

## 再帰トーラス結合アーキテクチャ†

松山隆司<sup>††</sup> 青山正人<sup>††</sup>

本論文では、MIMD 型分散メモリ並列計算機の基本アーキテクチャとして、再帰トーラス結合アーキテクチャ (RTA) を提案し、その構成と通信距離に関する基本特性およびソート、サーチなどの基本並列処理アルゴリズムを示す。RTA は、トーラス状に結合されたプロセッサを結ぶ通信線上にスイッチを配置し、動的にプロセッサ間の接続関係を変化させることによって、トーラスの再帰的な分割を実現するもので、動的結合網となっている。RTA には、トーラスの幾何学的次元による、1次元 RTA、2次元 RTA、 $\dots$ 、 $n$ 次元 RTA という分類のほかに、スイッチ切り換えの同期方式により、同期式 RTA、非同期式 RTA という分類がある。本論文では、まず1次元同期式 RTA の構成とプロセッサ間の通信距離計算アルゴリズム、並列処理アルゴリズムを述べ、それらを基にすることによって  $n$ 次元同期式/非同期式 RTA の構成、通信距離計算および並列処理アルゴリズムが容易に導かれることを示す。また、2次元同期式 RTA と MOT (Mesh Of Trees)、Polymorphic Torus との通信距離特性の比較を行い、その性能を評価する。

## 1. はじめに

並列画像理解では、一様な像的データ (画像) を対象とした並列画像処理<sup>1)</sup>に加え、線分や領域、多面体などの幾何学的構造を持ったグラフィックデータ、さらには認識対象に関する知識を表したシンボリックなデータに対する並列処理が必要となる。われわれは、並列画像処理に適したメッシュ結合アーキテクチャに、グラフィックデータやシンボリックデータに対する並列処理機構を付加するという視点<sup>2)-4)</sup>から画像理解用並列計算機のアーキテクチャを検討している。

本論文では、並列画像理解への応用を目指した MIMD 型分散メモリ並列計算機<sup>5)</sup>の基本アーキテクチャとして、再帰トーラス結合アーキテクチャ (Recursive Torus Architecture, 以下 RTA と略す) を提案し、その構成と通信距離特性およびソート、サーチなどの基本並列アルゴリズムを示す。ここで述べる RTA は、汎用の MIMD 型分散メモリ並列計算機の抽象アーキテクチャであり、並列画像理解用の特殊な機構を備えているわけではない。

以下、まず2章では、RTA の基本的な考え方を述べたのち、同期式 RTA と非同期式 RTA の区別を示す。3章では、1次元の同期式 RTA を対象に、プロセッサ間の最短通信距離に関する特性を証明し、簡単な並列アルゴリズムを示す。4章では、3章の議論を基にして、2次元の同期式 RTA の構成、通信距離特性、並列アルゴリズムを示し、その性能を RTA と類似

したアーキテクチャを持つ MOT (Mesh Of Trees)<sup>3)</sup>および Polymorphic Torus<sup>2)</sup>と比較する。5章では、3次元以上の RTA の構成を述べたのち、3、4章で示した同期式 RTA の特性が非同期式 RTA の場合にもそのまま成り立つことを示す。6章では、RTA のハードウェアによる実現方式について考察を加える。

## 2. 再帰トーラス結合アーキテクチャの概要

行列や画像といった一様な像的データの並列処理を目指した分散メモリ型並列計算機では、CPU と局所メモリからなる PE (Processor Element) を、通信線によって2次元メッシュやトーラス状に結合したアーキテクチャがよく用いられる<sup>1),5)</sup>。しかし、メッシュやトーラス結合では、PE 間の通信距離が長い<sup>6)</sup>うえ、木やグラフといった構造を持ったデータの並列処理や再帰構造を持ったアルゴリズムの並列化は困難である。この問題に対し、MOT<sup>3)</sup>では、2分木状に結合された PE 群を縦横方向に配列することによって、各木の葉ノードに位置する PE から構成される2次元配列構造と、2分木による階層構造の融合を図っている。

本論文で提案する RTA は、トーラス状に結合された PE を結ぶ通信線上にスイッチを配置し、動的に PE 間の接続関係を変化させることによって、トーラス結合に再帰的な階層構造を埋め込もうとするアーキテクチャである。すなわち、MOT が静的結合網であるのに対し、RTA は動的結合網となっている。また、トーラス結合ではなく、PE の再帰的な正三角形結合を基にしたアーキテクチャとして、FIN<sup>6)</sup>が提案されている。

RTA の最も代表的な例である2次元同期式 RTA

† Recursive Torus Architecture by TAKASHI MATSUYAMA and MASAHITO AOYAMA (Department of Information Technology, Faculty of Engineering, Okayama University).

†† 岡山大学工学部情報工学科

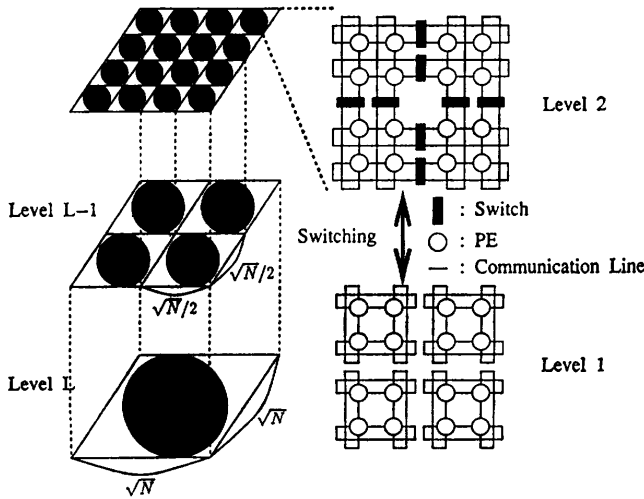


図1 2次元同期式 RTA の概念図  
Fig. 1 Concept of 2D-SRTA.

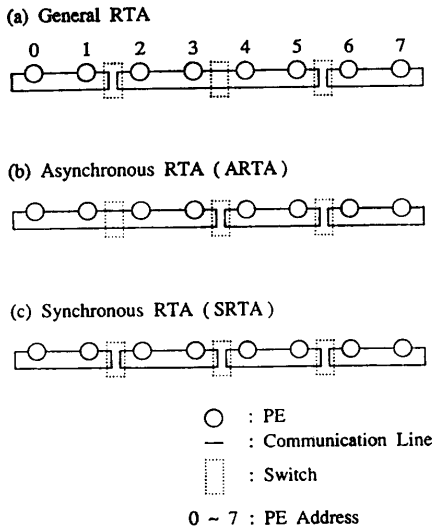


図2 1次元 RTA のクラス分け  
Fig. 2 Classification of 1D-RTA.

(2D-SRTA, 後述) の構成原理を 図1 に示す.  $N (= 2^{2L}, L$  は  $L \geq 2$  の整数) 個の PE からなる 2D-SRTA では,  $\sqrt{N} \times \sqrt{N}$  の 2次元トラスが, 最も大きなトラス結合状態であり, これをレベル  $L$  の状態にあるという. スイッチ切り換えによって, レベル  $L$  のトラスは, 4つの  $(\sqrt{N}/2) \times (\sqrt{N}/2)$  の部分トラスに分割される. 以下同様のスイッチ切り換えを再帰的に繰り返していくと, 最終的にレベル1では,  $2 \times 2$  のトラスが  $2^{L-1} \times 2^{L-1}$  個存在することになる. トラス結合状態のレベル変化は, 通信線上に配置されたスイッチの動的切り換えによって実現されており, あるレベルから任意のレベルへと直ちに状態遷移

が可能である.

RTA には, トラスの幾何学的次元による, 1次元 RTA, 2次元 RTA, ...,  $n$ 次元 RTA という分類のほか, スイッチ切り換えの同期方式による分類がある (図2). まず, 一般的な RTA では, 各スイッチの切り換えが完全に独立して行え, 図2(a)のような結合状態が実現できる. これに対して, スイッチ切り換えに制約(同期機構)を持たせたものとして, 同期式 RTA (Synchronous RTA, SRTA) と非同期式 RTA (Asynchronous RTA, ARTA) がある (厳密な定義は後述する). SRTA では, スイッチ切り換えによって生成されるすべての部分トラスが同じ構造を持つようにスイッチ群の同期が取られる (図2(c)). 一方, ARTA では, 図2(b)のように, SRTA における各部分トラスを他の部分トラスとは独立に再分割可能とすることによって, 大きさの異なるトラスの存在を許す. 以上のことから, RTA 特殊化 ARTA 特殊化 SRTA が言え, 図2(a)の結合状態は, ARTA や SRTA では実現できない.

### 3. 1次元同期式 RTA

#### 3.1 定義

定義1 [表記に関する文字の使用法]

定数...1字の大文字アルファベット

変数...1字の小文字アルファベット

関数名...小文字のアルファベット列

関係名...大文字のアルファベット列

RTA 固有の関数名...大文字で始まる小文字のアルファベット列

PE...1字の大文字ボールド体アルファベット

定義2 [divide, quotient, residue]

$x$  は  $x \geq 0$  の整数,  $y$  は  $y = 2^k (k: k \geq 0$  の整数) を満たす整数とする.

$$\text{divide}(x, y) \stackrel{\text{def}}{=} (x \text{ を } y \text{ で割ったときの商})$$

$$\text{quotient}(x, y) \stackrel{\text{def}}{=} \text{divide}(x, y) \times y$$

$$\text{residue}(x, y) \stackrel{\text{def}}{=} x \bmod y \text{ (mod は非負最小剰余)}$$

定義3 [1次元 RTA (1D-RTA) の構成]

1D-RTA とは, 2本の通信線を持つ  $N$  個 ( $N = 2^L, L$  は  $L \geq 2$  の正整数の定数) の PE が1次元トラス (リング) 状に結ばれた結合形態を持ち, 図3下図のように通信線上にスイッチが配置されたアーキテクチャである. 1D-RTA では, 左から  $i$  番目 ( $0 \leq i \leq N-1$ ) の PE のアドレスを  $i$  とする.

Switching Level

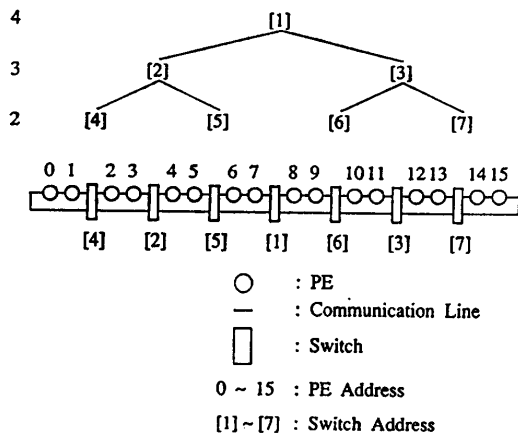


図3 1次元 RTA の構成  
Fig. 3 Architecture of 1D-RTA.

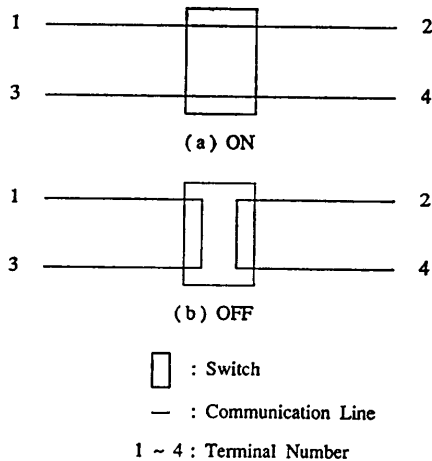


図4 スイッチの状態  
Fig. 4 State of switch.

スイッチは、図3下図のように、アドレスが  $2m-1$  と  $2m$  ( $1 \leq m \leq (N/2)-1$ ) の PE の間に配置され、スイッチの総数  $M$  は、 $M=(N/2)-1$  となる。スイッチには4つの端子があり(図4)、端子1, 2はそれぞれアドレスが  $2m-1$ ,  $2m$  の隣接する PE に接続され、端子3(4)は左(右)隣のスイッチの端子4(3)に接続される。ただし、左端のスイッチの端子3はアドレス0の PE に、右端のスイッチの端子4はアドレス  $N-1$  の PE にそれぞれ接続される。また、アドレスが  $2m$  と  $2m+1$  の PE はスイッチを介さずに直結される。

スイッチが ON の状態とは図4(a)の場合、OFF の状態とは同図(b)の場合をいい、ON $\leftrightarrow$ OFF のスイッチ切り換えは瞬時に行えるとする。単に 1D-RTA という場合は、各スイッチの状態変化が他とは独立に

行えるものを言う。

#### 定義4 [スイッチのアドレス付け]

スイッチのアドレスには、整数値  $j$  ( $1 \leq j \leq M$ ) を使うが、PE のアドレスとの混同を避けるため  $[j]$  と表記する。スイッチのアドレス付けは次のように行う。まず、図3上図に示すように、 $M$ 個のノードからなる完全二分木に対して、根ノードを1とした番号を順に付ける。番号付けされた二分木を中順走査したときのノード番号の並びを求め、それを 1D-RTA を左から走査したときのスイッチアドレスの並びとする。

#### 定義5 [スイッチングレベル]

図3上図の番号付二分木において、根ノードのレベルを  $L$  (PE の総数を  $N$  個としたとき  $N=2^L$ )、その子ノードのレベルを  $L-1, \dots$ 、葉ノードをレベル2としたとき、番号が  $j$  のノードが属するレベルのことをスイッチ  $[j]$  のスイッチングレベル  $l_{[j]}$  という。

#### 定義6 [1次元同期式 RTA (1D-SRTA)]

1D-RTA に、次のようなスイッチ切り換えの同期機構を加えたものを 1D-SRTA という。1D-SRTA において許されるスイッチ切り換えは、ある正整数  $l$  ( $1 \leq l \leq L$ ) に対して、スイッチングレベル  $l_{[j]}$  が  $l < l_{[j]}$  となるすべてのスイッチ  $[j]$  を同時に OFF にするとともに、 $l \geq l_{[j]}$  となるすべてのスイッチ  $[j]$  を同時に ON にするものに限る。

1D-SRTA の結合状態は、正整数  $l$  によって完全に指定することができ、 $l$  を 1D-SRTA の結合状態を表すレベルという。すなわち、1D-SRTA では、レベル  $L$  のときに  $N$  個の PE よりなる最も大きなトーラスが1つ存在する結合状態、レベル1のときに2個の PE よりなる最も小さな部分トーラスが  $N/2$  個存在する結合状態である。

#### 定義7 [1次元非同期式 RTA (1D-ARTA)]

1D-ARTA は、1D-SRTA における各部分トーラスを、他の同レベルの部分トーラスとは独立に再分割可能とするもので、以下のようにスイッチを同期切り換えする。1D-ARTA において、ON 状態にあるスイッチ  $[j]$  を OFF 状態にするには、スイッチの親子関係を表す二分木(図3)において、スイッチ  $[j]$  の祖先であるすべてのスイッチを同時に OFF にする(既に OFF 状態である)必要がある。逆に、OFF 状態にあるスイッチ  $[j]$  を ON 状態にするには、スイッチ  $[j]$  のすべての子孫であるスイッチを同時に ON にする(既に ON 状態である)必要がある。明らかに、1D-SRTA におけるスイッチの同期切り換え法は、

1D-ARTA の場合の特殊例となっており、前者は後者に含まれる。

**3.2 1次元同期式 RTA の通信距離特性**

ここでは、1D-SRTA の通信距離特性に関する定理を示すが、詳しい証明は文献 7) を参照していただきたい。

**定義 8 [PE 間の通信距離]**

2つの PE 間をデータが移動するときに通る通信線の数 (データが経由する PE の個数+1) の最小値をその PE 間の通信距離とする。つまり、スイッチ内にはラッチはなく、ON のときは左右、OFF のときは上下の通信線がそれぞれ直結されると考える。

**定義 9 [neighbor]**

$A$  を  $0 \leq A \leq N-1$  の整数、 $l$  は  $1 \leq l \leq L$  を満たす整数とすると、

$$\begin{aligned} neighbor(A, l) &= \{quotient(A, 2^l) + residue(A \pm 1, 2^l)\} \\ &\text{ただし, } \{\dots\} \text{は集合を表す.} \end{aligned}$$

この集合を返す関数  $neighbor(A, l)$  は、レベル  $l$  の結合状態にある 1D-SRTA において、アドレスが  $A$  である PE と通信線で結ばれている PE のアドレスを求めるもので、 $neighbor(A, 1)$  はスイッチを介さずに直接通信線で  $A$  と結ばれている PE (図 3 参照) のアドレス (1 個) を要素とし、すべての  $l$  に対して、 $neighbor(A, 1) \subset neighbor(A, l)$  が成り立つ。

**定義 10 [ADJACENT]**

$A, B$  は  $0 \leq A \leq N-1, 0 \leq B \leq N-1$  の整数、 $l$  は  $1 \leq l \leq L$  を満たす整数とすると、 $B \in neighbor(A, l)$  あるいは  $A \in neighbor(B, l)$  のとき、 $ADJACENT_l(A, B)$  の関係が成り立つといい、 $ADJACENT_l(A, B)$  と  $ADJACENT_l(B, A)$  は同値である。また、 $ADJACENT_l(A, B)$  が成り立たないことを、 $NOT-ADJACENT_l(A, B)$  と記す。

**定義 11 [Distance]**

2つの PE,  $S$  と  $E$  の間の通信距離を  $D$  とすると、

$$\begin{aligned} Distance(S, E) &= D \\ Distance(S, E) &= Distance(E, S) \\ Distance(S, E) &= 1 \text{ のとき, } ADJACENT_l(S, E) \end{aligned}$$

の関係が成り立つ  $l$  が存在する。

**定理 1**

レベル  $l (2 \leq l \leq L)$  の結合状態において存在する、ある部分トーラス  $T$  において、 $T$  に含まれる任意の PE を  $X$  とする。このとき、

$$\begin{aligned} X_1 &= quotient(X, 2^l) \\ X_2 &= quotient(X, 2^l) + 2^{l-1} - 1 \\ X_3 &= quotient(X, 2^l) + 2^{l-1} \\ X_4 &= quotient(X, 2^l) + 2^l - 1 \end{aligned}$$

となる 4 つの PE,  $X_i (1 \leq i \leq 4)$  は、次式を満たす。

$$\begin{aligned} Distance(X_1, X_2) &= 1 & Distance(X_1, X_4) &= 1 \\ Distance(X_2, X_3) &= 1 & Distance(X_3, X_4) &= 1 \\ Distance(X_1, X_3) &= 2 & Distance(X_2, X_4) &= 2 \end{aligned}$$

また、 $l=1$  のときは、

$$X_1 = X \quad X_2 \in neighbor(X, 1)$$

となる 2 つの PE,  $X_1, X_2$  は、 $Distance(X_1, X_2) = 1$  を満たす。ここで、 $l=1$  の場合に  $X_3, X_4$  を表記していないのは、 $X_3 = X_1, X_4 = X_2$  となるからである。

(証明略)

$X_1, X_2, X_3, X_4$  は、特定の PE,  $X$  に依存するのではなく、レベル  $l$  の部分トーラス  $T$  においてユニークに定まる PE で図 5 のような接続関係にあり、

$$\begin{aligned} ADJACENT_l(X_1, X_2) & \\ ADJACENT_l(X_2, X_3) & \\ NOT-ADJACENT_l(X_1, X_3) & \\ NOT-ADJACENT_l(X_3, X_4) & \\ ADJACENT_{(l-1)}(X_1, X_2) & \\ ADJACENT_{(l-1)}(X_3, X_4) & \\ NOT-ADJACENT_{(l-1)}(X_1, X_4) & \\ NOT-ADJACENT_{(l-1)}(X_2, X_3) & \end{aligned}$$

の関係が成り立つ。また、 $X_1, X_4$  は、トーラス  $T$  を含むレベル  $l+1$  のトーラスにおいて、 $X_1^{l+1}, X_2^{l+1}$  の組か、 $X_3^{l+1}, X_4^{l+1}$  の組になる。すなわち、レベル  $l$  の 2 つのトーラス  $T_1, T_2$  を併合してレベル  $l+1$  のトーラスを構成する場合、 $T_1, T_2$  内においてそれぞれ  $X_1$  と  $X_4$  に対応する 4 つの PE 間で通信線の切り換えが行われることになる。

**定義 12 [move, Path, Length]**

2つの PE,  $A$  と  $B$  の間に  $ADJACENT_l(A, B)$

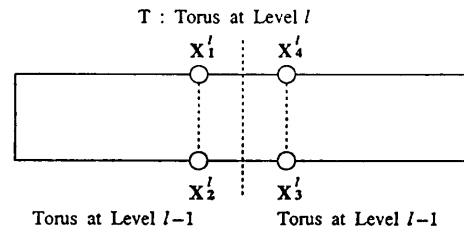


図 5 トーラス  $T$  に対する  $X_i (1 \leq i \leq 4)$  の幾何学的配置  
Fig. 5 Geometric arrangement of  $X_i (1 \leq i \leq 4)$  in torus  $T$ .

の関係が成り立つとき、レベル  $l$  において  $A, B$  を結ぶ通信線を用いた  $A$  から  $B$  へのデータ転送を  $move_i(A, B)$  と表記する。また、2つの PE,  $S$  と  $E$  の間の通信パスを

$$Path: L_1 m_1 L_2 m_2 \dots L_K m_K$$

と表す。ただし、

$$L_i: 1D-SRTA \text{ の結合状態のレベル } (1 \leq i \leq K)$$

$$m_i = move_{L_i}(S_i, E_i) \quad (1 \leq i \leq K)$$

$$E_j = S_{j+1} \quad (1 \leq j \leq K-1) \quad S_1 = S \quad E_K = E$$

$Length(Path) = Distance(S, E)$  となる  $S$  と  $E$  の間の通信パス  $Path$  を最短パスという。ここで、

$$Length(L_1 m_1 L_2 m_2 \dots L_K m_K) = K$$

### 定義 13 [Shortest-path]

関数 *Shortest-path* は、2つの PE,  $S$  と  $E$  の間の最短パスを求める。

$$Shortest-path(S, E)$$

$$= (Length(Path))$$

$$= Distance(S, E) \text{ を満たす } S, E \text{ 間の } Path$$

$$Shortest-path(S, E) = Shortest-path(E, S)$$

### 定理 2

レベル  $l$  ( $2 \leq l \leq L$ ) の結合状態において存在する、ある部分トラス  $T$  を考え、それを2等分したレベル  $l-1$  の2つの部分トラスのうち、 $T$  における  $X_1^l$  と  $X_2^l$  を含むものを  $T_1$ ,  $X_3^l$  と  $X_4^l$  を含むものを  $T_2$  とする。 $T_1$  内の PE,  $Y$  から  $T_2$  内の PE,  $Z$  へ至るすべての最短パスは、 $X_1^l$  と  $X_4^l$ , または、 $X_2^l$  と  $X_3^l$  の間の通信線を必ず通る。すなわち、

$$\begin{aligned} Length(Shortest-path(Y, Z)) &= Length(Shortest-path(Y, X_1^l)) \\ &\quad + Length(Shortest-path(X_1^l, X_4^l)) \\ &\quad + Length(Shortest-path(X_4^l, Z)) \end{aligned}$$

または、

$$\begin{aligned} Length(Shortest-path(Y, Z)) &= Length(Shortest-path(Y, X_2^l)) \\ &\quad + Length(Shortest-path(X_2^l, X_3^l)) \\ &\quad + Length(Shortest-path(X_3^l, Z)) \end{aligned}$$

(証明略)

### 補題

レベル  $l$  ( $2 \leq l \leq L$ ) の結合状態において存在する、ある部分トラス  $T$  を考える。 $T$  に含まれる任意の PE,  $Y$  から、 $T$  における  $X_i^l$  ( $1 \leq i \leq 4$ ) への通信距離を  $Distance(Y, X_i^l)$  ( $1 \leq i \leq 4$ ) とする。このうち、 $Distance(Y, X_i^l)$  の値が最小であるとしても一般性を失わない。その値を  $d$  とすると、

$$Distance(Y, X_2^l) = d + 1$$

$$Distance(Y, X_3^l) = d + 2$$

$$Distance(Y, X_4^l) = d + 1$$

(証明略)

### 定理 3

レベル  $l$  ( $3 \leq l \leq L$ ) の結合状態において存在する、ある部分トラス  $T$  に含まれる PE,  $Y$  から  $T$  における  $X_i^l$  ( $1 \leq i \leq 4$ ) への通信距離を  $Distance(Y, X_i^l)$  ( $1 \leq i \leq 4$ ) とし、 $T$  を分割してできるレベル  $l-2$  の4つの部分トラスを  $T_1, T_2, T_3, T_4$  としたとき、 $Distance(Y, X_i^l)$  が最小となる  $X_i^l$  を含むレベル  $l-2$  の部分トラス  $T_i$  に  $Y$  が含まれる。 $l=2, 1$  のときは、 $Y = X_i^l$  となる。(証明略)

### 3.3 PE 間の通信距離計算アルゴリズム

1D-SRTA において、任意の2つの PE,  $S, E$  間の通信距離計算アルゴリズムは、次のように構成される。

[STEP 1] レベル  $L$  から始め、トラスを順次分割しながら  $S$  と  $E$  が初めて異なった部分トラスに属すようになるレベル  $l^*$  を求める。レベル  $l^*+1$  において  $S$  と  $E$  が共に含まれる部分トラスを  $T$  とすると、 $S, E$  間の通信距離は部分トラス  $T$  内のみで考えれば良い。

(妥当性の証明)

定理 1 およびその後示した性質より明らか。

[STEP 2]  $T$  を4分割してできるレベル  $l^*-1$  の部分トラスのうち  $S$  が属すものを  $T_1$ ,  $E$  が属すものを  $T_2$  とする。 $T$  における  $X_i^{l^*+1}$  ( $1 \leq i \leq 4$ ) のうち  $T_1$  に属すものを  $X_S^{l^*+1}$ ,  $T_2$  に属すものを  $X_E^{l^*+1}$  とする。 $X_S^{l^*+1}, X_E^{l^*+1}$  間の通信距離を定理 1 より求める。

[STEP 3] レベル  $l^*-1$  の部分トラス  $T_1$  および  $T_2$  を対象に、 $X_S^{l^*+1}$  を新たな  $E$ ,  $X_E^{l^*+1}$  を新たな  $S$  として、 $S$  と新たな  $E$ ,  $E$  と新たな  $S$  に対してそれぞれ [STEP 1] と [STEP 2] の計算を再帰的に繰り返し、各再帰計算の [STEP 2] で求めた通信距離を積算したものが、 $S, E$  間の通信距離となる。ただし、[STEP 2] において、 $X_S^{l^*+1} = S$  あるいは  $X_E^{l^*+1} = E$  となったときには、 $T_1$  あるいは  $T_2$  を対象にした再帰計算は停止する。

(妥当性の証明)

これは、次の二通りの場合を確認すれば良い。

(1)  $X_S^{l^*+1}$  と  $X_E^{l^*+1}$  が、 $T$  における  $X_i^{l^*+1}$  と

$X_4^{l^*+1}$  または、 $X_2^{l^*+1}$  と  $X_3^{l^*+1}$  の関係にある場合：  
 定理 2 より、レベル  $l^*-1$  の部分トーラス  $T_1, T_2$  においてそれぞれ  $S$  と  $X_S^{l^*+1}$ 、 $E$  と  $X_E^{l^*+1}$  の通信距離を求め、それと  $X_S^{l^*+1}$  と  $X_E^{l^*+1}$  の間の通信距離を足したものが  $S, E$  間の通信距離となる。

(2)  $X_S^{l^*+1}$  と  $X_E^{l^*+1}$  が、 $T$  における  $X_1^{l^*+1}$  と  $X_3^{l^*+1}$  または  $X_2^{l^*+1}$  と  $X_4^{l^*+1}$  の関係にある場合：  
 $X_S^{l^*+1} = X_1^{l^*+1}$ 、 $X_E^{l^*+1} = X_3^{l^*+1}$ 、 $S$  と  $E$  の間の最短パスが  $X_1^{l^*+1}$  と  $X_4^{l^*+1}$  を通ると仮定しても一般性を失わない。定理 2 より、

$$\begin{aligned} & \text{Length}(\text{Shortest-path}(S, E)) \\ &= \text{Length}(\text{Shortest-path}(S, X_1^{l^*+1})) \\ & \quad + \text{Length}(\text{Shortest-path}(X_1^{l^*+1}, X_4^{l^*+1})) \\ & \quad + \text{Length}(\text{Shortest-path}(X_4^{l^*+1}, E)) \end{aligned}$$

となる。定理 3 と補題より、 $\text{Distance}(E, X_3^{l^*+1}) = d$  とすると、 $\text{Distance}(E, X_4^{l^*+1}) = d+1$  となるので、上式の右辺第 3 項は、

$$\begin{aligned} & \text{Length}(\text{Shortest-path}(X_4^{l^*+1}, E)) \\ &= 1 + \text{Length}(\text{Shortest-path}(X_3^{l^*+1}, E)) \end{aligned}$$

となる。したがって、 $X_1^{l^*+1}$  と  $X_4^{l^*+1}$  の通信距離 1 と、 $X_4^{l^*+1}$  と  $X_3^{l^*+1}$  の通信距離 1 および、レベル  $l^*-1$  の部分トーラス  $T_1, T_2$  において  $S$  と  $X_S^{l^*+1}$  ( $= X_1^{l^*+1}$ )、 $E$  と  $X_E^{l^*+1}$  ( $= X_3^{l^*+1}$ ) の通信距離を求め、それらを足したものが  $S, E$  間の通信距離となる。

**定理 4**

PE の総数が、 $N$  個の 1D-SRTA における Diameter, Average は、表 1 のようになる。ここで、Diameter は PE 間の通信距離の最大値、Average は通信距離の平均値を表す。

**3.4 基本並列アルゴリズム**

一般に RTA では、任意の PE 間で自由に (並列的に) 通信が行えるが、そうした無制限な通信方式を基に効率的な並列アルゴリズムを構成するのは困難である。これは、任意の PE 間での自由な通信を仮定すると、宛先 PE のアドレスが付けられた多数のペケットが PE 間で並列して転送されることになり、そうしたペケットの並列転送を効率よく行うためのスイッチ切

表 1 1D-SRTA における Diameter, Average  
 Table 1 Diameter and Average of 1D-SRTA.

	Diameter	Average
1次元同期式 RTA	$2 \log_2 N - 2$	$\frac{N(2 \log_2 N - 3) + 4}{2(N-1)}$

り換えアルゴリズムを考えるのがむずかしいことによる。そこで、RTA による並列アルゴリズムでは、スイッチの切り換えに同期した組織的な並列データ転送を考える必要がある。ここでいう組織的並列データ転送とは、

- (1) スイッチを切り換え RTA をある結合状態にする。
  - (2) その状態で各 PE が並列的に隣接する PE にデータを転送する。
  - (3) スイッチを切り換え RTA を新たな結合状態にする。
- という動作を繰り返すものをいう。

ここでは、1D-SRTA を対象に、最も基本的な組織的並列データ転送手順を示し、それを用いた簡単な並列アルゴリズムの時間計算量を示す。計算量は、1単位のデータが隣接する PE 間で転送されたときに 1 と数えるものとする。すなわち、通信時間 (= <通信パスの長さ> \* <通信データ量>) を計算量として評価する。

**3.4.1 基本並列データ転送手順**

PE の総数を  $N$ 、初期状態として各 PE に 1 つずつデータが与えられているとする。

1D-SRTA における基本並列データ転送手順とは、図 6 に示すように、トーラスの結合状態のレベル  $l$  を  $1, 2, 3, \dots, L$  に順次設定し、各レベル  $l$  においてスイッチングレベル  $l+1$  のスイッチの端子 1(2) に接続

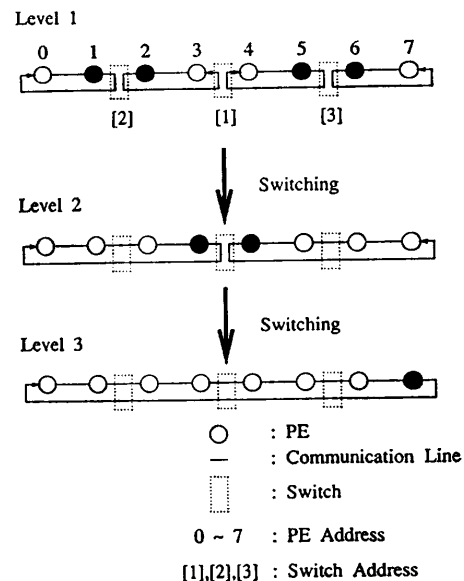


図 6 1次元 SRTA における基本並列データ転送手順  
 Fig. 6 Fundamental parallel communication scheme in 1D-SRTA.

している PE は、それぞれスイッチを介して右(左)隣の PE にデータを同時に転送する(図では黒丸の PE が矢印の方向にデータを送る)。データ転送の後、データを送った PE はアイドル状態になり、データを受け取った PE が演算  $O$  を実行する。演算終了後、結合状態のレベルが  $l+1$  に切り換えられ、最後は、レベル  $L$  でアドレス  $N-1$  の PE が右隣(アドレス  $0$ ) の PE にデータ転送する。

### 3.4.2 並列アルゴリズム

#### 最大値・最小値・総和のアルゴリズム

基本並列データ転送手順において、演算  $O$  として比較演算あるいは加算を用い、2つのデータの大きい方、小さい方、あるいは和の値を隣接する PE に転送することによって、 $N$  個のデータの最大値、最小値、総和を計算する  $O(\log_2 N)$  の並列アルゴリズム(結果はアドレス  $0$  の PE に得られる)が構成される。

#### 並列ソーティング・アルゴリズム

基本並列データ転送手順において、レベル  $l$  でデータを受け取った PE が、自分が保持しているデータと受け取ったデータからソート列を作り、それをレベル  $l+1$  の状態において隣接する PE に転送することによって、 $O(N)$  の並列ソーティング・アルゴリズム(結果はアドレス  $0$  の PE に得られる)が構成できる。

## 4. 2次元同期式 RTA

### 4.1 定義

**定義 14** [2次元 RTA(2D-RTA)の構成]

2D-RTA とは、同じ大きさの 1D-RTA を水平、垂直の2方向に配列したもので、各 PE は4本の通信線を持ち、水平、垂直方向の2つの 1D-RTA に同時に属することになる(図7)。2D-RTA は  $\sqrt{N} \times \sqrt{N}$  個 ( $\sqrt{N}=2^L$ ,  $L$  は  $L \geq 2$  の正整数の定数) の PE で構成され、上から  $i$  行目 ( $0 \leq i \leq \sqrt{N}-1$ )、左から  $j$  列目 ( $0 \leq j \leq \sqrt{N}-1$ ) の PE のアドレスを  $(i, j)$  とする。スイッチの配置は水平、垂直方向とも 1D-RTA のときと全く同じで、PE の総数が  $N$  の 2D-RTA では、スイッチの総数は、 $\sqrt{N}(\sqrt{N}-2)$  となる。以下では、水平、垂直方向の1つの 1D-RTA に含まれるスイッチの総数を  $M(=(\sqrt{N}/2)-1)$  とする。

**定義 15** [スイッチのアドレス付け]

スイッチのアドレスには、整数値  $u, v, w$  からなる3つ組  $[u, v, w]$  を用いる。 $u$  は、そのスイッチが含まれる 1D-RTA が水平、垂直方向いずれのものであるかを識別するために使用し、 $u=0$  は水平、 $u=1$  は

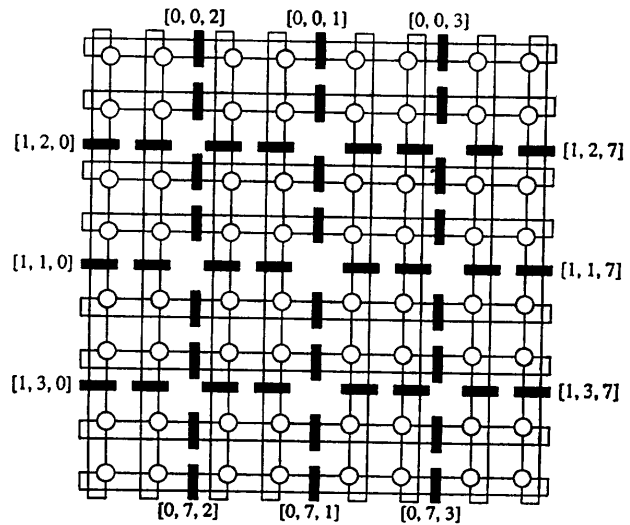


図7 2次元 RTA の構成 (PE: 64 個)

Fig. 7 Architecture of 2D-RTA (PE: 64).

垂直方向の 1D-RTA を表す。 $v$  と  $w$  は、スイッチの 2次元アドレスを表し、次のように定義される。図7に示すように、水平方向の 1D-RTA 内のスイッチに対しては、 $[u, v, w]=[0, (\text{その 1D-RTA が上から何行目かを表す番号}), (\text{その 1D-RTA 内のスイッチのアドレス})]$  ( $0 \leq v \leq \sqrt{N}-1, 1 \leq w \leq M$ )、垂直方向の 1D-RTA 内のスイッチに対しては、 $[u, v, w]=[1, (\text{その 1D-RTA 内におけるスイッチのアドレス}), (\text{その 1D-RTA が左から何列目かを表す番号})]$  ( $1 \leq v \leq M, 0 \leq w \leq \sqrt{N}-1$ ) とアドレスを付ける。なお、1D-RTA 内のスイッチアドレスは、定義4で与えたものをそのまま用いる。

**定義 16** [スイッチングレベル]

スイッチ  $[u, v, w]$  が属す水平あるいは垂直方向の 1D-RTA 内でのそのスイッチのスイッチングレベルを、2D-RTA におけるスイッチ  $[u, v, w]$  のスイッチングレベル  $l_{[u, v, w]}$  とする。

**定義 17** [2次元同期式 RTA (2D-SRTA)]

2D-RTA に、次のようなスイッチ切り換えの同期機構を加えたものを 2D-SRTA という。2D-SRTA において許されるスイッチ切り換えは、ある正整数  $l$  ( $1 \leq l \leq L$ ) に対して、スイッチングレベル  $l_{[u, v, w]}$  が  $l < l_{[u, v, w]}$  となるすべてのスイッチ  $[u, v, w]$  を同時に OFF にするとともに、 $l \geq l_{[u, v, w]}$  となるすべてのスイッチ  $[u, v, w]$  を同時に ON にするものに限る。2D-SRTA の結合状態は、正整数  $l$  によって完全に指定でき、 $l$  を 2D-SRTA の結合状態を表すレベルという。

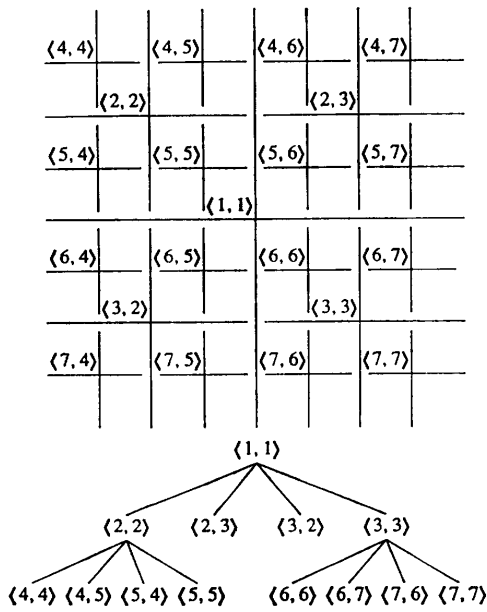


図 8 スイッチングクラス (PE: 256 個, 上図の各十字は, 水平, 垂直方向のスイッチの並びを表す. 図 7 参照).  
 Fig. 8 Switching class (PE: 256, each cross in the upper figure represents a group of switches, see Fig. 7).

定義 18 [スイッチングクラス]

図 7 に示すような 2D-RTA の構成図において, 各  $v$  に対して  $[1, v, j] (0 \leq j \leq \sqrt{N}-1)$  のスイッチ群を通る水平方向の直線  $line_v$  と, 各  $w$  に対してスイッチ  $[0, i, w] (0 \leq i \leq \sqrt{N}-1)$  のスイッチ群を通る垂直方向の直線  $line_w$  を考え,  $l_{[1, v, *]} = l_{[0, *, w]}$  を満たす (\* はスイッチングレベルには無関係)  $line_v$  と  $line_w$  の交点をスイッチングクラス  $\langle v, w \rangle$  という (図 8). 各交点  $\langle v, w \rangle$  から見て,  $line_v$  と  $line_w$  に沿って上, 下, 左, 右各々  $2^{(l_{[1, v, *]}-1)}$  個のスイッチ群をスイッチングクラス  $\langle v, w \rangle$  に属すスイッチ群という. また, 図 8 下図に示すように, スイッチングクラス  $\langle v, w \rangle$  をノードとする,  $(4^{l-1}-1)/3$  個のノードからなる完全 4 分木を構成する.

定義 19 [2次元非同期式 RTA (2D-ARTA)]

2D-ARTA では, 以下のようにスイッチを同期切り換えする. 2D-ARTA において, ON 状態にあるスイッチ  $[u, v, w]$  を OFF 状態にするには, 図 8 の 4 分木において, スイッチ  $[u, v, w]$  の属すスイッチングクラスのすべての祖先スイッチングクラスに属すスイッチを同時に OFF にする (既に OFF 状態である) 必要がある. 逆に, OFF 状態にあるスイッチ  $[u, v, w]$  を ON 状態にするには, スイッチ  $[u, v, w]$

の属すスイッチングクラスのすべての子孫スイッチングクラスに属すスイッチを同時に ON にする (既に ON 状態である) 必要がある. 2D-SRTA におけるスイッチの同期切り換え法は, 2D-ARTA の場合の特殊例となっており, 前者は後者に含まれる.

定理 5 [2次元通信パスの 1 次元通信パスへの分割]

2D-SRTA における任意の 2 つの PE,  $S$  と  $E$  の間の通信パスを

$$2D-Path: L_1 m_1 L_2 m_2 \dots L_k m_k$$

とする. ここで,  $L_i$  は 2D-SRTA の結合状態のレベル,  $m_i$  はレベル  $L_i$  において隣接している 2 つの PE 間でのデータの移動を表し,  $2D-Path$  の長さ  $2D-Length(2D-Path)$  を  $K$  とする. 2D-SRTA の構成から明らかのように,  $m_i$  は, 水平, 垂直方向移動の 2 種類に分けられる. そこで,  $2D-Path$  を, 水平方向の  $m_i(m_i^h)$  のみを順に並べたパス  $Horizontal-path$  と, 垂直方向の  $m_i(m_i^v)$  のみを順に並べたパス  $Vertical-path$  に分ける.

$$\begin{cases} Horizontal-path: L_1^h m_1^h L_2^h m_2^h \dots L_H^h m_H^h \\ Vertical-path: L_1^v m_1^v L_2^v m_2^v \dots L_V^v m_V^v \end{cases}$$

ここで,  $L_i^h, L_i^v$  はそれぞれの  $m_i^h, m_i^v$  に対応する  $2D-Path$  中のレベルを表す. このとき,  $Horizontal-path$  と  $Vertical-path$  はそれぞれ,  $S(E)$  を含む水平方向の 1D-SRTA 上の  $S(E)$  を始点 (終点) とする 1 次元の通信パス,  $E(S)$  を含む垂直方向の 1D-SRTA 上の  $E(S)$  を終点 (始点) とする 1 次元の通信パスと考えることができ,

$$\begin{aligned} 2D-Length(2D-Path) \\ = Length(Horizontal-path) \\ + Length(Vertical-path) \end{aligned}$$

が成り立つ. (証明略)

定理 6 [2D-SRTA における通信距離]

定理 5 より, 2D-SRTA における任意の 2 つの PE,  $S$  と  $E$  の間の通信距離を求めるには,  $S$  を含む水平 (垂直) 方向の 1D-SRTA と  $E$  を含む垂直 (水平) 方向の 1D-SRTA に同時に含まれる PE を  $M$  とし, 水平 (垂直) 方向の 1D-SRTA 内での  $S$  と  $M$  の間の通信距離と, 垂直 (水平) 方向の 1D-SRTA 内での  $M$  と  $E$  の間の通信距離を, 前述のアルゴリズムによってそれぞれ求め, それらを加えればよい.

4.2 基本並列データ転送手順と並列アルゴリズム

2D-SRTA における PE の総数を  $N$ , 初期状態として各 PE に 1 つずつデータが与えられているとする. 2D-SRTA における基本並列データ転送手順で



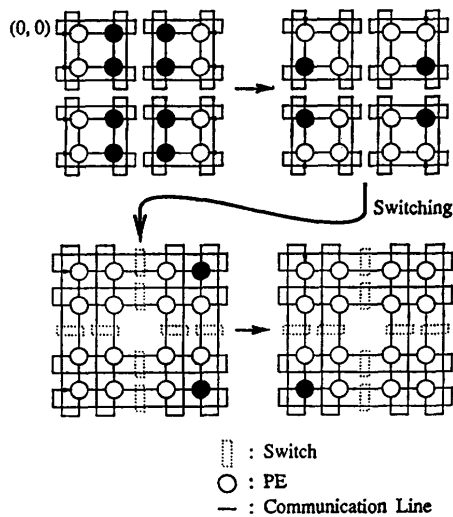


図9 2次元 SRTA における基本並列データ転送手順  
Fig. 9 Fundamental parallel communication scheme in 2D-SRTA.

は、図9に示すように、トーラスの結合状態のレベル  $l$  を  $1, 2, 3, \dots, L$  に順次設定し、各レベル  $l$  において以下のデータ転送、演算を実行する。まず、水平方向の各 1D-SRTA においてスイッチングレベル  $l+1$  のスイッチの端子 1(2)に接続している PE が、それぞれスイッチを介して右(左)隣の PE にデータを同時に転送する。データ転送の後、データを送った PE はアイドル状態になり、データを受け取った PE は演算  $O$  を行う。次に、垂直方向の各 1D-SRTA において、スイッチングレベル  $l+1$  のスイッチの端子 1(2)に接続している PE が、それぞれスイッチを介して下(上)隣の PE にデータを同時に転送する。データ転送の後、データを送った PE はアイドル状態になり、データを受け取った PE は演算  $O$  を行う。

1D-SRTA の場合と同様に、この基本並列データ転送手順を用いることにより、最大値・最小値・総和および並列ソーティングのアルゴリズムが構成でき、前者の時間計算量は  $O(\log_2 N)$ 、後者は  $O(N)$  となる。

る。

### 4.3 性能比較

2D-SRTA と類似したアーキテクチャを持つものとして、2次元格子状に配列された PE の上に完全2分木状の静的通信網を付加した Mesh Of Trees (MOT) と、2次元トーラス結合に動的な通信線切り換え機構を付加した Polymorphic Torus (PT)<sup>2)</sup>がある。PE 間通信距離および基本並列アルゴリズムに関するこれらのアーキテクチャの性能比較を表2および表3に示す。(表2では  $N$  は PE の個数、表3では処理対象データの個数を表す。)

まず、MOT と 2D-SRTA を比較すると、表2より両者は Diameter, Average とともにオーダは同じであるが、2D-SRTA の Average は MOT の約  $1/2$  となる。さらに、2D-SRTA は MOT に比べ、次の点で優れている。

(1) 表2中の  $N$  は、2D-SRTA の場合はすべての PE の数であるのに対し、MOT の場合は2次元格子状に配列された2分木の葉ノードにあたる PE の総数を表し、これらほかにデータ転送を中継するための PE が  $2\sqrt{N}(\sqrt{N}-1)$  個余分に必要となる。

(2) MOT は、完全2分木を基に構成されているため、2次元格子状上で隣接する PE 間で通信を行う場合でも、2分木の1つ上の中間ノードを経由しなければならない。画像処理や行列計算のように隣接する PE 間での通信頻度が高いアルゴリズムでは性能が落ちる。

表3 並列アルゴリズムの速度の比較  
Table 3 Speed evaluation of parallel algorithms.

	最大値・最小値・総和	ソーティング
2次元同期式 RTA	$O(\log_2 N)$	$O(N)$
MOT	$O(\log_2 N)$	$O(\log_2 N)$
Polymorphic Torus	$O(\log_2 N)$	$O(N)$

表2 通信距離の比較  
Table 2 Evaluation of communication distance.

	Diameter	Diameter の関係にある PE の組	Average
2次元同期式 RTA	$4\{(\log_2 \sqrt{N})-1\}$	8	$[N\{(\log_2 N)-3\}+4\sqrt{N}]/(N-1)$
MOT	$4 \log_2 \sqrt{N}$	$N^2/8$	$[2N\{(\log_2 N)-2\}+4\sqrt{N}]/(N-1)$
Polymorphic Torus	2	$N(\sqrt{N}-1)^2/2$	$2\sqrt{N}/(\sqrt{N}+1)$

MOT における  $N$  は leaf processor の総数

一方、表3を見ると、MOTによるソーティングが他のものより優れているが、MOTでは $N$ 個のデータの処理を行うのに $N^2$ 個のPEが必要となるという問題がある。

2D-SRTAとPTとの比較を行うと、表2より、PTの通信距離が他のものより非常に短いことがわかる。これは、PTではshort-circuitというスイッチ切り換え機構によって、1つの行あるいは列内の任意のPEを直接通信線でつなぐことが可能であるためである。しかし、並列アルゴリズムではこうした1対1の通信をそのまま使うことはできず、2D-SRTAの場合と同様に組織的な並列データ転送を行う必要がある。このため、表3ではPTと2D-SRTAとは同じ性能になっている。

2D-SRTA, MOT, PTはいずれもメッシュ結合に付加機能を付けたEnhanced Meshといえる。Enhanced Meshとしては、バス結合やブロードキャスト機能を備えたものもあり<sup>3),8)</sup>、最大(小)値・総和の計算が $O(N^{1/6})$ で実行できるものもある<sup>9)</sup>。現在並列画像理解という観点から、こうしたEnhanced MeshとRTAとの性能比較を行っており、その結果は稿を改めて報告する予定である。

### 5. $n$ 次元RTAへの拡張と非同期式RTAの通信距離特性

4章で述べた、1D-RTAに基づく2D-RTAの構成法と、2次元通信パスの1次元通信パスへの分割法、2D-SRTAにおける基本並列データ転送手順を基にすることによって、 $n$ 次元RTAの構成、 $n$ D-SRTA上の通信パスの1次元通信パスへの分割、 $n$ D-SRTA上での基本並列データ転送手順が直ちに実現できる。

#### 定理7

$N$ 個のPEより構成される $n$ D-SRTAの通信距離のDiameterは、 $2(\log_2 N) - 2n$ となる。

また、これまでの議論ではすべてSRTAを対象としてその通信距離特性を述べてきたが、次の定理よりARTAに対しても同様の性質が成り立つことが分かる。

#### 定理8

1対1のPE間の通信においては、通信距離特性、通信経路分割に関して、SRTAとARTAは同等である。

#### (証明)

ARTAの定義より、ARTAはSRTAを含んでお

り、SRTAにおいて可能なすべてのスイッチの同期切り換えはARTAにおいても可能である。逆に、3, 4章で述べた1対1のPE間の通信距離に関する議論では、他のPE間の通信との整合性を考慮する必要がなく、SRTAのようにすべての部分トラスが同一の構造であるか、ARTAのように異なる大きさの部分トラスの存在を許すかの差は生じない。

一方、基本並列データ転送手順は、トラスの結合状態のレベルというSRTA固有の特徴に依存して、多数の組のPE間で通信が同期実行されるアルゴリズムであるので、SRTA固有のものとなる。

### 6. おわりに

本論文では、 $n$ 次元トラスを再帰的に分割可能とすることによって、PE間の通信距離を短縮し、均一な像的データの並列処理に加え、構造を持ったデータに対しても効率的な並列処理が可能となることを目指した再帰トラス結合アーキテクチャを提案し、その構成法と通信距離特性および基本的な並列アルゴリズムを示した。これからの課題としては、次の3つが挙げられる。

(1) RTAは、MIMD型分散メモリ並列計算機の抽象アーキテクチャであり、今後は、PE、通信線、メモリ、スイッチ等に関する具体的なハードウェア仕様の設計を行う必要がある。特にスイッチのハードウェア構成とスイッチ群の制御方式は、RTAの有効性を実証するための重要な要因であり、現在検討中のハードウェア実現方式を簡単に述べる。

まず、RTAでは多数のスイッチを介してデータが転送される(たとえば1D-RTAの場合、アドレス0と $N-1$ のPE間の通信では、 $(N/2)-1$ 個のスイッチを経由してデータが転送される)ことがあり、スイッチでの遅延が問題となる。この問題に対しては、スイッチはゲートのみによって構成し、ラッチを持たせないようにすることでスイッチの遅延時間を最小限にとどめることを考えている。

スイッチ群の制御方式としては、各PEがそれぞれ接続しているスイッチを切り換える分散制御方式と、スイッチ群全体を制御するスイッチコントローラを設ける集中制御方式が考えられるが、RTAでは後者の方式を採用する。現在検討中のスイッチ制御方式では、各ビットが個々のスイッチのON/OFF状態を表すビットマップを設け、スイッチコントローラがビットマップにスイッチの切り換えパターンを書き込むこ

とによってスイッチ切り換えを実現する。この方式では、ビットマップから各スイッチに制御信号線を1本ずつ接続するだけでよい。図3や図8に示した木構造は、論理的なものであり、スイッチコントローラがそうした階層的なスイッチ切り換えパターンをビットマップ上に生成すればよいわけである。

(2) 3, 4章で述べた基本並列データ転送手順は、データを転送するたびにアクティブなPEの数が1/2になるという問題を含んでおり、SRTA, ARTAに対してさらに効率的な並列データ転送手順の開発を行う必要がある。

(3) RTAによる並列画像理解を実現するには、画像理解システムにおける画像処理、特徴抽出、マッチングの各解析過程において有効に働く並列アルゴリズムの開発が必要である。

### 参 考 文 献

- 1) 坂上勝彦, 木戸出正継: イメージプロセッサの最近の動向, 電子通信学会誌, Vol. 67, No. 1, pp. 90-98 (1984).
- 2) Li, H. et al.: Polymorphic-Torus: A New Architecture for Vision Computation, *Proc. of Workshop on Computer Architecture for Pattern Analysis and Machine Intelligence*, pp. 176-183 (1987).
- 3) Parasanna-Kumar, V.K. et al.: Parallel Architecture for Image Processing and Vision, *Proc. of Image Understanding Workshop*, pp. 609-619 (1988).
- 4) Weems, C.C.: Some Sample Algorithms for the Image Understanding Architecture, *Proc. of Image Understanding Workshop*, pp. 127-138 (1988).

- 5) 富田眞治: 並列計算機構成論, 昭晃堂 (1986).
- 6) 菅谷光啓ほか: 自己相似型ネットワークを有するマルチプロセッサシステム上での並列アルゴリズム, 電子情報通信学会論文誌, Vol. J74-D-I, No. 11, pp. 847-855 (1990).
- 7) 青山正人: 再帰トーラス結合アーキテクチャの研究, 岡山大学工学部情報工学科卒業論文 (1991).
- 8) Stout, Q. F.: Mesh-Connected Computers with Broadcasting, *IEEE Trans. Comput.*, Vol. C-32, No. 9, pp. 826-830 (1983).

(平成3年7月3日受付)

(平成3年12月9日採録)



松山 隆司 (正会員)

昭和51年京都大学大学院修士課程修了。京都大学工学部助手, 東北大学工学部助教授を経て, 平成元年より岡山大学工学部教授。京都大学工学博士。昭和57~59年米国メリランド大学客員研究員。画像理解, 人工知能, 並列処理の研究に従事。本学会創立20周年記念論文賞, 平成2年人工知能学会論文賞受賞。著書「SIGMA: A Knowledge Based Aerial Image Understanding System」など。



青山 正人

1969年生。1991年岡山大学工学部情報工学科卒業。同年岡山大学大学院工学研究科情報工学専攻修士課程入学。電子情報通信学会会員。