

並列処理システム-晴-の ソフトウェアシミュレータによる評価

山名早人 萩原 孝 草野義博 村岡洋一

早稲田大学 理工学部

-晴-の機能レベルのソフトウェアシミュレータを用いて、データフロー実行における待ち合わせ記憶のバンク化と-晴-の大域的メモリである共有メモリについて性能評価を行う。

データフロー実行では、待ち合わせ記憶が性能上の隘路となり所期の性能を達成することが困難となる。この問題に対して、待ち合わせ記憶のバンク化を行い処理性能の向上を図る。また-晴-では共有メモリに、マルチリードメモリを採用している。このため、要素プロセッサ数の増加に伴い莫大なハードウェア量が必要となる。そこで、このハードウェア量削減の可能性をさぐるために、要素プロセッサ数とポート数及びバンク数の関係についてシミュレータにより検討を行う。

Evaluation of a Simulated Parallel Processing System -Harray- (in Japanese)

Hayato YAMANA, Takashi HAGIWARA,
Yoshihiro KUSANO and Yoichi MURAOKA

School of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku, Tokyo, 169 JAPAN

The purpose of this paper is to evaluate a banked matching memory unit and a global memory unit of the -Harray- system by software simulation.

A matching memory unit is a bottleneck in the data flow execution. The bottleneck resulting from complicated data handling is cleared by banking its unit in the -Harray- system.

The global memory unit in the -Harray- system is constructed with multi-access memory scheme. The scheme make the amount of the hardware large. Therefore, the relation between the amount (the number of port and bank) and the performance is investigated to reduce the amount of the hardware.

1. はじめに

科学技術計算の高速実行に対する要求は、計算機の処理性能が向上するにつれて、ますます増大している。こうした要求に応えるため、プロセッサを多数台並べて並列処理を行い、高速処理を実現しようとするマルチプロセッサシステムの研究がさかに行われている。その中でもデータフローコンピュータはプログラムに潜在する並列性を自然に抽出し、かつ並列処理を自然な形で実現出来ることから、注目されている。しかし、実現に際して(1)並列性を明示的に制御することが出来ないので、並列度の爆発による資源の枯渇が発生する、(2)データバケットの一致を調べる、いわゆる待ち合わせ記憶の性能がシステム全体の処理性能に大きな影響を及ぼすといった問題があることも知られている。

また、マルチプロセッサシステムにおいては、各々のプロセッサ間の結合方式も重要な問題となる。結合方式には大別してメモリ共有方式とメッセージ交換方式がある〔1〕。メモリ共有方式では、各プロセッサ間で共有のメモリ領域を持つことが出来るため、配列に代表される大域的データをP/E間で共有出来る。しかし、プロセッサ数が増加するに伴いアクセス競合も増大するといった問題がある。これに対してメッセージ交換方式では、プロセッサ間通信をI/Oポートを通してメッセージ通信の形で行うので、メモリアクセス競合といった問題が生じない。また、データが流れていくという意味から、データフローコンピュータとの整合性もよい。

これらの背景のもとに、並列処理システム-晴-では、プログラムの並列性を自然に引き出し高並列実行を行うために、前述のデータフロー方式を取り入れている。そして先に述べたデータフローにおける問題を解決するため、CDフロー方式(Controlled Data Flow)〔2〕を提案している。CDフロー方式はコントロールフローとデータフローを融合させた制御方式であり、マクロブロックと呼ぶ処理単位内でデータフロー実行を、マクロブロック間ではコントロールフロー制御を行い、明示的に並列性を制御する。また、待ち合わせ記憶の高速化のためにWMのバンク化を提案している〔3〕。

また、プロセッサ結合方式には、データフローコンピュータとの整合性のよいメッセージ交換方式を用いる。しかし、科学技術計算では大規模な配列等の構造データを扱うことが多く、この方式だけでは十分といえない。そこで、局所的なデータはメッセージ交換により通信し、大域的なデータは共有メモリを用いる方式を取り入れている。また、この共有メモリには、アクセス競合を避けるためにマルチリードメモリを採用し、さらに書き込み時のアクセス競合を軽減するためにバンク化を行う。これまでに、このような共有メモリに対する評価はいくつか行われているが、書き込み・読み出しのアクセス比率をパラメータとしたもの〔4〕〔5〕〔6〕が大半であ

り、実際のプログラムを使用したもの〔7〕は、ほとんどない。また、各プロセッサが完全にアクセス競合なしに読み込みを可能とし、書き込み時の競合も低く抑えるためには、ハードウェア量が莫大なものとなる。

本稿では、-晴-の機能レベルのソフトウェアシミュレータHSSV2(Harray Software Simulator Version 2)を用いて、実際のプログラムを動作させ、(1)待ち合わせ記憶のバンク化の評価を行いその有効性を示す。次に(2)共有メモリのポート数及びバンク数と要素プロセッサ数との関係を調べ、共有メモリにおけるハードウェア量削減の可能性をさぐる。

以下、第2章では-晴-の概要を説明する。第3章ではソフトウェアシミュレータについて述べ、第4章でシミュレーション結果を示す。

2. 並列処理システム-晴-

本章では、並列処理システム-晴-の概要を述べ、今回の評価対象である待ち合わせ記憶(WM:Waiting Memory)と共有データメモリ(GM:Global Memory)について詳しく説明する。

2.1. -晴-の実行方式

-晴-では、CDフロー方式(図1)を採用している。この方式は、プログラム全体をいくつかのマクロブロックと呼ぶタスクに分割し、マクロブロック内ではデータフロー実行をマクロブロック間ではコントロールフロー制御を行うものである。各マクロブロックは、その内部の並列度に応じて、プロセッサ群へ割り付けられる。例えば、並列度の高いベクトル演算などを含むマクロブロックは多数のプロセッサに分散され、並列に処理される。

このCDフロー方式によって、並列性を陽に秩序付けることができ、資源管理が容易になる。またIF文等によって起こる条件分岐先の先行計算をマクロブロック単位で積極的に行うことが出来る。すなわち、プログラムの制御の流れに沿ってマクロブロックを複製・融合させ、各コントロールフロー毎にプロセッサを割り付け高速化を図るフローグラフ展開〔8〕がこの一例である。

2.2. -晴-の構成

-晴-は図2に示すように6つの大きなユニットからなり、プログラムのコンパイルや初期データのロード及び結果出力のため、ホストコンピュータと接続される。

集中制御機構(GCU:Global Control Unit)は、-晴-システム全体を制御する機構であり、各要素プロセッサ(PE:Processing Element)に対し、マクロブロック実行

制御命令を出す。この命令は、起動放送ネットワーク(BCN:BroadCast Network)を通じて各PEに放送される。PEは、GCUからの命令に従ってマクロブロックの実行を開始し、その終了を同期メモリ(SYM:SYnchroNize Memory)に報告する。SYMは、複数のPE間同期を高速に実行するためのものであり、GCUはSYMの同期完了報告をチェックしながら制御を続ける。

PE間のデータ交換のためには通信ネットワーク(DCN:Data Communication Network)を用意している。また、配列等の大域的なデータは共有データメモリ(GM:Global Memory)に格納し処理を行う。

PEの構成を図3に示す。PEはGCUからのマクロブロック実行命令を起動放送制御機構(BCU:BroadCast Control Unit)が受信することにより実行を開始する。まず、データメモリ(DM:Data Memory)内にある起動マクロブロックに属するデータ(以後バケットと呼ぶ)を待ち合わせ記憶(WM:Waiting Memory)に転送する。WMではバケットの待ち合わせを行い、対となったバケットを実行機構(EX:EXecution unit)に送る。EXでは、演算を実行すると共に新しいバケットを作成し、スイッチ機構(SW:Switching unit)へ送る。SWはDCNとのインタフェースであり、他PE宛のバケットを処理する。デー

タメモリ制御機構(DMC:Data Memory Controller)は、現在実行中のマクロブロックに属するバケットをWMへ、その他のバケットをデータメモリ(Data Memory)へ格納する。これにより、WMには、現在実行中のマクロブロックに属するバケットのみが存在することになり、WMの負荷を軽くすることが出来る。

2.3. WMの構成

データフロー実行では、WM(待ち合わせ記憶)の高速化がシステム全体の高速化において重要なポイントとなる。この理由は次の2点である。

- (1) 連想検索処理が必要であり、現在のハードウェア技術では、大容量の高速連想機能の実現は困難である。
- (2) データフローでは、次にどの命令が発火するかは不確定であり、フォンノイマンコンピュータにみられるレジスタを使用出来ない。このため、メモリを直接アクセスすることになり高速な処理が望めない。

まず(1)の問題点に関しては、マクロブロック単位で実行する「晴」の方式によって、WMに溜まるバケット量を少なくすることが出来る。このため、WMの少容

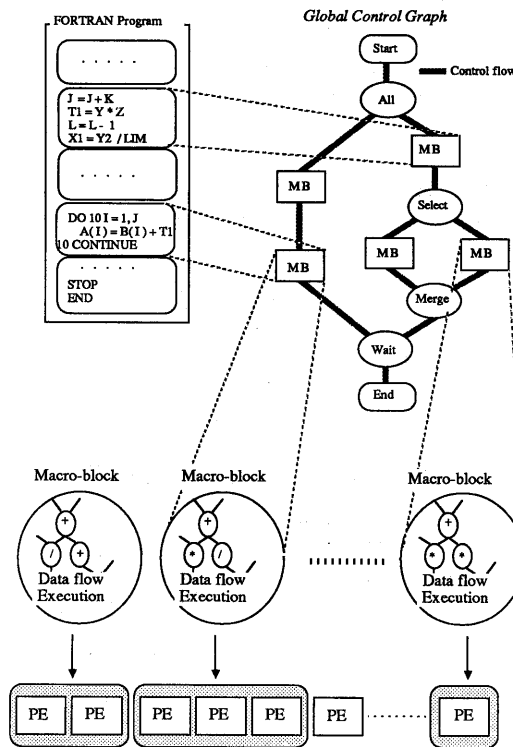


図1 Controlled Data flowの概念

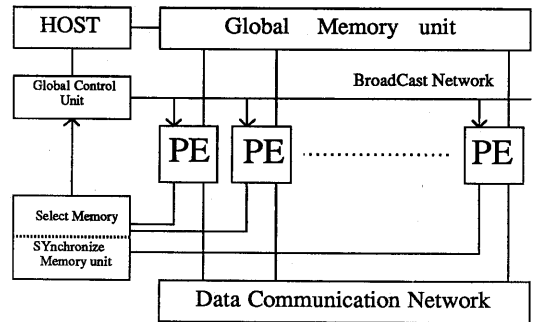


図2 晴の全体構成

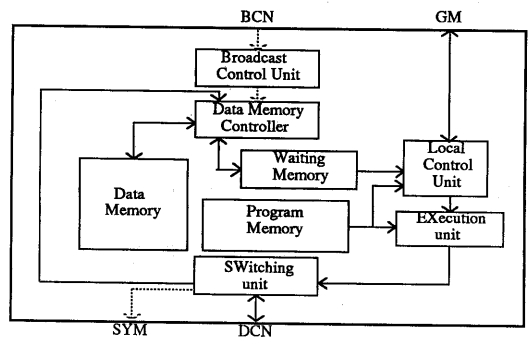


図3 PEの構成

量化が可能となる。これは、実行に際して現在必要とされないバケット（現在実行中でないマクロブロックのバケット）をDMへ格納し、WMにはすぐに必要なバケットのみを置いているからである。WMの具体的な容量に関しては、現在検討を行っている段階であり、別途報告する予定である。

(2)の問題点に関しては、WMのバンク化によって対処する。(以後、共有メモリのバンク化と区別するために、WMのバンクのことをWMバンクと呼ぶことにする)WMのバンク化とは、図4に示すようにWMに入って来るバケットを各WMバンクに分散し、スルーブット向上を図る方法である。このようにWMをバンク化した場合、各WMバンクへの負荷を出来る限り等しくすることが、WMのバンク化効率を上げるために必要となる。すなわち、入力バケットの各WMバンクへの分散方法が重要となる。

この分散方法を決定するにあたり考慮すべき点は、各WMバンクが独立に動作出来るようにバケットを分散することである。すなわち、対となるバケットは、いずれも同一のWMバンクに割り付け、そのWMバンク内で検索処理が完結するようにならなければならない。この点を考慮し、バケットの行き先ノード番号の下位2ビット(4バンク時)によってバンクを決定する方法を用いて、これまで考察して来た〔3〕〔9〕。この方法によると、対となるバケット同士は、必ず同一の行き先ノード番号を持っているため、同一のバンクに分散される。

しかし、この方法で分散を行うとノード番号の付け方による負荷のかたよりの影響を大きく受ける。特にカラーを用いたプログラムの場合には、カラーの異なった同一ノード行きのバケットが存在する為、同一WMバンクに負荷が集中しやすく、WMバンク化の効率が低下すると考えられる。そこで、カラーを考慮した、WMバンクへの分散方法を提案する。

カラーを考慮したWMへの分散は、以下の手順で行う。

(1) 行き先ノード番号とカラーの下位2ビットの和をとり、それをWMバンク番号(0~3)とする。〔4バンク化時〕

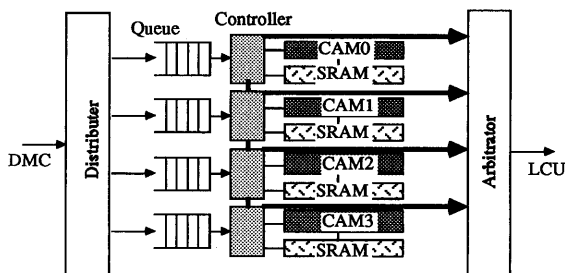


図4 WM構成(4バンク化時)

(2) スティックバケットはカラーを無視して待ち合わせを行うため、全WMバンクに対してコピーを送る。

この方法により、同じノードを行き先とするバケットもカラーが異なれば違ったWMバンクへ割り振られる。これにより、カラー使用時のノード番号の付け方による負荷のかたよりの影響を小さくすることが出来ると思われる。

2.4. GMの構成

科学技術計算では、大規模な配列等を利用したものが多く、その為、プロセッサ間結合方式としてメッセージ交換方式のみを採用した場合には、通信のトラフィックが大きくなり、配列等のデータを効率的に扱うことが出来ない。そこで、局所的なデータ通信にはメッセージ交換を用い、大域的なデータ通信には共有メモリを用いる方式を一瞥-では取り入れる。しかし、共有メモリの形態をとった場合には、複数プロセッサからのアクセス競合が問題となる。そこで、このアクセス競合を回避する解決法の1つとしてマルチリードメモリを採用する。マルチリードメモリの実現方法には、1つのメモリに複数のポートを設けたマルチポートメモリと、同一の内容をもった複数のメモリモジュールを用意し、同時読み出しを可能とするものがある。

一瞥-では、後者の方法を用いる。図5に一瞥-の共有データメモリGMの構成を示す。図5では、同一の内容をもったメモリモジュールを各PEが持ち、読み出しのアクセス競合をなくしている。しかし、書き込み時には、全メモリモジュールに対して書き込みを行い、メモリモジュール間でのデータの同一性を保証をする。このため、書き込み時にはアクセス競合が起こる。これを軽減する為に一瞥-ではバンク化を行い、異なったバンク間での並列書き込みを可能とする。

しかし、このようなメモリ形態をとると、

(1) 同一バンクへのデータ書き込みに対する調停が必要となる。

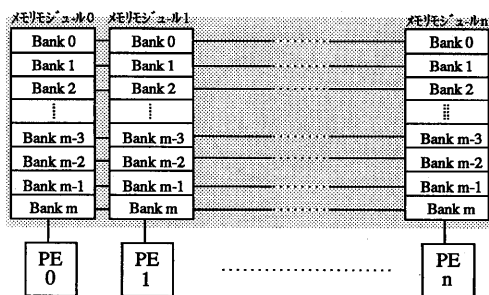


図5 GMの構成

(2) メモリモジュール数及びメモリモジュール間の接続数が膨大になるため、特に実装面で問題がある。

等が挙げられる。

これらの欠点はプロセッサ数が増加するに伴って、より困難となるため、性能をある程度犠牲にしてもハードウェアの簡略化を図ることが必要である。このため、第4章において一晴-の共有データメモリGMのポート数およびバンク数とプロセッサ数との関係を調べ、ハードウェア量削減の可能性をさぐる。

3. ソフトウェアシミュレータ

3. 1. シミュレータの概要

一晴-のアーキテクチャの妥当性を検証するため、第2章で述べた一晴-の構成を機能レベルで実現するソフトウェアシミュレータHSSV2 (Harray Software Simulator Version 2)を作成した。このシミュレータは、Micro-VAXII/GPX 上の Modula-2 言語で記述され、ソースコードで約14,000行である。入力は一晴-の機械語で記述されたプログラムである。このプログラムは、通常のFORTRANで書かれたプログラムをデータフローグラフに直した後、これを一晴-のアセンブラで記述し直し、HIAS (Harray Integrated Assembler)により機械語に翻訳したものである。

このHSSV2の主要パラメータを表1に示す。現在、各プロセッサ間の通信ネットワークDCNについては検討中であり、シミュレータでは暫定的にマルチバスを想定している。したがって、今回のシミュレーションでは、PE間での通信にDCNを使用せず、全てGMを介して通信を行っている。

表1 HSSV2の主要パラメータ

項目	設定値
システムクロック	10 MHz
PE内モジュール間 パケット通信速度	200ns/Packet
ネットワーク(DCN)通信速度	3MPackets/s
演算時間 (除算)	300 ns
(上記以外)	1500 ns
待ち合わせ記憶(WM)	
処理時間(平均)	700 ns/Packet
バンク数	4/PE
ALU数	2/PE
メモリ・アクセス時間	
Program / Data Memory	100 ns
Global Memory	400 ns

3. 2. 実行プログラム

科学技術計算においては、大規模な非対称行列を係数行列に持つ連立一次方程式を扱う場合が多い。そこで、連立一次方程式解法に使用されるプログラムの中から3つを取り上げ解析の対象とする。連立一次方程式の解法には大きく分けて、直接法と反復法がある。直接法から(a) LU分解と(b) ガウスの消去法を、反復法から(c) 双対共役勾配(BCG)法を取り挙げる。各プログラム共に、 2×2 、 4×4 、 8×8 、 16×16 の行列を扱った場合についてシミュレーションを行う。

4. シミュレーション結果

4. 1. WMのバンク化の評価

WMのバンク化により処理速度がどの程度向上するかを調べるために、1台のPEを用いてシミュレーションを行う。そして、WMバンク化しない時の実行時間を基準(=1)とした時の処理速度比を求める。プログラムは、(a) LU分解、(b) ガウスの消去法、(c) BCG法について、 2×2 及び 8×8 行列を実行させる。

また、各WMバンクへの分散方法は、

N) ノード番号の下位ビットをそのままWMバンク番号とするこれまでの方法。

N+C) ノード番号とカラー番号の下位ビット同士の和をとり、それをWMバンク番号とする方法。

の2種類を用いる。図6にこの結果を示す。

まず、 8×8 行列を実行した時の結果について考察する。 8×8 行列を実行した時、(a) (b) (c) いずれの結果においても、カラーを考慮した分散方法の方が、ノード番号のみで分散する方法に比較して効率がよい。また、WMバンク数が2の時と4の時とを比較すると、WMバンク数が2の時の方がこの傾向が顕著に表れており、WMバンク数4の時は、どちらの分散方法を用いた場合でもほぼ同じ効率を示している。これは、プログラムの性質によるものと考えることが出来る。実際に、WMをバンク化せずにWMの処理速度を4倍に設定してシミュレーションを行ったところ、3種類のプログラムについていずれも、4バンク化した時の値とほぼ一致した。WMの処理速度を4倍にした結果と4バンク化した時の結果がほぼ等しいということは、プログラムの性質(並列度の変化等)によって4バンク化した時に、これ以上の高速化が望めないことを示す。すなわち、4バンク化時にはパケットの分散方法よりも、プログラムの性質によって高速化が妨げられているといえる。

次に 2×2 行列を実行した場合の結果について考察す

る。この場合、(a) (b) のプログラムでは、カラーを考慮しない分散方法の方が、カラーを考慮した方法に比較して性能がよい。また、(c) のプログラムの場合においても、カラーを考慮した場合としない場合での差がほとんど見られない。これは、 2×2 行列の場合、並列度が小さいために、カラーを考慮した分散方法を行うことによるオーバーヘッドが表れるためだと考える。このオーバーヘッドはスティッキーバケットをWMが全バンクにコピーして渡すことによって生じる。

最後に 2×2 行列と 8×8 行列の結果を比較すると 8×8 行列の方が処理性能がよい。これは、プログラムの並列度が増すことにより、WMのバンクが有効に利用されることに起因する。(C) のBCG法では、 2×2 と 8×8 の時の性能の差が、他の2つのプログラムに比較して小さい。これは、BCG法は処理の途中で N^2 に比例して並列度が出る部分もあるが、総計算のために逐次に実行しなければならない部分もあり、平均すると並列度があまり増加しないことに起因している。

以上の結果より、カラーを考慮したバケットの分散方法は、考慮しない方法に比較して、プログラムの並列度があまり小さくない場合において効率のよい分散を実現出来ることがわかる。また、4バンク化した時には、分散方法よりもプログラムの性質によって処理性能が抑えられるという傾向があることがわかる。したがって、WMのバンク化を考えた場合、2バンク化の時はカラーを考慮した分散方法により1.7倍前後の処理性能向上が見込まれるが、4バンク化した時は、プログラムの並列度が十分になればコストパフォーマンスがよくないことがわかる。

4. 2. GMの評価

マルチリードメモリに関する諸特性のうち、パフォーマンスに大きな影響を与えるものに、メモリへのポート数(メモリモジュール数)、バンク数、メモリモジュール間でのデータ転送時間等が挙げられる。ここでは、この内のポート数とバンク数についてPE数との関係を調べ、一時的共有データメモリであるGMのハードウェア量削減の可能性を調べる。

ポート数とは、一度にメモリにアクセス出来るPEの数である。図5では、各PEが1つのポート(メモリモジュール)を持っており、PE数=ポート数である。このポート数を少なくした時の性能に与える影響を調べる。

バンク数とは、書き込みバンク数のことであり、同一バンクに対して書き込みの出来るPE数は、同一時期には1台に限られる。ただし、読み出しは、そのバンクに対しての書き込みが行われていなければ、バンク数に関係なく全てのPEがアクセス可能であるとする。また、配列の各バンクへの分割は、配列の添え字番号をバンク数で割った余りでグループ化した。ここでは、このバン

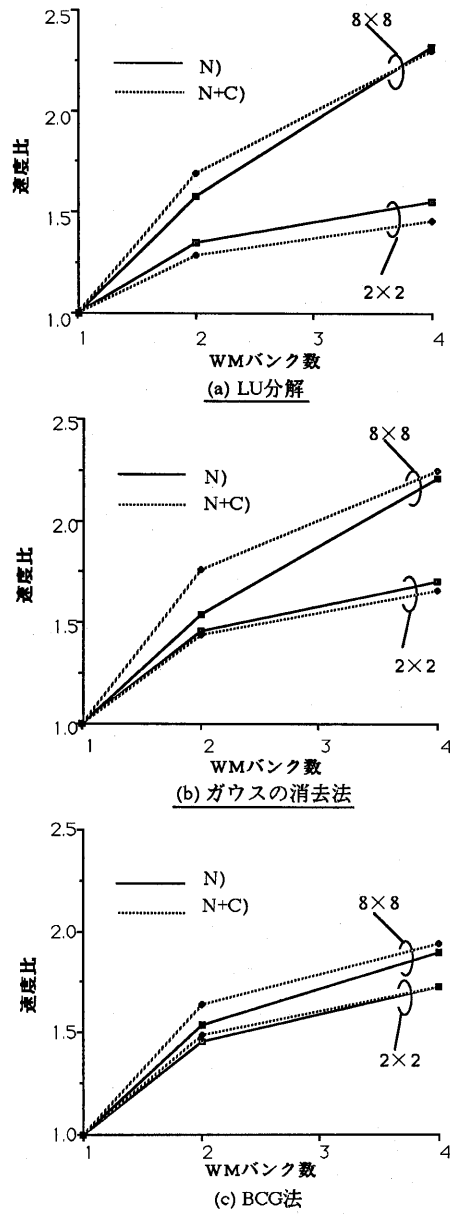


図6 WMバンク化による処理速度向上

N) ノード番号による分散方法
 N+C) ノード番号とカラーによる分散方法
 (WMバンク化しない時を基準とした速度比を示す)

ク数を変化させた時の性能に与える影響を調べる。

メモリモジュール間でのデータ転送時間は、今回は考慮せず0と仮定した。しかし、現実には0にすることは不可能であり、モジュール数の増加に伴い、モジュール間での遅延が生じるので、今後検討していく。

a) ポート数の影響

結果を図7に示す。横軸に、PE数で正規化したポート数を取り、縦軸には、ポート数=PE数とした時の実行時間を基準とした時の速度比をとる。また、この時バンク数の影響を排除するために、バンク数は無限大(1ワード1バンク)と仮定した。また、ポート数による影響はプログラムの並列度にも依存しているため、PE数とプログラムの最大並列度の比をパラメータとした。

この結果より、(a) (b) (c) いずれのプログラムにおいても、ポート数がPE数の2分の1以上であれば、処理速度にほとんど影響を与えないことがわかる。また、プログラムの最大並列度にPE数が近づくにつれ、ポート数はさらに少なくてよいことを、これらの結果は示している。これは、PE数が増大すると1PEあたりのメモリアクセス量が減少するため、全てのPEが同時にポートを要求する確率が減少することに起因していると考えられる。

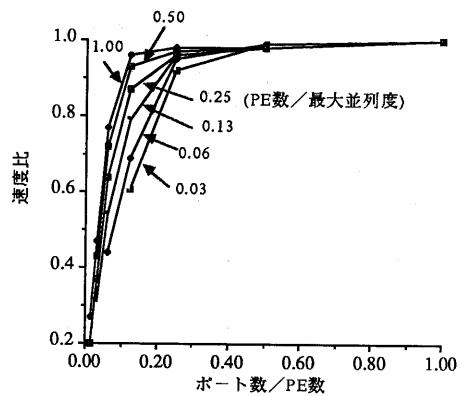
b) バンク数の影響

結果を図8に示す。横軸に、PE数で正規化したバンク数を取り、縦軸には、バンク数を無限とした時の実行時間を基準とした時の速度比をとる。また、この時のポート数による影響を排除するために、ポート数は無限大と仮定した。また、PE数が8, 16, 32台の場合について、行列が 8×8 、 16×16 の場合についての結果を示す。

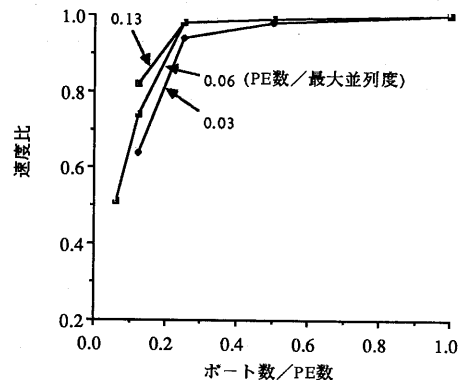
この結果より、バンク数がPE数の2分の1以上であれば、処理速度にほとんど影響を与えないことがわかる。また、1割程度の処理速度低下を許せば、バンク数はPE数の4分の1程度でもよいことがわかる。

ポート数とバンク数が処理速度に与える影響を比較すると、明らかにポート数のほうが大きな影響を与えることがわかる。一般にメモリへの書き込みと読み出しの比率は、読み出しの方が大きいことを考えると、書き込み時のアクセス競合を軽減させるバンク数の増加よりも、読み込み時のアクセス競合を軽減させるポート数の増加の方が重要であることを示している。

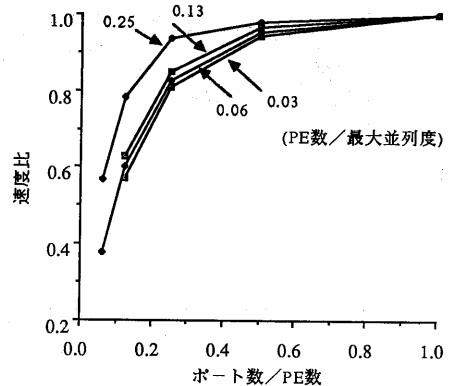
以上の結果より、ポート数はPE数の2分の1、バンク数もPE数の2分の1までであれば、処理速度に与える影響は無視出来ることがわかった。



(a) LU分解



(b) ガウスの消去法



(c) BCG法

図7 ポート数の影響

(バンク数=∞, ポート数=∞とした時を基準とした速度比を示す)

5. おわりに

本稿では、一晴一の機能レベルのソフトウェアシミュレータを用いて、(1) 待ち合わせ記憶WMのバンク化と(2) 共有データメモリGMの評価を行った。その結果、WMのバンク化については、バンク化時のバケット分散の方法として、カラーを考慮した方法が考慮しない方法に比較して効率がよいことを確認した。また、共有データメモリの評価においては、ポート数・バンク数共にPE数の2分の1までに削減しても処理速度にほとんど影響を与えないことを確認した。

今後は、さらに多くのプログラムについてこれらの結果を検証すると共に、このソフトウェアシミュレータを用いて一晴一の性能評価を進めていく予定である。

謝辞

シミュレータ作成及び本研究の遂行にあたり御支援いただいた本研究室の樋口和広氏に感謝いたします。

参考文献

- [1] 富田: "並列計算機構成論", 昭晃堂, PP.195-240, (1986)
- [2] H.Yamana, T.Marushima, T.Hagiwara, Y.Muraoka: "System Architecture of Parallel Processing System -Harray-", Proc of Int. Conf. on Supercomputing, pp.76-89, (1988)
- [3] 山名, 丸島, 草野, 村岡: "並列処理システム一晴一の要素プロセッサ構成", 情処研報, Vol.88, No.19, 88-CA-70-4, (1988)
- [4] 久野: "競合問題の解決を目差すメモリ構成法 -TRANSPARENT MEMORY", 信学論文誌, Vol.69-D, No.2, pp.128-135, (1986)
- [5] 内堀, 古谷, 西田: "マルチリード・メモリの動作解析", 第27回情処全大, 5N-2, (1982)
- [6] 天野, 地川, 吉田, 相磯: "マルチリード・メモリを用いた並列計算機の性能解析", 信学論文誌, Vol.J67-D, No.9, pp.1028-1035, (1984)
- [7] 曾和, 上村: "データフローコンピュータDFNR-2のシミュレーションによる性能評価", 信学論文誌, Vol.J69-D, No.7, pp.1054-1065, (1986)
- [8] 萩原, 山名, 村岡: "並列処理システム一晴一の並列化コンパイラの設計について", 信学技報, CPSY88-32, PP.43-48, (1988)
- [9] 山名, 草野, 萩原, 村岡: "並列処理システム一晴一の待ち合わせ記憶構成", 第37回情処全大, 1N-2, (1988)

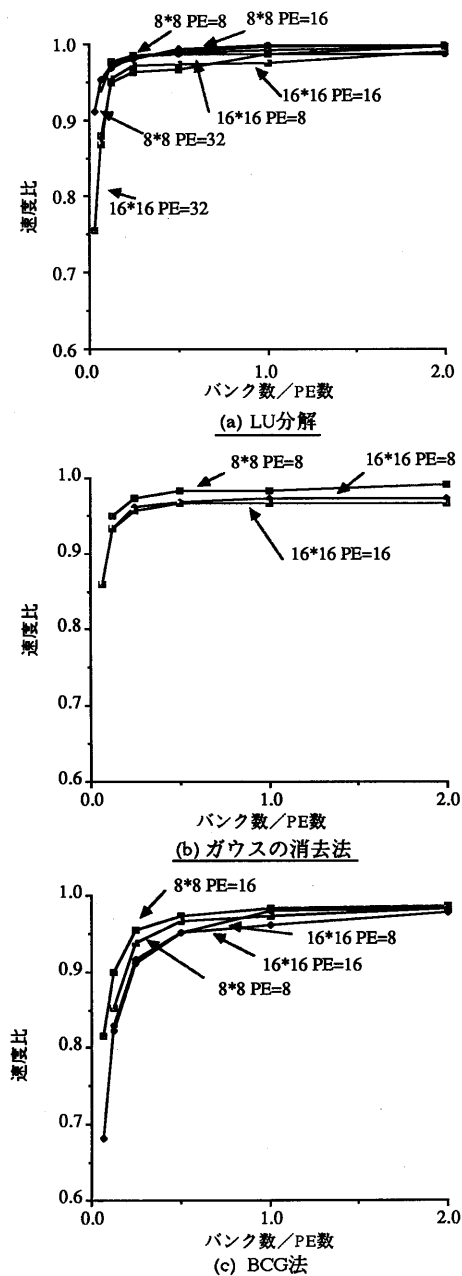


図8 バンク数の影響

(バンク数=∞,ポート数=∞とした時を基準とした速度比を示す)