

ドーナツ構造 Loop Structured Computer(DLSC) の プロトタイプの動作特性

石田 努 吉岡 良雄

弘前大学理学部情報科学科

概要

コンピュータ処理の高速化を目指して、各所で大規模なマルチプロセッサの研究が行われている。本報告ではマルチプロセッサ・システムの一つの方策として提案したドーナツ構造 Loop Structured Computer(DLSC) のプロトタイプの各部の動作時間、待ち時間、利用率等の測定結果について述べ、それを検討し、次に大規模マルチプロセッサ（ドーナツ型 LSC）の構成方法について述べる。

Characteristics of Doughnut Type Loop Structured Computer Prototype

Tsutomu Ishida Yoshio Yoshioka

Dept. of Information Science, Faculty of Science,
Hirosaki University

abstract

Recently, a massively multi-processors system has been studied in order to realize higher performance of computers. This paper presents the computing time, the waiting time and the utilization of shift-register for our a doughnut type multi-processors system which consists of several Loop Structured Computers. Moreover, we discuss the suitable system of a doughnut type multi-processor system.

1 まえがき

1940 年代にノイマン型計算機が提言されてから、電子回路素子の高速化、集積化が進み、計算機の開発は急速に発展した。この計算機は、1つのプログラム・カウンタで記憶装置から1命令づつを中央処理装置に取り込んで実行するために、処理のほとんどがデータ転送に費やされている。また、今日、計算機が扱うべきデータが大量になっていることから、ノイマン型計算機に代わる計算機の開発、研究が行なわれている。パイプライン計算機や、並列処理計算機等も現在開発され、実用化されている。我々はループ状に接続されたシフトレジスタに処理要素が接続した並列計算機(Loop Structured Computer(LSC) [2] [3] [4])をドーナツ構造に拡張した、ドーナツ構造 Loop Structured Computer [5] [6] を提案してきた。この LSC は、バスによる並列転送に疑問をもち、ある転送率以上では直列転送の方が有利であるという理論的検討 [1] の結果提案されたものであり、ループ状に接続したシフトレジスタ間を同軸ケーブルや光ファイバで接続し、素子の動作限界で直列転送を行なうものである。また、その LSC をドーナツ状に結合したドーナツ構造 Loop Structured Computer は、並列処理モデルがプロセッサ数増加に伴い、各プロセッサ間の通信の複雑さが増加することを防ぐために、パケットの転送方向を限定する事によって PE(Processing Element) の負荷を軽減し、より大規模な並列処理を行なうシステムである。

本論文では、ドーナツ構造 Loop Structured Computer のプロトタイプ (8 層; 1 層 9PE) の各部の通信時間、処理時間、待ち時間、利用率等について報告し、大規模マルチプロセッサ (ドーナツ構造 LSC) の構成方法について述べる。

2 プロトタイプのシステム構成

本報告では、図 1 に示す Loop Structured Computer(LSC) を図 2 のように接続したドーナツ構造 Loop Structured Computer(DLSC) を取り扱う。プロトタイプは、現在、8 層の LSC から成り立ち、1つの LSC には 8 個の処理 PE と、制御部と通信するための通信用 PE から成り立っている。システムの全体像としては、図 3 に示すように、ホスト計算機・制御部、DLSC から成り立ち、各 LSC の通信用 PE が制御部と接続され、また、制御部がホスト計算機と接続されている。全ての処理 PE は、ホスト計算機から送られてくる演算マッピングパケットによって処理が決定され、到着したデータパケットを静的データフローによって処理する。プロトタイプにおけるシフトレジスタ部によるデータ転送は、3.3msec 毎の NMI(Non-Maskable-Interrupt) 割り込みを利用して同期をとり、並列転送アダプタ PIA(Peripheral Interface adapter) を介して、ソフトウェアで仮想的に実現している。

次に、処理 PE は図 4 のような構成になっている。PE 内には、上層からの出力パケットをシフトレジスタ部にのせるための Queue(上層からの出力 Queue)、ループ方向にパケットを転送するシフトレジスタ部、この PE を目的とするパケットをシフトレジスタ部から取り出し格納する入力 Queue、入力したパケットを処理する演算部、その結果パケットや下層の LSC へのパケットを格納する出力バッファ等がある。

処理 PE に演算を割り当てるためには、ホスト計算機から送られてくる演算マッピングパケットの情報を、PE 内にある機能決定情報に書き込むことにより実現できる。演算部ではその情報を

元に、入力されてくるデータパケットを静的データフローによって処理し、結果パケットを目的 PE に転送する。ホスト計算機から、DLSC に転送されるパケットには、PE の機能を決定する演算マッピングパケット、PE の機能決定情報を解消するマッピング消去パケット、演算データを格納する変数データパケット、定数データパケット、演算結果を格納する結果データパケットなどがあり図 5(a)(b) で示すような 16byte のデータ構造である。

さらに、プロトタイプでは、ホスト計算機と制御部の一部を X68000、ホスト計算機と LSC の制御部との通信方法として RS-232C、または並列データ転送 IO ポートを利用している。また PE の構成は、プロセッサとして MC68B09、シフトレジスタ用として MC68B21、メモリとして 8KBRAM(HM6264)、32KBROM(HN27C256G)、その他 HD74LS14、HD74LS00 等となっている。

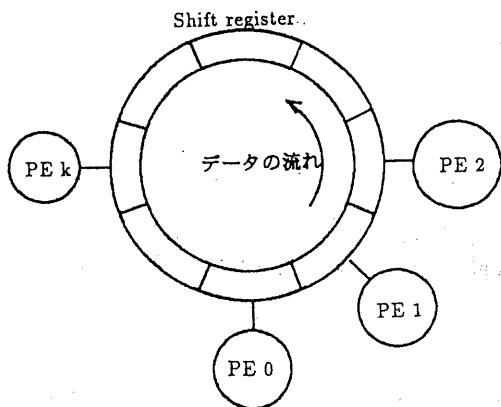


図 1 Loop Structured Computer

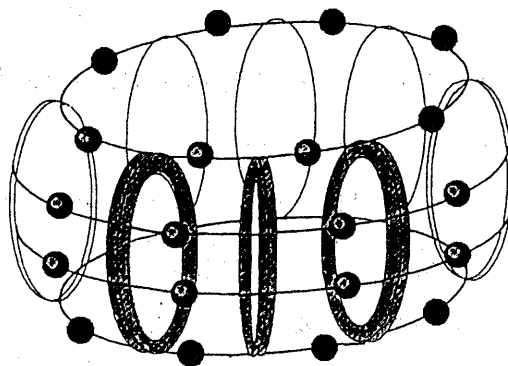


図 2 ドーナツ構造 Loop Structured Computer

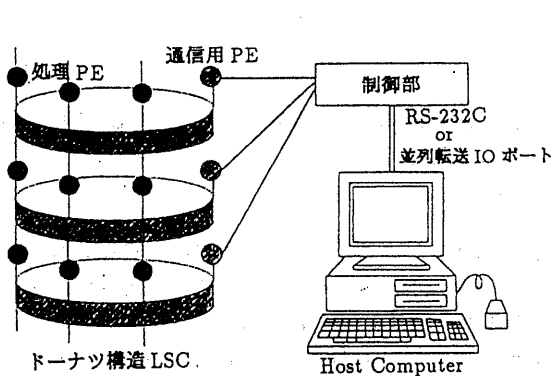


図 3 システム構成

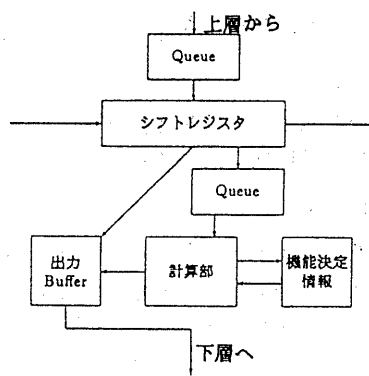


図 4 処理 PE の構成

送に使われることになり、残りはPEの演算処理等にあてられる。これゆえ、1パッケージがLSCを1周するのに必要な時間は $9 \times 3.3 = 29.7\text{msec}$ (1LSCのなかに9PE)となる。

5. 処理 PE

処理 PE では、約 3.3msec の NMI 割り込みによってシフトレジスタ部でパッケージ転送が行なわれるが、約 3.3msec のうち約 60% が演算にとられる。整数型演算の処理時間については加算、減算では約 8msec、乗算では約 20msec、除算では約 23msec 必要としている。また、下層へのパッケージ転送は割り込みを使用していないので平均 40msec 必要としている。なお、この時間はシフトレジスタ部でのパッケージ転送時間を含んでいる。

以上から、1パッケージがDLSCで何らかの処理(加算)をされ結果パッケージとして制御部に転送されるまでは、全体で約 130msec が必要になり、ほとんどがパッケージ転送に費やされていることがわかる。ここで注意しなければならないことは、これらの時間測定は全てスカラー演算の測定である。よって、これらのスカラー演算を単純に繰り返すようなプログラムでは通信のオーバーヘッドが増大し効率的な処理が期待できない。すなわち、ベクトル演算のようにパイプライン的に処理するほうが効果的である。

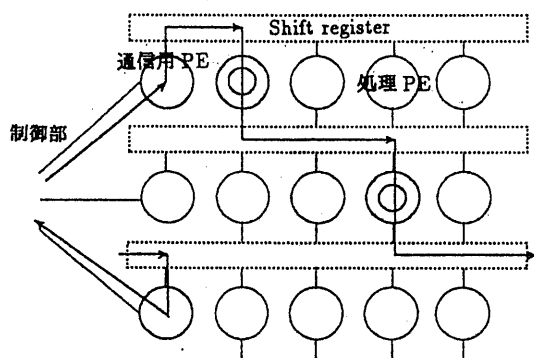


図6 DLSC 内のパッケージの流れ

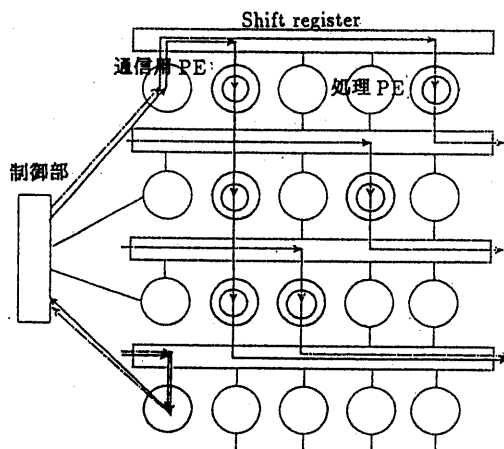


図7 最適割付と最悪割付

3.2 ハードウェア特性の検討

この節では、シフトレジスタの利用率、層方向の転送速度について検討する。

1. シフトレジスタの利用率

まず、1つのLSCにおいて考えてみる。LSC内では1パッケージが3.3msecでシフトレジスタ間を移動するのに対して、ホスト計算機から送られる1パッケージ転送時間が3.3msec以上かかるならば、シフトレジスタ部には空の状態が起こる。すなわち、RS-232Cで1パッケージの転送時間は約20msec必要としているので、パッケージをすきまなく順次転送しても1LSC上に2~3パッケージしかのっていないことになりシフトレジスタの利用率は $3.3 \div 20 = 0.17$ にまでしか至っていないことがわかる。

パケットの転送時間を約 3.3msec よりも速くすれば、シフトレジスタ部にパケットを乗せる通信用 PE に待ちができることになる。そして並列転送 IO ポートでの転送時間は 1.5msec(実質 250Kbps) であり、従って通信用 PE に待ちが起こる。以上の事を考慮すると、プロトタイプは 8 層の LSC からなるので、シフトレジスタの利用率は前の値の $\frac{1}{8}$ である。従って、シフトレジスタの利用率は RS-232C で 0.02、並列転送 IO ポートでは 0.28 となる。

2. 層方向の転送速度

シフトレジスタ間の転送は 3.3msec 毎の割り込みで実現できるのに対して、層方向の転送速度は約 40msec 必要となっている。これは、処理 PE でシフトレジスタ部のパケット転送が割り込みで行なわれるのに対して、層方向のパケット転送は割り込みを利用しないで、パケット解析、演算等を平行動作させているためである。従って、層方向のパケット転送時間を速くするためには、割り込みを利用したり、平行動作の優先度を高くする工夫が必要であろう。

つぎに、シフトレジスタ部による転送速度と層方向の転送速度について考えてみると、シフトレジスタによる転送速度が速い事で得られる利点は、プログラムの PE 割付が非常に楽になることである。つまり、図 7 のように最適なマッピングを行なわなくても遅延時間は高々 29.7msec(LSC をパケットが 1 周する時間) となる。逆に、シフトレジスタによる転送速度が層方向のそれより低かった場合は、LSC をパケットが 1 周する時間が多くなり、その結果、プログラムの PE 割付が制約され、数箇所割付ミスが性能を大幅に低下させることにつながる。このプロトタイプでは、最大遅延 29.7msec、層方向の転送時間 40msec となっている。

3.3 マッピングによる動作時間の検討

前の節でも述べたように、シフトレジスタによるパケットの転送速度と演算処理の時間の差は、プログラムの PE 割付に直接影響してくる。そこで、ここでは各 PE の通信の同期を取るための割り込み (NMI) を 3.3msec から 6.6msec に変えて、具体的にプログラムをマッピングし解析してみた。プログラムは簡単に $x = a + b + \dots$ として、スカラー演算とベクトル演算を計算する。マッピング方法は最適な PE 割付と最悪な PE 割付、そしてランダムに PE に割付を試みた。また、ホスト計算機からの転送は並列転送 IO ポートを利用する。

その結果を図 8(a)(b) に示す。NMI 割り込みを 6.6msec とした方は、最適な PE 割付においてはこれら全ての測定の内滞留時間が最小であったが、最悪な PE 割付とランダムな PE 割付は元の 3.3msec の NMI で同期を取った方よりはるかに滞留時間が長い。これは、処理 PE で演算中に、シフトレジスタ部のパケット転送のための割り込みが少なかった事が、最適な PE 割付において処理時間が短縮した原因と思われる。しかし、PE 割付が最適に行なわれなかった場合は、パケットがループ状に接続されたシフトレジスタを一方方向にしか流れないので、ループ方向の転送に費やす時間がより多くなり、処理全体としても遅くなっている。スカラー演算について、3.3msec の NMI で同期を取った方は滞留時間の最小と最悪の比が 1:1.25 に対して、6.6msec の NMI で同期を取って層方向の出力の優先順位を上げたほうは 1:2.1 となり、さらにベクトル演算も 1:1.17 に対して 1:1.81 となっている。この事から、シフトレジスタ転送と層方向の転送速度の比に対して、最大性能を引き出すためにマッピングアルゴリズムが重要な要素となる。

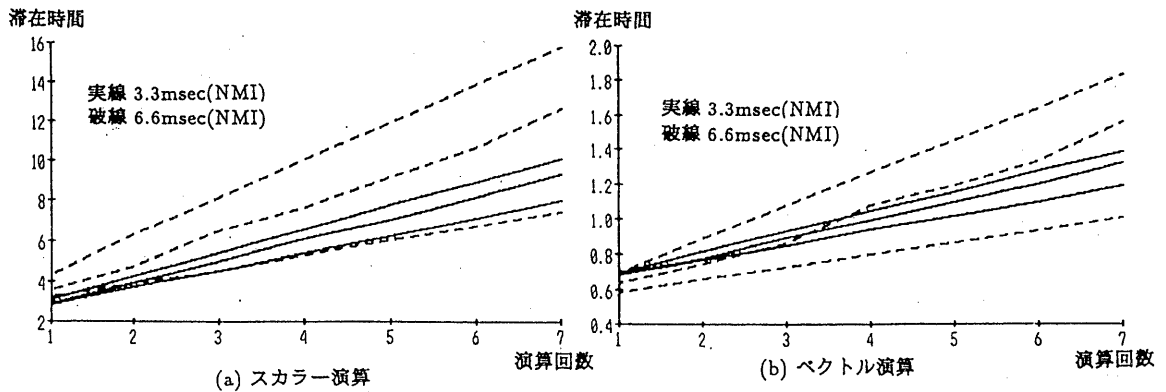


図8 演算回数による滞在時間の変化

3.4 ドーナツ構造 Loop Structured Computer の構成の検討

プロトタイプ動作の特性から、LSCの構成方法を考えてみる。まず、各演算PEの機能レベルとしては中～高を考えている。これは、機能レベルをより低くし、さらなる並列化を計っても通信のオーバーヘッドが増大し、効率的ではないと思われるからである。

またPE内の滞在時間における層方向のバケット転送時間の占める割合が大きくなっている。そこで、前節でも述べたように、処理PEの層方向の転送を割り込みを利用したり、プログラムの改良をすすめる必要がある。

次に考えることは、PE間の通信時間を下げることである。現在のプロトタイプでは1層のLSCに8個の演算PEと1個の通信用PEが接続されているが、将来的に1LSC内に128個の演算PEを接続するつもりである。このために現在の3.3msec毎の同期でシフトレジスタ部による転送では通信のオーバーヘッドが大き過ぎ、1LSCの滞在時間ばかりが大きくなり有効的な処理が期待できない。そのためにはシフトレジスタ部のバケット転送速度をもっと上げなければならない。このシフトレジスタ部によるバケット転送速度は、数百～数千Mbpsとなることが予想され、そのためにはシフトレジスタ部ハードウェア化し、回路の集積化(1チップ化)して、同軸ケーブルや光ファイバ等で高速通信することが必要になってくる。

また、ホスト計算機から制御部へのデータ転送方法も考慮しなければならない。現行のようにホスト計算機から一つずつバケットを転送していたのでは、現在のプロトタイプのPE数でシフトレジスタの利用率が0.28を考えると、将来的にPE総数32768個(256層のLSC、1LSC内に128PE)を考えた場合、シフトレジスタの利用率を上げるためには数Gbpsでホスト計算機と制御部が通信されなければならない。この転送速度を実現するのは、ホスト計算機の負荷も高くなることから困難であり、それゆえ制御部を完全にホスト計算機から切り外し、ホスト計算機からはバケットを送るのではなく、演算命令や演算データの情報等を制御部に送ることにし、制御部でバケットの生成やPEの状態管理等を行なうようにしたい。

4 まとめ

本報告では、Loop Structured Computer(LSC) をドーナツ構造に拡張した大規模マルチプロセッサのプロトタイプ通信時間、処理時間、待ち時間、利用率について報告し、それらの検討から大規模マルチプロセッサの有効的な構成方法はどのようにすればよいかを述べた。その結果、現プロトタイプでは、演算は滞在時間の約10%程度しか占めず、大部分がパケット転送や待ちに費やされていることが分かった。そこでまず、演算PEの負担を軽くし、層方向の出力を高速にすることが有効的であるという結論を得た。

今後は、まず層方向のプログラムの改良を行なう。さらに、プロトタイプへのマッピングアルゴリズムや高級言語からのコンパイラなどの開発を行なっていきたい。

参考文献

- [1] 吉岡 良雄、' 計算機におけるバスのデータ転送方式について'、情報理論とその応用研究会(第7回シンポジウム) 資料、1984年11月
- [2] 熊沢、吉岡 'Loop Structured Computer について'、情報処理学会・計算機アーキテクチャ研究会資料、56-1、1985年1月
- [3] 吉岡 良雄 'Loop Structured Computer の特性解析'、1989年 並列処理シンポジウム JSPP 論文集、1989年2月
- [4] 吉岡 良雄 'Loop Structured Computer のトラヒック解析'、電子情報通信学会論文誌 D-I、Vol.J72-D-I、3、pp.149-156、1989年3月
- [5] 吉岡 良雄 '超並列処理機構(ドーナツ型 Loop Structured Computer) の提案'、平成3年度第1回情報処理学会東北支部研究会、3-1-1、1991年5月
- [6] 石田 吉岡 'ドーナツ構造 Loop Structured Computer(LSC) の提案とそのプロトタイプについて'、情報処理学会・計算機アーキテクチャ研究会資料、1992年1月