

並列推論マシン PIE 64 の相互結合網の作製および評価†

高橋 栄一†† 小池 汎平†† 田中 英彦††

本研究は、大規模な知識処理の高速実行を目的として研究を進めている並列処理マシン PIE 64 の相互結合網の開発に関するものである。一般に並列計算機において、相互結合網は計算機アーキテクチャの良否を決定する重要なファクタの一つであり、相互結合網の性能や特徴は、システム全体の処理能力と処理方式に重要な影響を与える。PIE 64 におけるプログラムの実行は、細粒度のプロセスを動的に生成し、かつ割り付けることにより行われ、この実行過程で発生するプロセッサ間通信を効率的に支援するような特性を有する相互結合網を構成する必要がある。本稿では、まず PIE 64 の相互結合網としてどのような構成のネットワークが最適かを考察し、(1)回線交換、(2)ノンバッファリング、(3)多段網、(4)動的負荷分散支援、(5)二重構成(同一構成の独立した二つのネットワークを用意)などの特徴を有するネットワークが PIE 64 の相互結合網として妥当であることを述べる。次に、相互結合網ハードウェアの実装方法を検討し、実際の実装過程について説明する。最後に、作製した相互結合網ハードウェアの予備評価として、経路設定や転送遅延など基本的な機能や信号伝送路の品質などの電気的特性の測定結果を検討し、PIE 64 の相互結合網として十分な性能を持つことを示す。

1. はじめに

PIE 64¹⁾ は、記号処理を高速実行することを目的として開発を行っている並列知識処理マシンである。PIE 64 が有する並列計算機アーキテクチャ上の大きな特徴は、要素プロセッサである 64 台の推論ユニット²⁾ が、2 系統の均質な相互結合網²⁾ で接続されている点である。

相互結合網^{12),13)} は、並列計算機アーキテクチャにおける基本的なチョイスポイントのひとつであり、その性能や特性は、システム全体の処理能力と処理方式に直接反映する。

我々は、PIE 64 の相互結合網として、

- 回線交換
- 多段網
- 動的負荷分散支援³⁾

などの特徴を有する相互結合網を 2 系統用意した。

相互結合網の実現には、構成単位となるスイッチングエレメントをゲートアレイで作成し、その LSI チップを用いて 64 ポート×64 ポートの多段網を構成した。多段網は構造的には規則的であるが、他のトポロジと比較して配線パスが多くかつ集中しており実装が容易ではない。この点に関し PIE 64 では、実装方法を工夫して解決した⁹⁾。

現在、PIE 64 のハードウェアの開発は、完成した相互結合網ハードウェアのテストが終了し、推論ユニットの作製が進んでいる段階である。相互結合網のテストは、専用ハードウェアデバッグ¹⁰⁾を用いることにより、効率的で実際の動作状態に近いレベルでの機能チェックとデバッグ、テスト系自身のテスト、電気的特性の測定、および、放熱状態の測定などを行うことができた。

本稿では、まず、PIE 64 の相互結合網の構成を検証する。次に、相互結合網ハードウェアの実装方式を検討し、それに従って実装した結果を報告する。最後に、完成した相互結合網ハードウェアの電気的特性の測定結果について考察する¹¹⁾。

2. 並列推論マシン PIE 64

PIE 64 は、Committed Choice 型言語 Fleng⁴⁾ およびオブジェクト指向言語 Fleng++⁵⁾により記述された大規模知識処理ソフトウェアの高速実行を目的とする並列記号処理マシンである。PIE 64 では並列処理技術の基本は相互結合網にあるものと考え、強力な相互結合網により多数台のプロセッサを結合するタイプのアーキテクチャを目指して研究開発を行ってきた。

PIE 64 のアーキテクチャ上の特徴は、推論ユニットと呼ぶ 64 台の要素プロセッサが、同一構成の 2 系統のネットワークからなる相互結合網で結合されている点である。PIE 64 の全体構成を図 1 に、相互結合網 (1 系統分) を図 2 に示す。

推論ユニットの構成を図 3 に示す。各機能ブロック

† Implementation and Evaluation of an Interconnection Network of a Parallel Inference Machine PIE 64 by EIICHI TAKAHASHI, HANPEI KOIKE and HIDEHIKO TANAKA (Department of Electrical Engineering, Faculty of Engineering, The University of Tokyo).

†† 東京大学工学部電気工学科

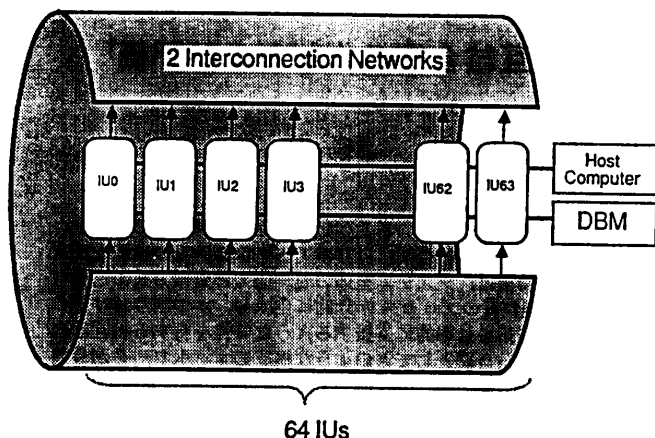


図 1 PIE 64 のアーキテクチャ

Fig. 1 Architecture of a parallel inference machine PIE 64.

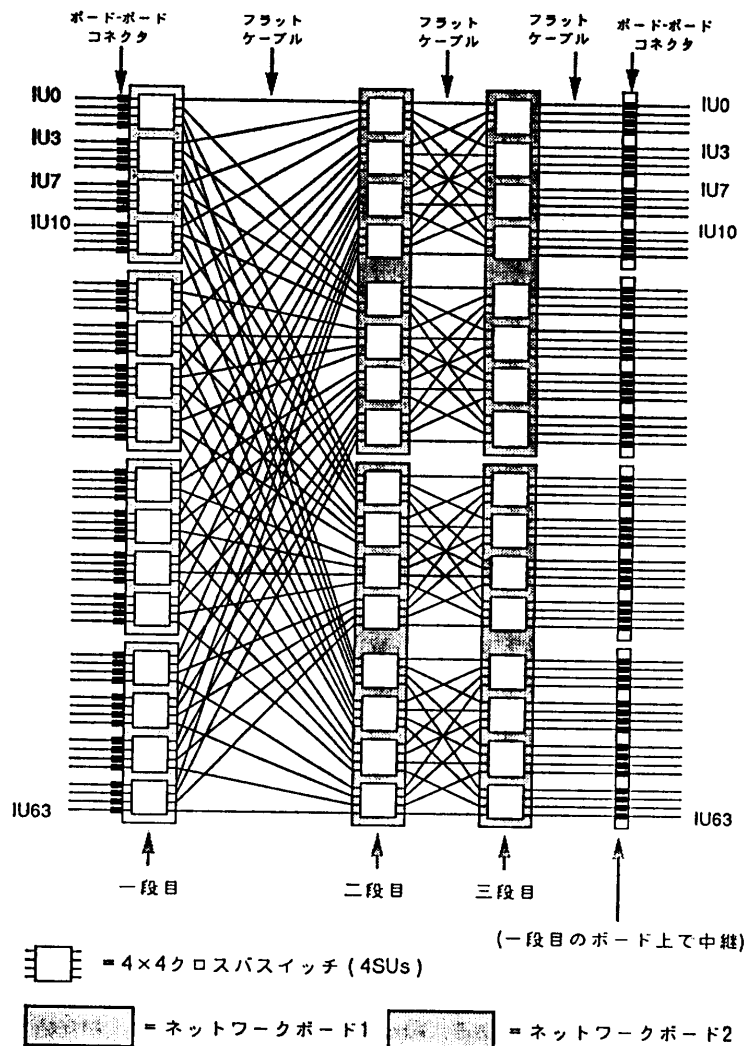


図 2 PIE 64 の相互結合網

Fig. 2 Interconnection network topology of PIE 64.

について簡単に説明する (カッコ内は図 3 中の名称)。

●ユニファイア/リデューサ
(UNIED)⁶⁾

Fleng の実行の中心であるユニフィケーション, リダクション処理を高速に実行する専用プロセッサである。

●ネットワークインタフェースプロセッサ (NIP)⁷⁾

相互結合網を介して 2 台の推論ユニット間で行われるリモートデータアクセス, プロセス間同期などの並列処理を支援する通信制御ハードウェアである。

●管理プロセッサ (SPARC)

負荷分散や実行のスケジューリングなどのゴール管理, システム述語の実行, 分散ガベージコレクションを担当する。

●ローカルメモリ (Local Memory)

バンク分けされ, 管理プロセッサおよび UNIED, NIP からアクセス可能なメモリである。

●ホストインタフェース (Host I/F)

ホストプロセッサであるワークステーションと PIE 64 とのインタフェースを行う。

●I/O インタフェース (I/O I/F)

データベースマシンやグラフィックプロセッサなどを接続することを想定したインタフェースである。

PIE 64 上における Fleng プログラムの実行モデルは, Committed Choice 型言語である Fleng の計算モデルにほぼ忠実であり, Fleng プログラムの実行に伴って動的に生成されるゴールがその処理単位となり, また要素プロセッサへの割付けもゴール単位で行われる。したがって, 相互結合網を介してやり取りされるデータは, ほとんどが変数 (1 ワード) やリストセル (2 ワード), あるいはアリティの小さい

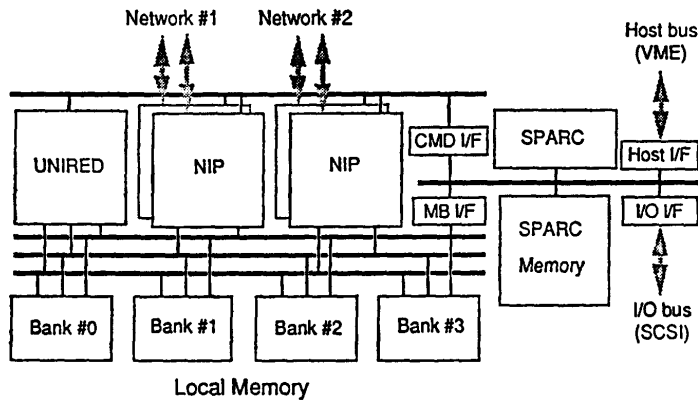


図 3 推論ユニット

Fig. 3 Block diagram of an inference unit, a processing element of PIE 64.

ベクタ (n ワード) であり、かつこれらの小さいデータの通信処理が多数行われることが予想される。また、送ったデータに対して戻り値を要求するような通信 (リモートデータの読み込みや書き込みアドレスを返させるような場合) では、処理の粒度が細かいことから、レイテンシが低いことも必要である。

3. PIE 64 の相互結合網

前章で述べたように、PIE 64 の相互結合網は高バンド幅であると同時に低レイテンシでなければならない。また、並列計算機ではプロセッサ台数が増えるに連れ、相互結合網のハードウェアの割合が増大し実装が困難となる¹⁴⁾傾向にあるので、実験機として実際に実装可能であるようにできるだけ単純な構成である方が望ましい。そこで、こうした要求を満足するようなネットワークとして、

- 回線交換
- 多段網
- ノンバッファリング
- 動的負荷分散支援
- 二重化されたネットワーク
- 分散経路制御

という特徴を持つネットワークを提案した。

まず、PIE 64 の相互結合網の大きな特徴である、

- 回線交換
- 多段網

について考察する。

回線交換 PIE 64 では、パケット交換方式ではなく、回線交換方式を採用したが、これは次のような考察に基づく。

- 相互結合網を流れるのは経路制御情報とデータだけであり、パケットの順序制御など通信のための付加的な情報によるオーバーヘッドはない。
- 機能レベルが原始的であり、相互結合網のレベルでデッドロックが生じることはない。
- 接続を保ったままの双方向通信が可能で、データをリモートに書き込み、その結果何らかの返事を持つようなアクセス、つまり、低レイテンシが要求されるような場合に有利である。

- 接続のオーバーヘッドは大きい、接続後の転送のコストは小さく、大きなデータ転送に有利である。

- 同程度のデバイステクノロジーを用いた場合、機構が単純なのでハードウェアの小型化、処理の高速化が期待できる。
- 開発という観点から見た場合にも、設計およびデバッグが容易である。
- 相互結合網を外側から見た場合、機能が原始的である分、通信制御ハードウェアである NIP による最適化が期待できる。

また、いくつかの問題点を指摘することができるが、それらは次のように対応可能である。

- 「閉塞網を用いる場合、閉塞状態に対する何らかの対策が必要である」
相互結合網では閉塞状態によるタイムアウトを扱わず、通信制御ハードウェアでタイムアウトを検出する。こうすることにより、相互結合網のハードウェアが単純になり、また通信制御ハードウェアではネットワークの外側から閉塞状態に対する効率的な処理が期待できる。
- 「閉塞網を用いた場合、相互結合網が混雑してくると急激に閉塞率が上がりスループットの低下を招く」
相互結合網を 2 系統使用することにより、「耐閉塞性」を高める。また、ネットワークトラフィックを低減するような静的、および動的負荷分散ストラテジを併用する。
- 「推論ユニット間を結ぶ通信経路の物理的な距離が長くなると、信号伝達路としての品質が低下し、誤り制御などのために実質的な転送速度が低

下する」

推論ユニットは各々の間の物理的な距離が最短になるように配置する。

多段網 同様に、PIE 64 のネットワークポロジとして採用したスイッチ結合型の多段網について、ハイパキューブやメッシュ、クロスなどの他のトポロジと比較した場合の特徴を挙げる。

- 「階層構造を持たない均質な相互結合網であり、すべてのプロセシングエレメント間の距離が等しい」

ある種の問題を扱うのに有利な固定的な接続には向いてないので、必ずしも最適ではないが、広範囲な問題を扱うのに適している。また、負荷分散を考える際の処理モデルが単純化され、動的な負荷分散処理のためのオーバーヘッドを抑えることができる。ハイパキューブやメッシュでは、動的な負荷分散処理はオーバーヘッドが大きくなり、必ずしも有効性を保証できない。

- 「プロセシングエレメント間の平均距離が小さい」
二つのプロセシングエレメント間の平均距離（転送経路を構成するアーク数、ただしスイッチノードも中継点と見なす）は、プロセシングエレメント数を n とすると、

$$k \text{ 次元メッシュ} \dots O(\sqrt[k]{n})$$

$$(k \leq 3)$$

$$\text{ハイパキューブ} \dots O(\log_2 n)$$

$$\text{多段網} \dots O(\log_k n)$$

$$(k=2, 4, 8, \dots)$$

$$\text{クロスバ} \dots O(1)$$

であり、多段網は比較的小さい（対数の底 k はスイッチノードのポート数）。特にハイパキューブと比較した場合、ノンバッファリング・回線交換の多段網ではデータの転送遅延は純粹に信号の伝達遅延のみになる。これによりデータの高速転送が可能になるが、実現のためのハードウェア量、特にプロセシングエレメントやスイッチ間を接続するパス数が増大する。クロスバスイッチでは、実現のためのハードウェア量がさらに増大する。

- 「シングルパスの閉塞網である」

多段網としては最小のハードウェア量で構成でき、冗長経路を持たずルーティングが単純になる利点がある。しかし、高負荷時に急激に閉塞率が上昇するので、この現象を抑えるような静的/動的

的負荷分散処理が必要になる。特に、ホットスポットの発生によって閉塞率が上昇している場合には負荷分散処理が有効である。

- 「ネットワーク構成（ノード数、ポート数）に対し拡張性がある」

メッシュ、ハイパキューブ、多段網などは同一のハードウェアを用いてプロセシングエレメント台数のバリエーションに対応することができる。特にメッシュであれば、持続できるプロセシングエレメント台数に制限はない。ハイパキューブや多段網では、プロセシングエレメントの台数を増やすためには、ノードに対して特別な機構が必要となる。

PIE 64 相互結合網のその他の特徴 次に、PIE 64 の相互結合網が持つその他の特徴について考察する。

- 「ノンバッファリング」

通信網としてみた場合、比較的小規模の回線交換網では途中でデータをバッファリングせずに伝達することができる。これにより低レイテンシのネットワークが実現できる。

- 「動的負荷分散支援（図 4）」

通信に使用していない転送パスとスイッチを用いてコンパレータを構成し、各推論ユニットの負荷を示すデータをこのコンパレータで比較する。最小の負荷量を報告した推論ユニットを相互結合網が記憶し、同時にこの最小値を通信を行っていない

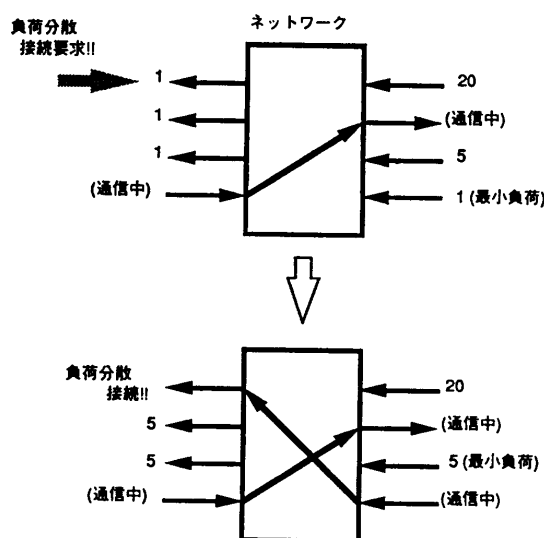


図 4 動的負荷分散支援機能

Fig. 4 The way how an interconnection network supports to balance loads dynamically.

い各推論ユニットに配る。負荷最小の推論ユニットへの接続は相互結合網が行う。この方式は単純な機能であるが、システム中での負荷の最小値の取得、および最小負荷推論ユニットへの接続を瞬時に実現し有効である。通信に使用していない資源を利用しているので、通信処理へのオーバーヘッドはない。負荷量としてどんな値を用いるかがこの機能を利用する上での鍵となる。

●拡張性

ビットスライス構成により、8ビットずつビット幅を拡張することができる。また、4段までの多段構成をサポートしており、最大256台の推論ユニットが接続できる。これにより推論ユニットの台数、および推論ユニット間の転送データ幅に自由度を持たせることができる。

●二重化された相互結合網

2系統の相互結合網を使用することにより閉塞率を抑えることができる。また、各推論ユニットの負荷情報を2次元で表現することができ、よりきめの細かい動的負荷分散支援が可能になる。ただし、どのように各要素プロセッサの負荷を算出するか、また、2系統のネットワークが各々持つ負荷分散支援機能をどう使い分けるかは今後の研究課題である。

4. 相互結合網の実装

前章では、PIE 64 が採用した相互結合網の構成に関して考察を行い、その妥当性を検討した。その結果、定性的な評価においては妥当性が確認されたものの、定量的な判断が必要な部分についてはある程度実機での評価が必要である。

本章では、PIE 64 の相互結合網の実現について、問題点およびその解決法を述べる。

以下、

1. 基本構成
2. 推論ユニットの物理配置
3. 相互結合網の実装
4. スイッチノードの配置
5. 組立

の順に説明する。

基本構成 まず、PIE 64 の基本構成として、

- 実用規模の並列マシンという目標に従

い64台の推論ユニットを作成する。これは、むやみに推論ユニット台数を増やすよりも、高性能の推論ユニットを用意してシステム全体としても実用的な性能を達成しようという主旨からである。

- 相互結合網は、閉塞率を抑えスループットを向上するために2系統用意する。

を決定した。

特に前章で述べた回線交換の多段網を実現するために、4×4でデータ幅8ビットのクロスバスイッチをゲートアレイ (Switching Unit, SU) を用いて開発した。このSUは、

- 多段構成をサポートする分散型のルータ
- 8ビットずつのビットスライス構成でデータ幅拡張可能
- 動的負荷分散機能を支援するコンパレータ内蔵
- 内部診断用のスキャンパス

などの機能を持つ。推論ユニット内の処理が32ビットで行われるので、相互結合網のデータ幅も32ビットとした。

推論ユニットの物理配置 PIE 64 の相互結合網は間接結合網¹³⁾であり、周りを推論ユニットが取り巻くような論理構成を有する。したがって、物理的にも図5のように相互結合網を中央に、推論ユニットを周囲に配置するのが合理的である。

相互結合網の実装 相互結合網の基板は、推論ユニットボードのバックプレーンにあたる位置と、これらに対し箱の上面や底面にあたる位置に配置する。2系統の相互結合網は、同形状に組み立てたものを上下に併置する。また、基板間の接続は基板間コネクタで行う

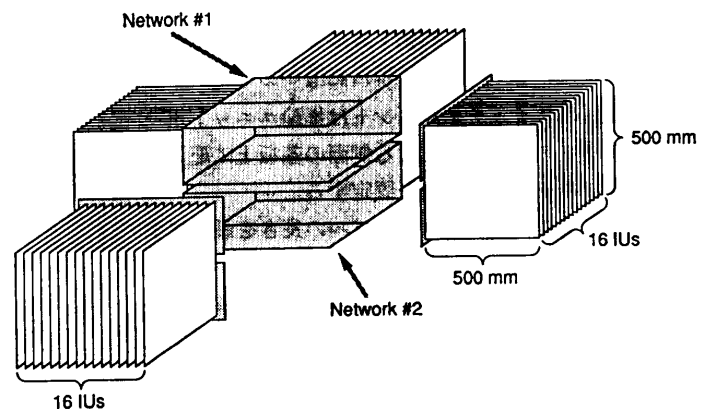


図5 PIE 64 の実装

Fig. 5 A detail of PIE 64 hardware.

のが確実であるが、

- 基板間で結線すべき信号線の本数が多く、通常の密度のコネクタが使用できない。
- 特殊な高密度のコネクタを使用すると、取り付け部分での基板上の配線が困難になる。
- 相互結合網のチェックは全体に対してのみ可能である。

また、基板上の配線に対し、

- 信号線への雑音の回り込みやクロストークが問題となりやすい。
- 32 ビット分の信号線の遅延特性を揃えるのは困難である。

などの問題点が存在する。

これに対し PIE 64 の相互結合網では、スイッチ間の接続をすべてフラットケーブルを用いて実現した。この方式は、

- フラットケーブルの伝達特性は安定している。
- スイッチ間の遅延を全接続に対して揃えることができる。
- 各スイッチノードが独立しているので、チェックしやすい。
- スイッチノードの配置を工夫することにより、最長のフラットケーブルを最短にすることができる。

などの利点を持つ。逆に、接続の正しさや接続部分に対する信頼性が低いという指摘があるが、専用のメンテナンス機構「タコ」(次段落参照)を用いた効率的なテスト環境によりこの問題を解決した。

スイッチノードの配置 PIE 64 の相互網は、スイッチノード(四つの SU で構成される 32 ビットクロスバスイッチ)を実装した 6 枚の基板を直方体の箱状に配置し、各段のスイッチノード間を接続するフラットケーブルをその箱の中に収納する形で実現する。基板上のスイッチノードの配置には自由度があるが、

- 1 段目のスイッチエレメントは推論ユニットボードのバックプレーンとなる側面の基板上に配置し推論ユニットとの接続を固定する。
- 必要な最長のフラットケーブルの長さを最短にする。
- ステージ間のシャッフルが四つに分割できることから、分割されたケーブル群が入り交じらないようにする。

という条件に従い決定する。

組立 組立(図 6)には困難が予想されたが、



図 6 相互結合網の組み立て

Fig. 6 The interconnection network (under building).

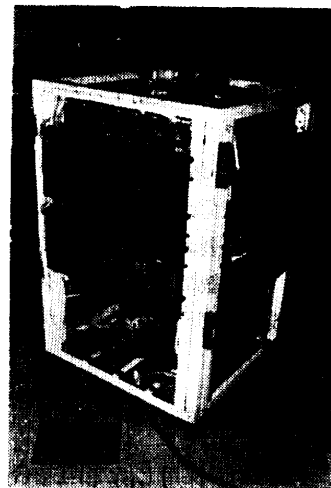


図 7 相互結合網(完成時)

Fig. 7 The interconnection network (finished building).

表 1 相互結合網の機能

Table 1 Specifications of the interconnection network.

—二系統
—回線交換方式
—ノンバッファリング
—多段網(3 段構成, スwitchノードは 4×4 のクロスバ)
—入力側 64 ポート×出力側 64 ポート
—転送方向を切り替えて双方向通信可能
—動的負荷分散支援
—データ線 1 ワード=32 ビット(他に通信制御線 6 ビット)
—動作クロック 10 MHz
—分散制御型のルーティング機能内蔵
—ビット幅, ポート数が拡張可能
—マルチキャスト接続(一対多接続)機能

- 彩色して各ケーブルの論理的な位置を明確にするとともに、誤配線を防ぐ。
- オープンなスペースであらかじめケーブルの配置を決定しておき、そのままラック内に移動して、コネクタの接続を行う。

などの対策を講じた結果、非常に効率よく作業を進めることができた。

完成した相互結合網の外観を図 7 (写真) に示す。また、本章で述べた相互結合網の機能をまとめて表 1 に示す。

5. 相互結合網のテストおよびメンテナンス

PIE 64 の相互結合網のテストやデバッグを行う方法について述べる。

相互結合網はスイッチエレメントごとにテストすることができるが、

- 基本的にスイッチ素子であるので監視すべき端子数が多い (1 ボード 38 本×8 ポート)。
- 32 ビット幅のデータ線は双方向端子である。
- 実働状態と同じ 10 MHz のクロックを使用してのテストが必要。

などの理由から、通常のロジックアナライザではテストが困難である。

そこで、PIE 64 のホストインタフェースおよびクロックジェネレータを組み込みスイッチエレメントのテストに必要な 38 ビット×8 本のプローブを持つ PIE 64 の相互結合網専用ハードウェアデバッグ「タコ」を開発した。

この「タコ」は、

- 推論ユニット内の NIP の機能を一部エミュレーションすることができ、これにより 10 MHz のクロックで相互結合網を駆動することが可能となる。
 - 相互結合網上を流れるデータをモニタする。
- という二つの機能を持ち、これにより

- SU チップのテスト
- スwitchエレメントのテスト
- フラットケーブルのテスト
- 相互結合網基板のテスト
- 多段接続状態でのテスト
- 相互結合網のモニタ

などに使用できる。

タコの操作はワークステーション上で対話的に行うことができるように X ウィンドウ上に相互結合網のテ

スト支援環境を作成した。

実際のテストは、

- ホストから対話的にタコを動作させ、基本的な機能を確認する。
- バッチ的に動作させ、システムティックに全状態のテストを行う。

という手順で、非常に効率よくテストを行うことができた。

特に、タコがシステムクロックを操作し、またゲートアレイ内のスキャンバスを自由に扱える機能により、ソフトウェアのデバッグのような気軽さで様々なハードウェアのテストを行うことができた。

6. 相互結合網の電気的特性の評価

本章では、作製した相互結合網ハードウェアに対する評価の第一段階として、本来の使用形態とは独立な基本パラメータを実際に動作しているハードウェアを測定することにより取り出し、検討を加える。

まず、次のような二つのステップで基本的な動作確認を行った。

1. 相互結合網を構成するすべての 32 ビット幅、4×4 クロスバスイッチ (SU チップ 4 個により構成) に対する動作確認
2. 実装状態の相互結合網に対して、フラットケーブルにより実現されているすべてのステージ間接続が正しいノード間を結合しており、かつ、正常にデータ転送を行えることの確認

4 章で述べたように、すべてのクロスバスイッチは基板上で独立しているため、基板単体に対する 1. のテストは迅速に、そしてシステムティックに行うことができた。2. のステージ間接続に関するテストは、4096 通り (64 入力×64 出力) ある接続経路のうち、10% 以下の 384 通りの接続経路について接続テストを行うことにより達成できることが、ネットワークポロジのグラフを検討することにより分かる。この 1. と 2. のテストにより、全接続経路のテストを行うことなく、しかもそれよりも少ない手間で、相互結合網全体の機能確認を行うことができた。

次に、相互結合網の基本性能の測定を行った。以下、項目別にテストの内容と結果を示す。

• 経路設定 (接続) 多段網の場合には、システムクロックに同期して動作させても、経路接続にはスイッチノードごとに 1 クロックずつかかり、PIE 64 の相互結合網の場合 3 段の多段接続を行っているため、接

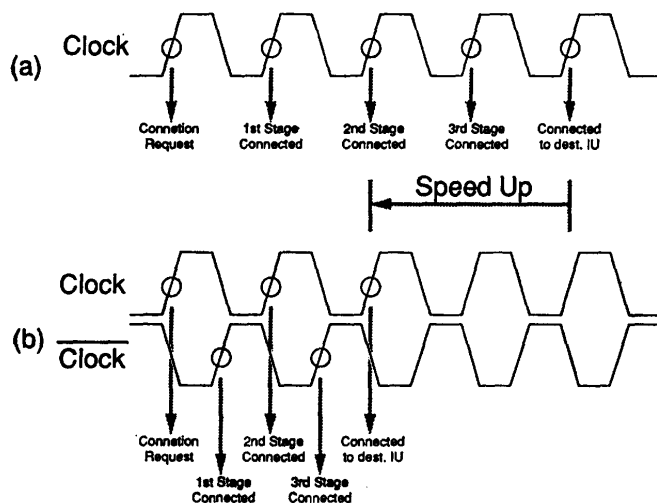


図 8 ステージ間のクロックの位相差を利用した高速化

Fig. 8 Speedup of connecting cycles using two phase clock.

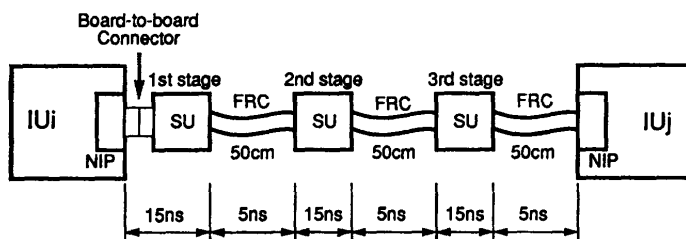


図 9 接続時のデータ転送遅延

Fig. 9 Delay time in transferring data in the interconnection network.

統先の要素プロセッサが接続要求を受け取るのは4クロック目である(図8(a)).

このような接続のためのオーバーヘッドは、一般に回線交換のデメリットの一つとしてあげられるが、PIE 64では1段目と3段目にシステムクロックに対して180度の位相差を持つクロックを与えることにより、接続にかかる時間が2クロック分短縮されることが動作テストの結果確認された(図8(b)).

●**データの転送** 回線交換方式であるため、通信経路が設定された要素プロセッサ間は直接データ転送が可能となる。したがって、相互結合網内でのデータや制御信号の伝達遅延は、純粋に電気的な信号の伝搬遅延のみとなる。測定の結果、図9のようにゲートアレイ内の遅延が約15ns、フラットケーブルの遅延が約5ns、全体で約60nsであった。この値はシステムクロックサイクル100nsに対して、要素プロセッサ同士が十分同期転送を行うことのできる数値である。

●**転送方向の反転** データの転送方向は、それを示す

制御線の電圧レベルにより指定されるが、この制御線もデータ線と同様回線交換方式により接続されているため、状態の変化(転送方向の反転)は約60nsの遅延時間で伝達される。各ステージのスイッチノードはこの変化に応じて非同期的に(システムクロックに関係なく)データ転送方向を変化させるので、結果として1クロック内で転送方向の反転が完了することがテストの結果確認された。

●**接続解除** 接続解除要求を示す制御線も、データ線と同様回線交換方式で接続されるので、解除要求はやはり約60nsの伝搬遅延時間の後、接続先の要素プロセッサに伝えられる。一方、各スイッチノードは接続解除要求をクロックの立ち上がりで受理し、1クロックサイクル分接続状態を保持した後接続を解除する。これは各ステージのスイッチノードが位相の異なるクロックで動作している場合でも、確実に接続を解除するためのシーケンスであるが、テストの結果その有効性が確認された。

●**信号伝送路としての特性** スイッチノードを構成するゲートアレイは

CMOSデバイスであり、CMOSのドライバがフラットケーブルを介してCMOSのゲートをドライブする構成になっている。10MHzの矩形波をデータ信号として転送した場合の信号波形は、図10に示すように比較的良好的なものであった(上の波形が相互結合網の入

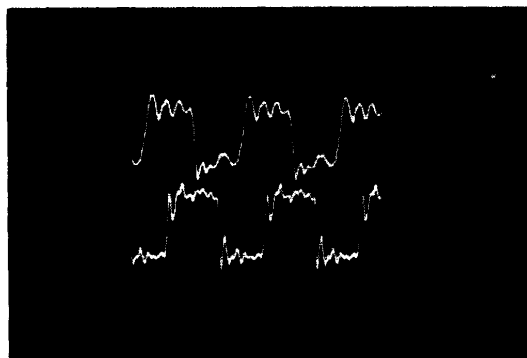


図 10 データ線の信号波形 (2V/div, 50ns/div)

Fig. 10 Wave form of a data signal in the interconnection network (2V/div, 50ns/div).

力、下が出力である。目盛は 2 V/div, 50 ns/div).
これは、

- 実装方式の工夫により、ステージ間を接続するフラットケーブル長を 50 cm と短くすることができたので、比較的遅い信号の立ち上がりで反射波が打ち消される。
- 信号転送に使用しているフラットケーブルには 1 本おきにグラウンド線をはさんであること、および転送がポイントーポイントで行われることにより転送路のインピーダンスが安定している。

などの理由によるものと考えられる。

また、転送時のビット誤りなどの発生率を測定する目的で、約 24 時間連続で接続・データ転送・接続解除を繰り返すテストを行ったが、ビット誤りはまったく観測されなかった。

さらに、実働状態で現れると思われる苛酷な状況での動作を確認するために、SU のマルチキャスト（一対多接続）機能を用いてネットワーク内の大部分のリンク（フラットケーブル 192 本中 149 本、約 80%）を一斉に振らせた状態、つまりこれらのリンクのすべてのビットに対して同時に“010101...”と変化するデータを与え、このような状況下で新たな回線の接続、転送、解除を行うテストを繰り返し、正常動作を行うことを確認した。

以上の評価によって得られた相互結合網の性能測定結果を表 2 にまとめる。このように、十分な性能を持った相互結合網を実現できたことが分かった。なお、今回の評価は相互結合網の基本パラメータについて行ったが、実働状況下での性能を知ることも当然重要である。これは、現在開発が進んでいる PIE 64 の要素プロセッサの完成を待って行い、別途報告することとしたい。

表 2 相互結合網の性能
Table 2 Performances of the interconnection network.

項 目	性 能
データ転送速度	40M バイト/s
バンド幅	相互結合網全体（2 系統）で 5G バイト/s
経路設定†	最短 2 クロック
データ転送遅延	約 60 ns
転送方向の反転	約 60 ns
接続解除	1 クロック

† 接続先指定時、負荷分散支援時ともに同値

7. おわりに

本稿では、

まず、PIE 64 の相互結合網の構成とその特徴である

- 回線交換
- 多段網
- ノンバッファリング
- 2 系統
- 動的負荷分散支援

について考察し、PIE 64 の相互結合網として適切であることを確認した。

次に、PIE 64 の相互結合網の実装方式である

- 相互結合網を中央に配置する。
- 多段網のステージ間接続にフラットケーブルを用いる。

について検討し、その実現の過程を報告した。

最後に、相互結合網のテスト、デバッグ、メンテナンス方式について述べ、テスト用ハードウェア「タコ」を用いて測定した電気的特性について評価し、基本的な特性については PIE 64 の相互結合網として適していることを示した。

今後の課題として次のような項目が挙げられる。

- PIE 64 上で実用的な規模のプログラムを実行したときのリモートアクセス特性を考慮したデータ転送能力の定量的な評価が必要である。これは、現在開発中の要素プロセッサに対する評価と合わせて行っていく予定である。
- プログラムに対する静的な解析方法や、動的な負荷分散戦略などを関連付けて、負荷情報をどう表現するか、また 2 系統のネットワークが独立に持っている動的負荷分散機能をどう有効利用するかについて考察する。

謝辞 相互結合網の構成要素となるゲートアレイ、SU の開発に当たり、多大なる御支援を賜った富士通研究所人工知能第三研究室の服部室長、並びに、久門氏、三宅氏に深謝いたします。

また、ハードウェアの作成やチェックに協力して頂きました研究生の松本氏（日本 IBM）と宮本氏（日立）、PIE 64 の相互結合網作成に当たり、必ず議論に参加し多くの貴重なアイデアを提供して下さった修士 2 年の日高君ほか、修士 1 年の中田君と毛利君、学部生の下國君、その他田中研究室の皆さんに感謝します。

さらに、相互結合網基板を作成してくださいました日立化成工業株式会社の山林氏、PIE 64 のラックの設計および作成、「タコ」の基板の作成を行ってくださいましたヨシキ電子株式会社の中野氏、橋本氏に感謝いたします。

なお本研究は、文部省特別推進研究 No. 62065002 の一環として行われている。

参 考 文 献

- 1) 小池, 田中: 並列推論エンジン PIE 64, 並列コンピュータアーキテクチャ, bit 臨時増刊, Vol. 21, No. 4, pp. 488-497 (1989).
- 2) Koike, H., Takahashi, E., Yamauchi, T. and Tanaka, H.: The High Performance Interconnection Network of Parallel Inference Machine PIE 64, *Computer Architecture Symposium IPS Japan*, pp. 65-72 (1988).
- 3) 坂井, 小池, 田中, 元岡: 動的負荷分散を行う相互結合網の構成, 情報処理学会論文誌, Vol. 27, No. 5, pp. 518-524 (1986).
- 4) Nilsson, M. and Tanaka, H.: Fleng Prolog—The Language Which Turns Supercomputers into Prolog Machines, in Wada, E. (ed.), *Proc. Japanese Logic Programming Conference, ICOT, Tokyo*, pp. 209-216 (June 1986). Also in Wada, E. (ed.), *Logic Programming '86, Springer LNCS 264*, pp. 170-179 (1986).
- 5) 中村, 小中, 田中: 並列論理型言語 FLENG に基づいたオブジェクト指向型言語 FLENG++, 日本ソフトウェア科学会オブジェクト指向計算に関するワークショップ WOOC '89 (1989).
- 6) 島田, 下山, 清水, 小池, 田中: 推論プロセッサ UNIREDI の命令セット, 情報処理学会計算機アーキテクチャ研究会, 79-5 (Nov. 1989).
- 7) 清水, 小池, 田中: 並列推論マシン PIE 64 の推論ユニット間通信, 情報処理学会計算機アーキテクチャ研究会, 79-4 (Nov. 1989).
- 8) 日高, 小池, 田中: 並列推論エンジン PIE 64 の推論ユニットのアーキテクチャ, 電子情報通信学会コンピュータシステム (CPSY) 研究会, SWoPP 琉球 '90, CPSY 90-44, pp. 37-42 (July 1990).
- 9) 高橋, 小池, 田中: 並列推論マシン PIE 64 の相互結合網の作成および評価, 並列処理シンポジウム JSPP '90, pp. 89-96 (May 1990).
- 10) 日高, 高橋, 小池, 清水, 田中: PIE 64 のネットワークメンテナンス, ホストインタフェース, クロック分配機構: タコ, 第 40 回情報処理学会全国大会論文集, 1L-7, pp. 1171-1172 (Mar. 1990).
- 11) 高橋, 小池, 田中: PIE 64 の相互結合網の電気的特性の評価, 第 41 回情報処理学会全国大会論

- 文集, 6P-3, 分冊 6, pp. 163-164 (Sep. 1990).
- 12) Bhunyan, L. N.: Interconnection Network for Parallel and Distributed Processing, *Computer*, Vol. 20, No. 6, pp. 9-12 (June 1987).
 - 13) 富田: 並列計算機構成論, 昭晃堂 (1986).
 - 14) Denneau, M. M., Hochschild, P. H. and Shichman, G.: The Switching Network of the TF-1 Parallel Supercomputer, *Supercomputing Magazine*, pp. 7-10 (Mar. 1988).

(平成 2 年 8 月 20 日受付)

(平成 3 年 2 月 12 日採録)



高橋 栄一 (正会員)

1963 年生。1987 年東京大学工学部電子工学科卒業。1989 年同大学大学院工学系研究科情報工学専攻修士課程修了。現在、同博士課程在学中。並列計算機アーキテクチャの研究に従事。IEEE 会員。



小池 汎平 (正会員)

昭和 36 年生。昭和 59 年東京大学工学部電子工学科卒業。平成元年同大学院工学系研究科情報工学専攻博士課程満期退学。同年東京大学工学部電気工学科助手。工学博士。平成 3 年東京大学工学部電気工学科講師。現在にいたる。並列計算機アーキテクチャ、および並列プログラミング言語に関する研究に従事。本会学術奨励賞受賞。日本ソフトウェア科学会、ACM 各会員。



田中 英彦 (正会員)

昭和 18 年生。昭和 40 年東京大学工学部電子工学科卒業。昭和 45 年同大学院博士課程修了。工学博士。同年東京大学工学部講師。昭和 46 年助教授。昭和 62 年教授。昭和 53 年～54 年ニューヨーク市立大学客員教授。現在に至る。計算機アーキテクチャ、並列推論マシン、知識ベース、オブジェクト指向プログラミング、分散処理、CAD、自然言語処理、等の研究を行っている。'計算機アーキテクチャ'、'VLSI コンピュータ I, II'、'ソフトウェア指向アーキテクチャ' (いずれも共著)、'情報通信システム' 著、電子情報通信学会、人工知能学会、日本ソフトウェア科学会、IEEE、ACM 各会員。