計算機アーキテクチャ　95－6
(1992. 8. 19)

グループ共有メモリを持った柔軟な格子結合型並列計算機（ＦＭＭ）

*塚原宏享　*岩根雅彦　**浦崎直彦

*九州工業大学，　**東芝北九州工場

ＦＭＭはいくつかのＰＵをグループとしてあつかい，システム上に複数のグループを生成できる．生成された各グループは，SPMD型処理，MIMD型処理，MIMD／SPMD融合型処理のいずれかをそれぞれグループごとに独立に処理を行う．本稿では，ＦＭＭのアーキテクチャ，ＰＵのグループ化および，グループにおける同期／排他制御，グループごとに独立な共有メモリ（ＧＳＭ）について述る．また，ＧＳＭについてはモデル化を行い，待ち行列によりその性能予測を行う．

Flexible mesh-network multi-microprocessors (FMM) with group shared memory (GSM)

Hiroyuki Tsukahara, Masahiko Iwane, Naohiko Urasaki

Kyusyu Institute of Technology
Kitakyusyu Works Toshiba Corporation

FMM (Flexible Mesh-network Multi-microprocessor) can be dynamically reconfigured to operate as the groups which consist of one or more SPMD, MIMD or Mixed machines.

This paper describes the architecture of FMM, grouping of the PUs, syncronize/mutual exclusion operation on the group, and the group shared memory. Moreover we model the GSM and evaluate the performance of GSM by the queuing theory.

1．はじめに

マルチマイクロプロセッサ並列計算機においては，個々のPU（要素プロセッサ）で均一な処理を行なうSPMD型処理，不均一で個々の処理が高い独立性をもつMIMD型処理がある．これらの処理を行なうとき計算機上の全ての要素プロセッサを使用して処理を行なうのが常に最適とは限らず，処理すべき問題の並列性，PU間通信やデータ分配などのオーバーヘッドによって処理するときの最適な要素プロセッサの数が決まる．

FMM（Flexible Mesh-network Multimicroprocessors）ではPUを有効に利用するため，いくつかのPUを一つのグループとして扱いシステム上に複数のグループを生成し，それぞれのグループを独立に動作させることができる．またグループにおいて，SPMD（Single Program Multiple Data）型処理，MIMD型処理，MIMD／SPMD融合型処理が動作する．各グループごとにグループ内でのみ有効な共有メモリ（GSM：Group Shared Memory）が用意され，またSPMD型処理のためのグループ内での同期制御，共有変数更新のためのグループ内での排他制御が行なえる．

本稿ではFMMのアーキテクチャ及びGSMの構成について述べ，次に GSMについてのシステムの性能予測を行なう．

2．ＦＭＭのアーキテクチャ

2.1 システム構成

FMMの構成を図１に示す．FMMはHC（Host Computer），IIU（Integrated Interface Unit），および６４台のPUから構成される．HC-IIU間はHCバスで，IIU-PU間はDTバス及びINTバスで，PU間はトーラス網及びGSMバスで結合している．

HCバスは HCとIIU間のデータ転送，DTバスはHCとIIU間のデータ転送，INTバスは PUからIIUへの割込みのときの割込み種別およびパラメータ転送，トーラス網はPU間の４隣接方向へのデータ転送，GSMバスは同一グループにおいてPUのGSMの内容を更新するために使用される．

2.2 HC（Host Computer）

HCは80386MPU，２MBのメモリ，入出力装置で構成され，OSとしてMS-DOSをマルチプロセス用に拡張したMS-DOS（MS-DOS Extention）が使用される．またHC上ではPUのグループ管理，プロセス管理，入出力管理，PU上へのプログラムのロード，ユーザーインターフェース，SPMD型処理の制御などが行なわれる．

2.3 IIU（Integrated Interface Unit）

IIUの構成を図２に示す．IIU は３つのユニットから構成されている．

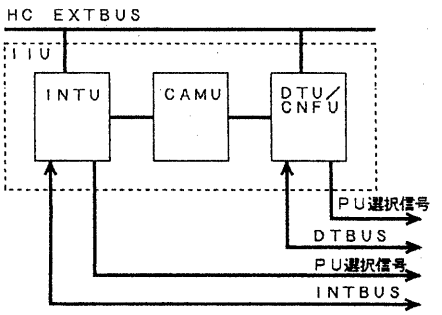

図２ IIU構成図

(1) DTU/CNFU（データ転送／構成制御ユニット）

DTUは HCメモリーPUメモリまたはPUメモリ間での DMA転送を行い，CNFUはHCからのCAMUの更新および DTUによるメモリ間転送のときのPUの選択に使用される．

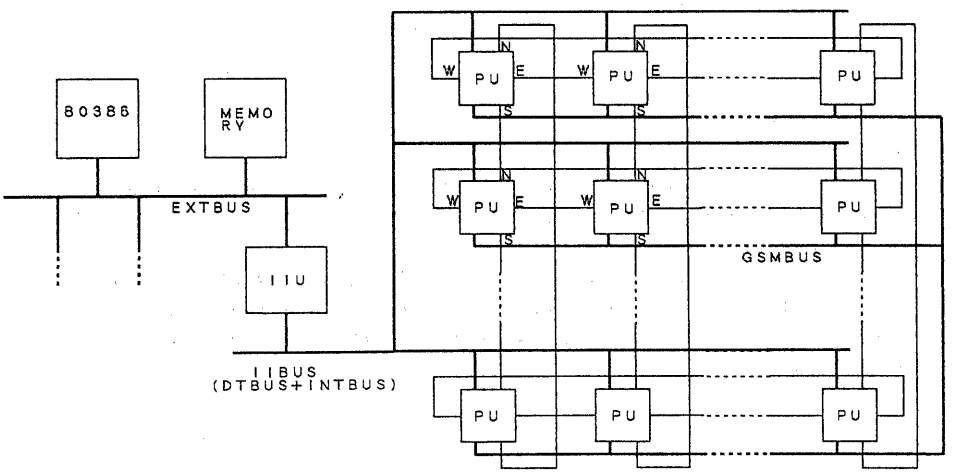

図１ システム構成図

(2) CAMU（ＣＡＭユニット）

このユニットは CAM（Content Addressable Memory）を用いて図３のように，グループの情報をグループの番号とそのグループに属するPUをビットパターンで表したPUパターン部により管理しており，CNFU，INTUから使用される．

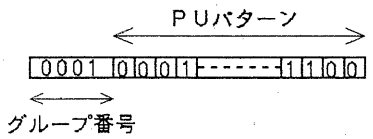

図３　グループ情報

CAMUはグループ番号が入力されるとそのグループに属するPUをビットパターンとして出力し，PUのビットパターンが入力されるとそのビットパターンに包含されるグループ番号が出力される．

(3) INTU（割込みユニット）

INTUはグループ化されたPUに対して同期や排他制御といった機能を行うためのユニットである．PUはINTUに対して割込みという形でこれらの機能を要求し，INTUは，割込んだPUから送られてくるPU番号またはグループ番号をもとにCAMUからグループ情報を引き出して機能を実現する．INTUがグループに対してサービスを行なう場合，INTUはCAMUより引きだしたPUパターンを用いPUを選択するためのPU選択信号を生成し，サービス対象のPUをそのグループのみに制限する．

INTUで処理されるPUからの割込みには以下のようなものがある．

(a) Lock　：同一グループ内の他のPUの実行を中断する．

(b) Unlock ：Lockで中断している同一グループのPUの実行を再開する．

(c) Barrier：グループ内PUの同期をとるために同一グループの全てのPUがBarrierを発生するまで実行を停止する．

(d) HSYNC　：グループ内PUの同期をとるために同一グループの全てのPUがHSYNCを発生するとHCに対して割込みを発生する．

(e) Service Call：PUからHCへのサービス要求で．INTUはPUからこの割込みを受けるとHCへ割込みを発生する．

2.4 PU (Processing Unit)

図４にPUの構成図を示す．PUは MPUとして8088，8087，メモリとしてLM（Local Memory），GSM（Group Shared Memory），ROMをもつ．

また各PUは内部に自分がどのグループに属しているかを表すGPNOレジスタ，PU番号を表すPUNOレジスタ，HCよりPUを制御するためのRCTLレジスタ，HCに対するサービスコールのときHCに渡すパラメータを格納するパラメータレジスタ，GSMとLMの境界を表すBOUNDレジスタをもつ．

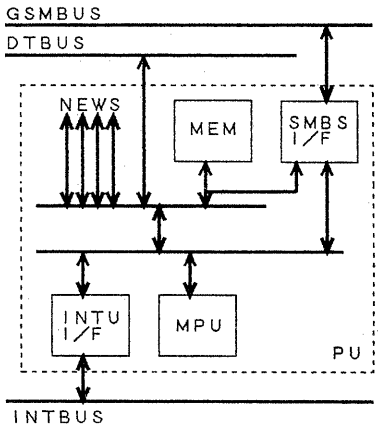

図４　PU構成図

3　グループ管理機能

3.1 グループ動作

FMM ではグループ単位に処理を行い，処理の形態は，MIMD型処理，SPMD型処理があり，SPMD型処理は，HCが処理の制御を行なうHC制御SPMD型処理，PUが処理の制御を行なうPU制御SPMD型処理の２つがある．

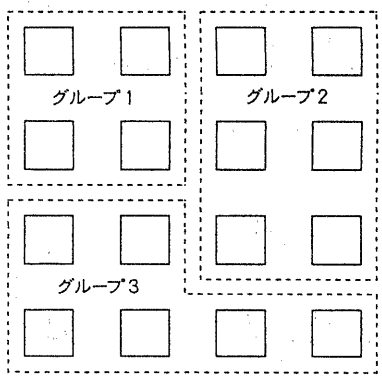

図５　複数グループ動作

図５はシステム上に複数のグループを生成した状態の例を表している．例えば図５の，グループ１はある問題をSPMD型処理を実行し，グループ２もまた別の問題に対しSPMD型処理を行い，グループ３はMIMD型処理で別の問題を解くといったことが同時にできる．またFMMでは MIMD型処理グループのあるPUで行なわれる処理をSPMD型処理で処理させるMIMD／SPMD融合型処理がある．これはMIMD

型処理中のあるPUが親プロセスとなり，子プロセスとしてSPMD型処理を行なうため，空きPUをサブグループとして獲得，サブグループへの問題の割り当て，結果の回収を行うことにより実現している．

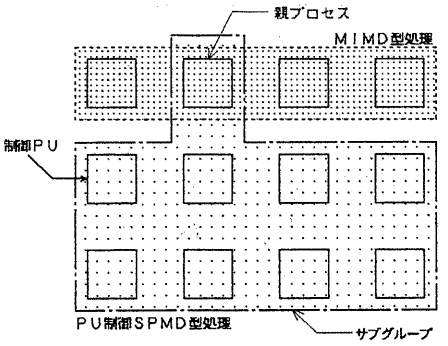

図６　MIMD／SPMD融合型処理

3.2　グループ内同期

PUはこれらの機能，HSYNC，BARRIER，を使用するためにINTUに対して割込みをかける．割込みがINTUに受け付けられるとPUは INTバスを介してINTUに対して割込み識別と自分のPUNOを送る．INTUは送られてきた割込み識別より同期割込みであることを判定する．つぎに，送られてきたPUNOをもとにINTU内部にあるCAMUのグループ情報と同じ構造をした，割込みレジスタのPUパターン部を更新する．そして，この更新された割込みレジスタのPUパターン部に包含されるグループの情報をCAMUより引き出す．このときグループ内のPUが全て同期割込みを行っていればINTUはCAMUからのグループ情報を引き出すことができ，INTUは引き出した情報を CAM読み出しレジスタに格納する．そうでないとき，INTUはCAMUからエラーを受け取る．INTUはCAMUからエラーを受け取ったときは何もせず次の割込みを待つ．一方，成功したときは，引き出されたグループ情報のPUパターン部を割込みレジスターのPUパターン部から削除し，HSYNCのときはINTUはHCに対して割込みをかけグループの処理の同期がとれたことをHCに知らる．BARRIER のときは引きだした情報のPUパターン部を用いてPU選択信号を生成し，グループのPUのみに対して処理の再開信号を送り処理を再開させる．

3.3　グループ内排他制御

グループ内排他制御とは，一つのグループ内のPU間で排他制御を行うための機能であり，これはグループが持っている共有メモリGSM を更新するときにPU間で排他制御を行うときに用いられる．

同期の時と同じようにPUはINTUへの割込みによってこれらの機能を使用する．

図中の割込み発生PUはINTUへ割り込んだ後，割込み識別と自分の属しているグループの番号GPNOをINTUに送る．INTUは割込み識別より割込みの種類を判定しLOCKまたはUNLOCKの処理を開始する．

まずINTUはPUより送られてきたGPNOを用い，このグループの情報をCAMUから引き出しCAM読み出しレジスタに格納する．つぎに引き出した情報のPUパターン部を用いてPU選択信号を生成しサービスの対象であるPUを選択する．PU選択後，INTUはLOCKの場合は処理を中断させる信号を，UNLOCKの場合は処理の再開の信号を送る．

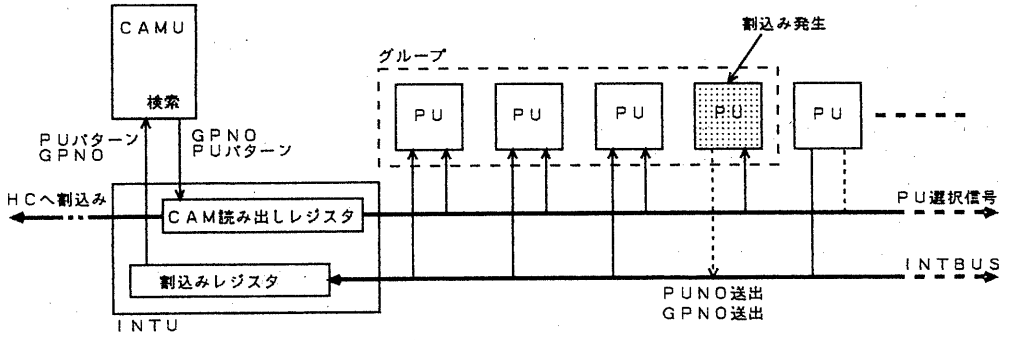

図７　グループ動作

4．ＧＳＭの構成

4.1 メモリ構成

ユーザーはPU内部のメモリをLMとGSMに 区別するために境界を表すBOUNDレジスタにGSMの開始アドレスを設定する・この開始アドレスはグループごとに任意に設定でき，図８のように各グループごとに異なったサイズの GSMを持たせることができる．

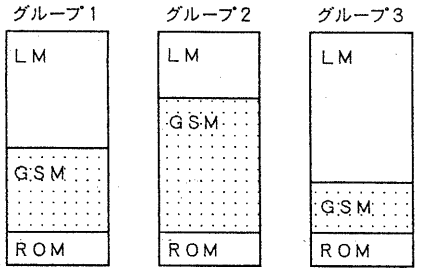

図８　グループごとのGSM

区別されたメモリでは，LMは実行されるプログラムやそのPU内部のみで使われるローカルなデータなどの格納，GSMは グループごとに独立なグループ内の共有メモリとして使用される．

図９に示すようにPU内部のメモリはDPM（Dual Port Memory）によって構成されており，メモリのポートの一つはPU内部バスへ，もう一方はGSMバスへ結合されている．またPU間はGSMバス との入出力インターフェースを持ち，GSMバス のアービトレーションは各PUの出力インターフェースがPUNOレジスタを用いてFuturebusの方法によって行なう．

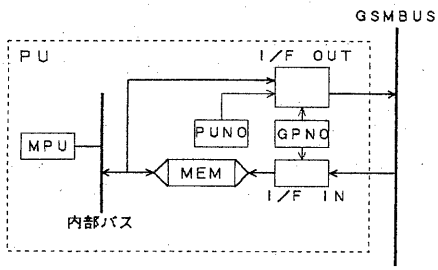

図９　GSM構成図

4.2 メモリ動作

GSMは，PU内部におけるMPUのメモリへの書込の監視，および書込発生のときの書込内容の同一グループの他のPUへのマルチキャストによって実現している．

PUにおいて GSMへの書込が発生するとPUの出力インターフェースが GSMバスの獲得を行なう．バスの獲得に成功すると次にGPNOレジスタの内容，書込アドレス・データをバス上に出力し書込信号を送る．一方，受け取り側のPUでは入力インターフェースが GSMバス上のGPNOと自分のGPNOと比較し同一のGPNOであればバス上のデータの書込を許可する．アプリケーションからの GSMの読み込みは，DPMのPU内部バス側から行なう．

5．ＧＳＭのモデル化

5.1 モデル化

GSMの性能を待ち行列による予測を行なうために以下の仮定の下でモデル化を行なう．

1) 各PUでは，ある一定のマシンサイクルの後に必ずメモリアクセスを行う．
2) PUのGSMバス要求の到着はポアソン分布に従うとする．
3) GSMのサービスは到着順に行われる．
4) DPM において同時に同一アドレスへの読み込みと書き込みは無視できるものとする．

各PUを客，GSMバスをサービスを行なう窓口としたM/D/1モデルを図９に示す．

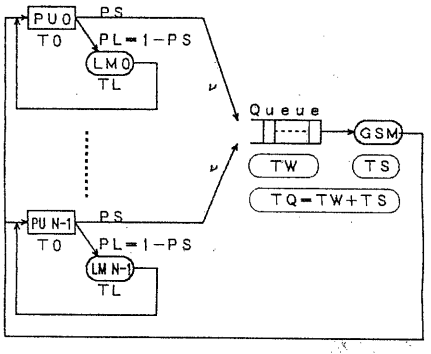

図９　ＧＳＭモデル図

ここで，
T0：PUにおける基本命令時間
TL：LM read/write時間
TS：GSMバスのサービス時間
TW：平均待ち時間
TQ：平均系内時間
PS：各PUのGSM write確率
PL：LM read/write確率　(PL=1-PS)
ν：窓口への到着率
N ：PUの数
ρ：窓口の利用率
である．

5.2 平均系内時間TQ

このモデルに対して，PUの数が無限であるとした無限母集団，有限であるとした有限母集団でそれぞれ待ちを含めた GSMバスの平均系内時間TQを

求める．

各PUのGSM書込を行なう間隔TUは，PL，PSより

$$TU = PL(T0 + TL) / PS + T0 \quad (1)$$

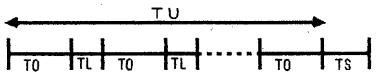

図10　到着間隔

と考えられる．これより，１台のPUの単位時間での窓口への到着率νは

$$\nu = 1 / TU \quad (2)$$

である．

無限母集団の場合，窓口の利用率ρは

$$\rho = TS \cdot N \cdot PS / (T0 + PL \cdot TL) \quad (3)$$

平均系内時間TQは，

$$TQ = TS + TS \cdot \rho / (2(1 - \rho)) \quad (4)$$

となる．

有限母集団の場合，窓口の利用率ρは

$$\rho = N \cdot Y / (N \cdot Y + Z) \quad (5)$$

平均系内時間TQは

$$TQ = TS(N - Z \cdot \rho) / \rho \quad (6)$$

但し，

$$Z = 1 / (TS \cdot \nu) \quad (7)$$

$$Y = 1 + {}_{N-1}C_1 (e^{1/Z} - 1) + {}_{N-1}C_2 (e^{1/Z} - 1)(e^{2/Z} - 1) \cdots\cdots + {}_{N-1}C_{N-1} (e^{1/Z} - 1)(e^{2/Z} - 1) \cdots\cdots (e^{N-1/Z} - 1) \quad (8)$$

6．性能予測

6.1　実行時間増加

PUがGSMに対して書込みを行ったとき，待ちが生じない場合のPUの平均実行時間をTEとするとTEは

$$TE = T0 + PS \times TS + (1 - PS)TL \quad (9)$$

GSM書込み時に待ちが生じた場合のPUの平均実行時間をTE'とするとTE'は

$$TE' = T0 + PS \times TQ + (1 - PS)TL \quad (10)$$

となる．これより待ちが生じた場合の実行時間の増加率を次のように表すことができる．

$$Tinc = (TE' / TE - 1) \times 100 \ (\%) \quad (11)$$

このTincを性能予測式とする．この式においてPUの台数，GSM書込み率を変化させシステムの性能予測を行ってみる．

6.2　性能予測

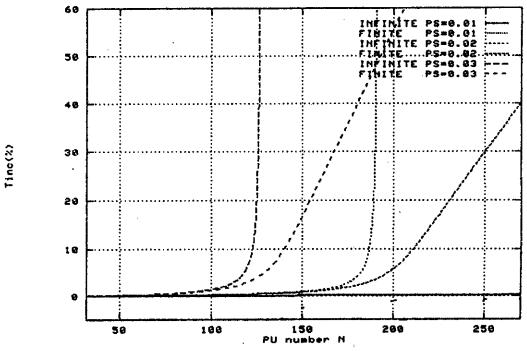

図11　ﾊﾞｲﾄｱｸｾｽ実行時間増加

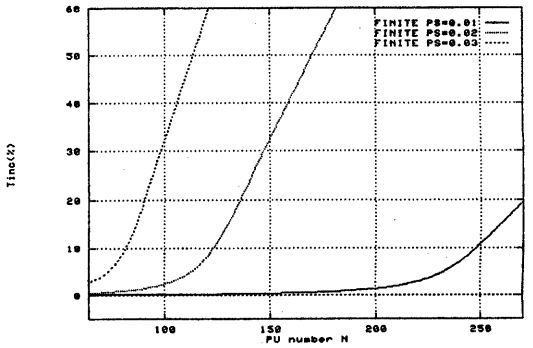

図12　ﾜｰﾄﾞｱｸｾｽ実行時間増加

FMMの場合の各パラメータは，バイトアクセスの場合，T0＝2382(ns)，TL＝500(ns)，TS＝750(ns)，ワードアクセスの場合，T0＝2382(ns)，TL＝1000(ns)，TS＝1500(ns)であり，図11はバイトアクセス，図12はワードアクセスの結果である．

図11は無限母集団，有限母集団の両方の結果を示している．この結果より，無限母集団として扱った場合は，PUの数を無限としているため GSMバス（窓口）の利用率ρが１となる付近で急激に実行時間が増加している．実際にシステムにおいて，このような急激な実行時間の増加は考えられず，また無限母集団は利用率ρ＜１を仮定しておりρが１を越える範囲では無限母集団の結果は適用できない．よってシステムの性能予測には有限母集団の平均系内時間TQを使用することにする．

図11の有限母集団の結果より，バイトアクセスの場合は，GSM書込み率PSが２％，PUの台数が２５６台であれば，実行時間増加は４０％程度で，一方，図12の結果からワードアクセスの場合は，GSM書込み率PSが２％のとき 実行時間増加が４０％となるのは，PUの台数が１６０台程度のときであ

る．このようにワードアクセスの実行時間増加がバイトの場合に比べ早いのは，GSMバスが8088MPUの特性に合わせて設計されているためである．バイトアクセスの場合は１回のバスサービスですむが，ワードアクセスの場合２回のサービスを必要とする．そのためワードの場合バイトに比べ１つのPUのバス占有時間が長くなり，GSMバスが混雑するためである．

現在のシステムではPUの台数は６４台でありGSM書込み率PSが ２％程度であればバイト，ワードアクセスともにほとんど実行時間の増加は無いと予測される．しかし今後システムを拡張しPUの台数を増やすならば，今のシステムではアプリケーションにおいて GSMへのワードアクセス増加するとシステムの性能低下が予測される．これを回避するためにはバイト，ワードアクセスともに１回のメモリサイクルで行える8086などに変更し，GSMバスもそれに合わせ，１回のバスサービスで行えるように変更する必要がある．

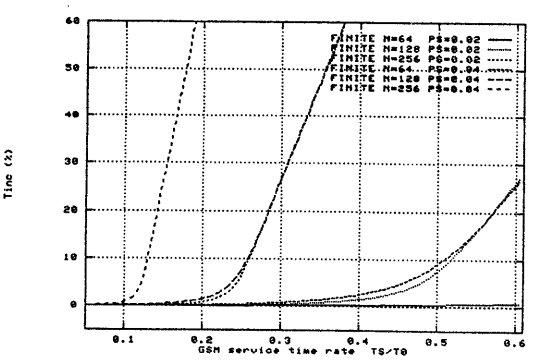

図13　T0,TS比による実行時間増加

図13は 基本命令実行時間T0とGSMバスサービス時間の比をもとにしたバイトアクセスの実行時間増加の結果である．現在のシステムは，TS/T0＝0.3，PU台数N＝64であり，GSM書込み率PS＝0.4のとき，図13の結果からは，ほとんど性能低下は見られない．

MPUを高速にした場合，基本命令実行時間T0は小さくなりTS/T0は大きくなる．例えば現在のシステムでMPUを倍の動作スピードにしたと仮定すると，TS/T0は２倍の0.6となり，GSM書込み率PS＝0.4の場合，グラフより２５％以上の実行時間増加が予測される．

また現在のシステム上のPUの台数を２５６台まで拡張した場合，GSM書込み率PS＝0.2であれば，実行時間増加をなくすためにはTS/T0＝0.2程度，つまりバスのサービス時間TSを1.5倍，速くしなければならないと予測される．

７．結び

FMMのアーキテクチャについて述べDPMを用いたGSMについての性能予測を行ってきた．GSMは，読込みについては通常のLMと同じであるので，書込み頻度が十分小さければ GSMバスは混雑せずシステムの性能低下も少ないと予測される．また今後，MPU を高速なものにし，PUの台数を増やすならばバス幅の変更及びバスのサービススピードの高速化が必要となるであろう．

今後の課題としては，特定のアプリケーションの性質を考慮した予測，モデルの予測結果と実際のシステムでの実測との比較，実行時間の増加と並列化による高速化とのトレードオフの検討，である．

現在，FMMのシステムは ４台のPUを用いデバッグ中である．

［参考文献］

[1] R.Duncan, A survey of Parallel Computer Architecture, Computer, Feb, 1990, p.p.5－p.p.16

[2] T.Hoshino, Invitation to the world of PAX, Computer, May, 1986, p.p.69－p.p.79

[3] H.J.Sigel, PASM:A Partitionable SIMD/MIMD system for Image Processing and Pattern Recognition, IEEE Trans Computer, Vol c-30, No.12, Dec, 1981

[4] H.Li and Q.F.stout, Reconfigurable SIMD Massively Parallel Computers, Proc.IEEE, Vol79, No.4

[5] S.F.Lundstrom, Applications Considerations in the System Design of Highly Concurrent Multiprocessors, IEEE Trans Computer, Vol.C-36, No.11, Nov, 1987, p.p.1292－p.p.1309

[6] 富野眞治著，＂並列計算機構成論＂，昭晃堂，1986

[7] 高橋義造編，＂並列処理機構＂，丸善，1989

[8] 森村英典，大前義次，＂応用待ち行列理論＂，日科技連，1975

[9] 村田他編，＂並列コンピュータアーキテクチャ（ビット臨時増刊），共立出版，1987