1 }
  NextNode = RecursiveRouting(n_cur, n_dst)
  if( NextNode == BUSY
      && Cond(n_cur, n_dst) == TRUE
      && |n_dst - n_src| > d_{l_cur}/2 ){
      NextNode = n_cur + dir × d_{l_cur}
  }
}
```

ここで $Cond(n_{cur}, n_{dst})$ は n_{cur}, n_{dst} が条件式 (3) を満たすか否かを判別する関数である．また，$|n_{dst} - n_{src}| > \frac{d_{l_{cur}}}{2}$) は行き過ぎによる目的ノードまでの距離が増大してしまうのを防ぐ為の条件である．だたし，どの程度まで増大を許容するかは任意である．

次にチャネルの使用法であるが，再帰ルーティングでは 2 つある仮想チャネルの使用法はあらかじめ決められいる．最初は番号の小さい方の仮想チャネルを使用し，ラウンドトリップループを通過したとき，もう一方の番号の大きい方のチャネルに切替える．しかし，この方法では最初に使用される方が使用頻度が大きくなる可能性がある．そこで本稿では，ラウンドトリップループを使用する場合は従来通りにチャネルを使用し，それ以外はパケットがネットワーク内に投入される時，空いている方を選択できる手法を採る．このチャネルの使用方法はデッドロックフリーが保証される．つまり monotonic order routing はラウンドトリップループを通過しない限り，仮想チャネルのクラスは固定である．したがって，パケットの投入の際，2 つあるチャネルのどちらを選択してもチャネル番号は一様に昇順となり，デッドロックフリーは保証される．

3.5 2D-SRT の適応型ルーティング

本稿で提案する同次元迂回ルーティングは 2D-SRT に対して，それを次元オーダで施すことで適用することが可能である．

2D-SRT の任意のチャネルに対しチャネル番号 (d, v, n_y, n_x, l) を割り当てる．ここで，d はチャネルが y 方向なら 1 を，x 方向なら 0 を与える．n_y, n_x はチャネルが正方向ならノード番号を，負方向ならノード番号の各次元のサイズに対する補数を与える．また，v は仮想チャネルの番号，l はリンクのレベルである．

ルーティングは次元オーダで行なわれるのでパケットが x 方向にある時は y 方向に関する番号を無視できるのでチャネル番号の体系は 1D-SRT のそれと等価である．また，y 方向にある時も同様でる．したがって，同次元迂回ルーティングを次元オーダで用いれば 2D-SRT に対してデッドロックフリーは保証される．

4 動的通信性能の評価

4.1 シミュレータ概要

デッドロックフリーな再帰ルーティングの性能を評価するために，C++によりシミュレータを作成し，性能評価を行なった．シミュレータはフリットレベルでのシミュレートが可能であり，トポロジ，ルーティングアルゴリズム，パケットサイズ，メッセージの発火確率などが変更可能である．

なお，シミュレーションは，Short-Span 型の SRT に対して行なう．これは，基本型や Long-Span 型では最大レベルリンクの基本トーラス上でのホップ数が $N/2$ であり，ルーティングが一方向に制限されているので，大半がそのリンクを使用しないからである．

4.2 適応型ルーティングの動的性能

ここでは，一様なランダム転送を対象とした，各ルーティングアルゴリズムの動的通信性能評価を行なう．シミュレーションはデッドロックフリーな再帰ルーティング (DLF) とそれに同次元迂回ルーティングを付加した適応型ルーティング (ADP) に対し

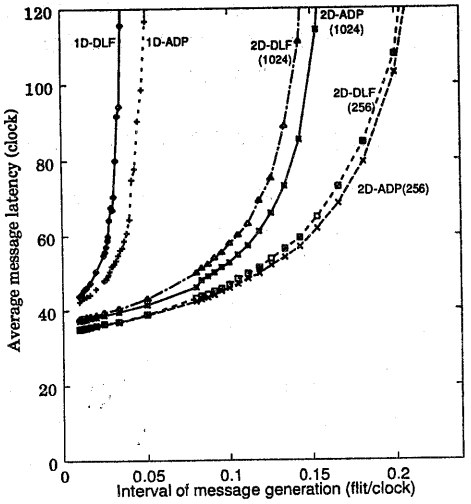

図 3: ランダム転送時の平均通信時間

て行なった．シミュレーションはネットワークサイズを 1D-SRT に対して 256PEs，2D-SRT に対して 256PEs と 1024PEs として行なった．パケット長は 16 フリット，仮想チャネル数は 2，フロー制御方式は Wormhole とした．シミュレーションはある発生確率でランダムな宛先にメッセージを発生させ，その平均通信時間を測定した．シミュレーション時間は各発生確率において 10000 クロック行なった．ここで，発生確率はフリットが毎クロック投入される確率を 1 とする．

図 3 にメッセージ発生確率と平均通信時間の関係を示す．図 3 から，256PE の 1D-SRT では適応型ルーティング (1D-ADP) の方が高い性能を示している．一方，2D-SRT では固定型 (2D-DLF)，適応型 (2D-ADP) の双方に大きい差が見られない．これはノードサイズが 256PEs 程度の場合，各次元のサイズが高々16PEs であるため同次元迂回ルーティングが可能となる領域が狭くなるためと考えられる．そのため，各次元のサイズが大きい 1024PEs のほうが性能が向上している．

5 まとめ

再帰シフトトーラスネットワークの適応型ルーティングの提案を行い，デッドロックフリーを証明した．この適応型ルーティングは，仮想チャネルを新たに付加する必要がなく，また，既存の再帰ルーティングとの併用が可能である．また，シミュレーションによる動的通信性能評価を行なった．その結果，1D-SRT では従来の再帰ルーティングに比べ高い性能を得ることができた．また，2D-SRT では ノードサイズが 32 × 32PEs の方が 16 × 16PEs に比べ性能が向上している．したがって，提案手法は大規模向きであると考えられる．

今後の課題はリンクやノードの故障に対する通信性能の解析，および仮想チャネルを付加した場合の性能評価である．

謝辞
本論文の一部は，文部省科学研究助成金，ならびにセコム科学技術振興財団研究奨励金を用いて行なわれた．関係各位に感謝する．

参考文献

[1] Y. Inoguchi and S. Horiguchi, "Shifted Recursive Torus Network for Mesh-Oriented Interconnections," Proc. 31st Conference on Information Sciences and Systems, Baltimore, Mar. 1997.

[2] Y. Inoguchi and S. Horiguchi, "Shifted Recursive Torus Interconnection for High Performance Computing," IEEE High Performance Computing in Asia Conference, Seoul, pp. 61–66, Apr. 1997.

[3] 楊 愚魯, 天野 英晴, 柴村 英智, 末吉 敏則, "超並列計算機に向き結合網:RDT," 信学論 (D-I) vol. J78-D-I no.2 pp.118-128, 1995.

[4] W.W. Kirkman and D. Quammen, "Packed Expornential Connections - A Hierarchy of 2D-Meshes," Proceeding of the Fifth International Parallel Processing Symposium, pp.464-470, Apl. 1991.

[5] L.M. Ni , L.P. McKinley , "A Survey of Wormhole Routing Techniquse in Derect Networks," IEEE Trans. on Computers, 1993.

[6] J.Duato , "A New Theory of Deadlock-Free Adaptive Routing in Wormhole Networks," IEEE Trans. on Parallel and Distributed System, Vol.4, No.12, December 1993

[7] J.Duato , "A Necessary and Sufficient Condition for Deadlock-Free Adaptive Routing in Wormhole Networks," IEEE Trans. on Parallel and Distributed System, Vol.6, No.10, December 1995

[8] Glass,C.J , Ni, L.M. , "Maximally Fully Adaptive Routing in 2D Meshes," Proceedings of ISCA92, pp.278-287(1992)

[9] 川井 雅之, 井口 寧, 堀口 進, "再帰トーラス相互結合網 SRT におけるデッドロックフリー・ルーティング", 平成 10 年度電気関係学会北陸支部連合大会論文集, p.264 (Oct.1998)