

## 発注／受領型並列計算機の例外処理とデバッグ支援機構

田村 仁 富澤真樹

東京農工大学 工学部

我々は並列手続き呼出しを計算モデルとした並列計算機の分散制御機構を提案している。これをハードウェアに組み込んだPEの通常状態では、ユーザの手続きが論理的に多重環境下で並行動作する。しかし例外処理は物理的に1つの環境で実行されなければならない。このために別の動作状態を設定する。この動作状態モデルと整合する例外処理機構を構築した。並列計算機の例外を例外発生PEだけで処理でき、他PEに影響しない『解析例外』と、全PEを凍結し並列計算機全体で処理する『凍結例外』に分ける。そしてこれらを任意に起こして並列計算機用ブレイクポイントとして利用したり、イベントヒストリーを取得するデバッグ支援機構を構築した。

### Exceptions Handling and Debugging Support Mechanism for the Demand/Accept Multiprocessor

Hitoshi TAMURA Masaki TOMISAWA

Faculty of Technology, Tokyo University of Agriculture and Technology

2-24-16 Naka-chou, Koganei-shi, Tokyo 184, Japan

We have proposed a distributed control mechanism based parallel procedure call for a multiprocessor. This mechanism generates multiple-environments for procedures in a PE, and executes cocurrently. But exceptions must be handled on a static environment. So, we make a "singular" mode of a PE. We design a exceptions handring mechanism matched this mode. We divide exceptions into "analysis" and "freeze". "Analysis" exception doesn't affect other PEs, and a PE caused the exception handles it. "Freeze" exceptions suspends all PEs, and a kernel PE handles it. We design a debugging mechanism used these exceptions as proves and breakpoints for the multiprocessor.

## 1. はじめに

我々は、分散共有メモリ型並列計算機の各プロセッサエレメント（以下、PE）に組み込むための制御機構[1]及び言語Cを拡張した並列記述言語C//[2]を提案している。その制御機構は、並列手続き呼出しの命令とインスタンスの実行管理をハードウェアとして提供する。この制御機構を持つ並列計算機を発注/受領型並列計算機と呼んでいる。本稿では、発注/受領型並列計算機の例外機構と、C//で記述された並列プログラムを発注/受領型並列計算機でのデバッグするための支援機構について報告する。

## 2. 発注/受領型並列計算機の概要

C//では関数呼出しを拡張し、図1のように並列呼出しを記述する。まず演算子『//=』を用いて関数を並列に呼び出し（発注）、『//』によって返り値を受け取る（受領）。この発注と受領の間を『Transfer変数』（図1の t1, t2, t）によって対応付ける。

C//の制御フローは図2に示すように、並列呼出しによって関数（以後テンプレート）のインスタンスを動的に生成し、返り値を受け取る時に同期が取られる。インスタンスは実行終了と同時に消滅する。1つの関数に対してインスタンスは複数生成されてよく、テンプレートを共有するインスタンスは並行実行される。

実際に発注/受領型並列計算機では図3に示すように、各PEは分散された共有メモリに割り当てられたテンプレートだけを実行する。すなわち、PE#1は、

```

void main()
{ int i, j, //t1, //t2, funca(int);
  ...
  t1 // = funca(512);
  t2 // = funca(1024);
  ...
  i = //t1;
  j = //t2;
  ...
}

int funca(int p)          int funcb(int p)
{ int i;                 { int ret;
  int //t, funcb(int);   ...
  ...                    return ret;
  t // = funcb(p+i);     }
  ...
  return i + //t;
}
    
```

図1 C//プログラム例

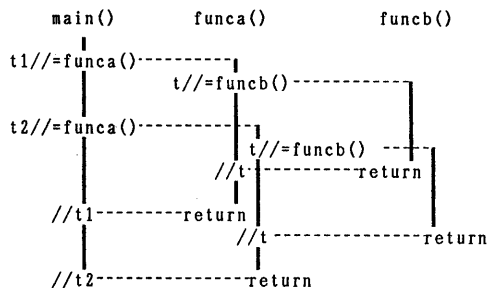


図2 制御フロー

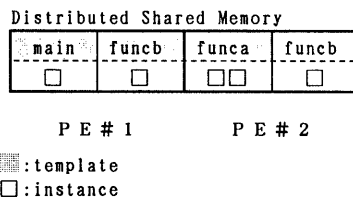


図3 テンプレートとインスタンスの空間配置

main と funcb だけのインスタンスを生成し並行実行する。main のインスタンスが、PE #2 の funca を呼び出すことによって、PE #1 と PE #2 とで図1のプログラムが実際に並列実行される。一つのPE内でのインスタンスの実行管理と並列呼出しは、3. で述べるようにハードウェアとして提供されている。

各PEの通常の実行状態は、インスタンスの実行管理はハードウェア化されているので、複数のインスタンスを並行実行している状態である。この状態は、プロセッサを多重化している状態と言える。一方、例外処理など物理的なハードウェアに依存する処理は、プロセッサを多重化しない状態で逐次的に実行しなければならない。このようにPEは、プロセッサを多重化する状態としない状態を持つ。2つの実行状態を基本とした保護機構については4. で述べる。

C//のプログラムは、main のインスタンスから並列関数呼出しによって拡散的に各PEにインスタンスが生成されゆく。このような実行環境下では、PEで発生する例外やインスタンスのデバッグも並列計算機全体で処理する必要がある。すなわち、例外処理やデバッグも各PEで協調して動作する必要があり、そのため例外機構については5. で述べる。

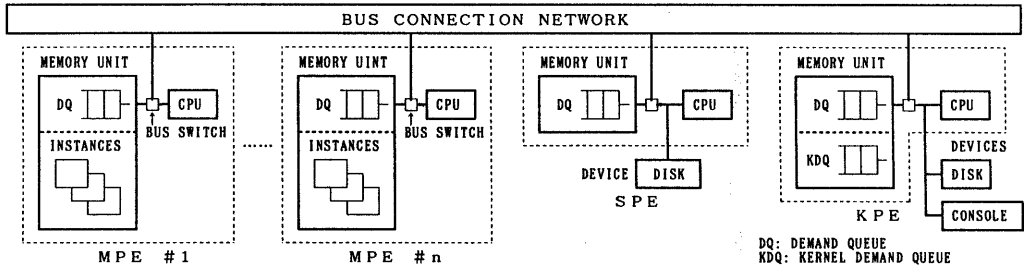


図4 並列計算機の構成

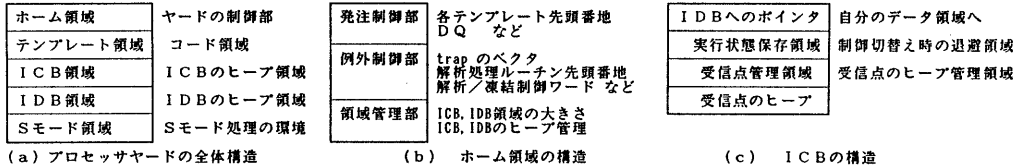


図5 プロセッサヤードの構造

3. 発注/受領型並列計算機の並列リンク機構

図4に示すように発注/受領型並列計算機は分散共有メモリを持つ。この各PEに分散された共有メモリ上のそれぞれに、プロセッサヤードと呼ぶPE固有の制御構造をおく(図5)。この構造はホーム領域、テンプレート領域、ICB (Instance Control Block) 領域、IDB (Instance Data Block) 領域およびSモード領域から構成される。

ホーム領域はPEの制御部であり、例外処理、発注制御、および領域管理用の領域である(図5b)。テンプレート領域は各PEのコード領域である。ICB領域とIDB領域はそれぞれインスタンスの制御部とデータ部のためのヒープ領域である。ICBとIDBは、インスタンスの生成時に環境として動的に割り当てられる。そしてインスタンスとは別に『Sモード』の処理(4.)のための静的な環境を用意しておく。

ホーム領域中の発注制御部にDQ (Demand Queue)を設定する。DQは他PEからの発注に使われ、DQには受信点(Recipient)のアドレスが発注として積まれる。受信点は発注するテンプレートの識別子とそれへの引き数を格納した領域のアドレス、およびそのインスタンスの状態フラグからなる。この受信点を介して発注/受領の同期を行う。受信点はICB内にあるヒープ領域から動的に確保される(図5c)。

この各PEで構築された構造を前提に、2.で述べた制御モデルを基本的な3つの並列リンク命令 DMND、ACPT および DLVR によって実現する。

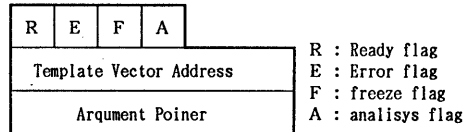


図6 受信点(Recipient)

表1 並列リンク命令

命令	ニモニック	機能
Demand	DMND	発注
Accept	ACPT	発注結果の受領
Deliver	DLVR	結果納入とインスタンス消滅
Select	SLCT	複数の発注の中から受領するものを非決定的に選択
Test & Set with Switch	TASS	失敗の場合は制御を切り替える TAS 命令
Retract	RTCT	発注取り消し

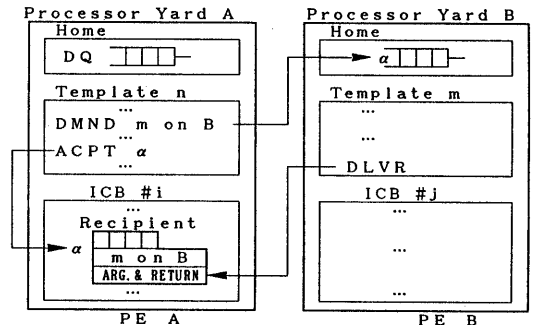


図7 並列リンク

まず DMND 命令で発注する。この命令では受信点を確保して値を初期化し、発注先の DQ に受信点のアドレスを積む。各 PE は DQ に受信点アドレスが積まれると、ICB と IDB 両領域から領域を取得して環境を設定し、受信点に示されたテンプレートの実行を開始する。インスタンスは DLVR 命令によって実行終了する。この命令で引き数領域に結果を返し、DQ から自分に対応する受信点アドレスを取り除き、環境を解放して消滅する。そして発注側で ACPT 命令により受領する。DMND 命令で生成された受信点のフラグを参照し、インスタンスが終了していれば引き数領域から実行結果を読み出し、受信点を解放する。TASS 命令は、単純なテスト&セット命令にインスタンスの制御を切り替える機能を付加したものである。その他の命令は本稿では触れない。

発注/受領型並列計算機では ACPT 命令で発注先のインスタンスがまだ実行中であった時や、TASS 命令でテストに失敗した時など同期待ちになると、インスタンスの制御を自動的に切り替える。DQ から新たに実行可能な発注のインスタンスを生成したり、サスペンドしていたインスタンスに制御を戻す。

以上のように発注/受領型並列計算機では、各 PE に分散して組み込まれた制御機構によって、手続きのインスタンスを動的に生成する並列呼出し型の制御フローを実現している。

#### 4. 発注/受領型並列計算機のモード

##### 4.1 PE の動作モード

発注/受領型計算機ではユーザの手続きに対する制御は多重化される。3. で述べたように PE は通常多重環境を持ち、手続きに対する複数のインスタンスを並行実行している。言い替えればインスタンスに対して個々に論理的な PE が与えられ、実行が行われる。

しかしエラー処理のようにただ一つの物理的な PE に依存したり、または入出力処理など物理的なデバイスに依存しなければならない処理がある。この処理では制御が途中で勝手に切り替えられ、他の関係する制御に追い抜かれては困る。

この論理的に多重な PE の状態と、物理的なイメージに依存した PE の状態の境界は明確に区分すべきである。そこで前者の動作状態を M (Multiplex) モード、

後者を S (Singular) モードと呼ぶことにする。

さらにこの他に並列計算機全体でただ 1 箇所に存在し、系全体を監視するモードが必要になる。これを K (Kernel) モードとする。

発注/受領型並列計算機の通常の PE は、メンバ PE (MPE) と呼ばれ、ユーザのインスタンスを M モードで処理するが、例外の発生などの要因で M モードから S モードに遷移する。

K モードで動作できる PE はハード的に特別な機能 (凍結用割込み元回路等) を持つカーネル PE (KPE) 1 つだけである。この KPE には計算機の立ち上げ用に入出力装置が複数接続される。この管理のため KPE には S モードも必要となる。KPE は通常は S モードでユーザからの入出力要求を処理し、系の中で例外が発生した場合に K モードに遷移する。

また入出力専用 PE や特殊なハードウェアを持つ PE が含まれる場合、それらを専用処理 PE (SPE) と呼び、S モード専用とする。

##### 4.2 特権レベルの設定

4.1 で述べたモードの性格から発注/受領型並列計算機の特権レベルは次のように設定できる。並列計算機の系全体を監視する K モードに対して最上位の特権レベルを与え、入出力などデバイスを扱う S モードは PE 内での特権を与える。そしてユーザの手続きを実行する M モードの順に特権レベルを階層化する。

さらに 1 つの PE に複数のデバイスが接続され、それを扱う複数の S モード環境が存在する場合、それらを管理するために、より上位の S モードが必要になるかもしれない。一般的にはさらに特権レベルの細分化が考えられるが、必要最小限としては K、S、M モードの境界上に設定すればよい。

単純なメモリ保護の例として図 8 のように設定できる。S モードまたは K モードからアクセスできる空間にホーム領域など制御構造おくことで、通常 M モードで動作する他 PE から制御構造を保護する。

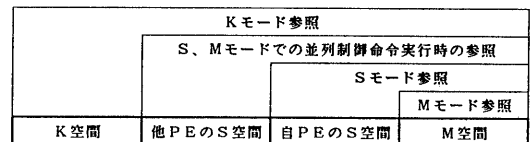


図 8 並列計算機のメモリ保護

#### 4. 3 モード間の交信・遷移

MPEではSモード管理下で複数のMモードインスタンスが多重に並行動作している。そして他PE上のSモード、MモードインスタンスおよびKモードで1つの並列プログラムの処理を実行する。これら1つのプログラムを処理するために構成されたMモードインスタンス間では空間は共有され、SモードとKモードの処理からもアクセス可能でなければならない。

PEがMモードの場合、発注によって動的に環境が割り当てられ、インスタンスが生成される。PEがSモードの場合は、インスタンスを生成せずに自分の環境を用いて処理を行う。Sモードの環境は静的に存在し、DLVR命令を実行しても消滅せず、DQに積まれた次の処理に制御が移る。5. で述べる例外処理の場合もSモードの環境下で処理され、RTE命令で例外処理から復帰するときも環境は消滅しない。

Sモードでは同期待ちが生じて制御を自動的に切り替えない。このことによりSモードのPEに対して発注された処理は、DQに積まれた順に処理が進むことを保証する。見かけ上は、Sモード上の処理は物理的なPEやデバイスに張り付く静的なプロセス(ただし空間は共有する)のように振る舞う。

図9にモード間の交信、遷移図を示す。Mモードインスタンスに対しては呼出しによる動的な生成消滅だけを用意し、Mモードインスタンス間の交信は制御機構としてサポートしない。Sモード、Kモードの処理に対しては、DMND命令による発注動作が疑似的に通信に対応する。MモードからSPEに対しては特定形式のDMNDを行うことができる(②)。Sモードでは同期待ちが起こっても制御が切り替わらず、飢餓状態に陥る危険性も持っている。これを避けるため、自分より上位の特権レベルを持つSモードへしか発注できないようにDMND命令の方向性を限定する。

モード間遷移は例外によって行われる。並列計算機における例外には当該PEだけで処理される『解析例外』と全PEを凍結する『凍結例外』の2種類がある。これらは5. で説明する。

ユーザ手続きのインスタンスをMモードで動作させていたMPEは、各種のエラーによる例外でSモードに遷移する(④)。解析例外ではSモードに入り、ホーム領域のベクタで示されたエラー処理手続きを行う。一方凍結例外では、SモードになったPEが直ちにK

PEに対して暗黙的なDMNDを行う(③)。KPEはこの時SモードからKモードに遷移する(⑤)。

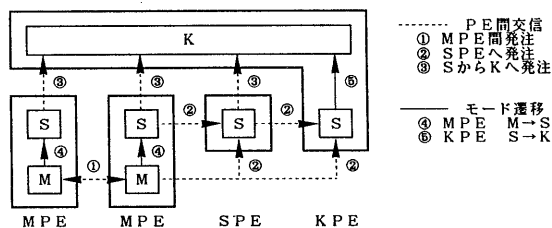


図9 モード間遷移・交信

#### 5. 例外処理機構

##### 5. 1 例外処理機構の概要

並列計算機の例外は、バスエラーやアドレスエラーのような並列計算機全体の例外として処理すべきものと、TRAP命令などのPE単体の例外として処理できるものの2つに分けて考える。前者を凍結例外と呼び、Kモードで動作するKPEを除く全PEを凍結して処理する。後者を解析例外と呼ぶ。これら2つとリセットに相当する立ち上げ処理を含めて、3つの例外処理がある。

解析例外に関しては、各PEが従来と同様の例外処理機構を持てばよいが、凍結例外と立ち上げ処理にはPE間を結び割込み機構が必要になる。

凍結例外が発生した場合、デバッグに用いるためにも、全PEに速やかに例外の発生を伝達する必要がある。このためMPE、SPEおよびKPEの間には、メモリバスとは別の凍結用割込みバスが必要である。MPEまたはSPEのどれか1つ以上が凍結例外を起こすと、オープンコレクタバスを通してKPEに伝達され、凍結フリップフロップ群がすべてセットされる。1つ1つの凍結フリップフロップはそれぞれのMPEまたはSPEに専用線を通して割込みを発生し、KPEを除く全PEを凍結状態にする。一方凍結の解除は、KPEがKモードのプログラムで、個々の凍結フリップフロップを個別に解除することにより行われる。

立ち上げ処理にも一斉に割込みがかかるようにしておく。次節で説明するKPEによる初期化処理の終了時にKPEが一斉にMPE、SPEに立ち上げ割込みをかける。



らは従来のトレース例外に対応する。

TASS 命令を特別にした理由は、共有変数へのアクセス動作だけをブレイクするためである。この機能を用意することで、プログラマが任意の共有変数へのアクセスだけをトレースできる。

また、AOF 照合ビットはブレイクポイント命令である AOF 命令実行時に参照される。AOF 命令の即値オペランドと、AOF 照合ビットの論理和がとられ、もし結果が 0 でなければそれぞれ解析／凍結例外を引き起こす。これから、AOF 照合ビットのそれぞれに対応して、系列の違う並列ブレイクポイントをビット数個分用意できる。

なおこの機構によって解析と凍結が両方発生し得る場合は、凍結例外が優先される。凍結例外が生じた場合、K P E への暗黙デマンドの引き数として、例外の要因すなわち凍結制御ワード中のどのビットで凍結例外が生じたかを示すコードなどを与える。

この解析／凍結制御ワードは従来の計算機の状態語の拡張であり、並列リンク命令の度に参照される。実行中の参照は、P E の立ち上げ時に K P E から与えられた値で初期化される特定レジスタから行うか、ホーム領域においたまま行うかは選択の問題になる。レジスタにのせると実行中ら K P E から制御できず、メモリにおくとアクセスの負荷が大きい。ホーム領域をキャッシュできれば負荷は減少するが、キャッシュの一貫性管理の負荷ができる。またレジスタにのせたままでも割込みを使用して K P E から制御できるが、P E 別の割込みが必要で、P E 数が多いと実現が面倒になる。

解析制御ワードの状態に関わらず解析状態に移行する命令として TRAP 命令を用意する。これは従来のソフトウェア例外に対応し、独自のベクタを参照して解析状態に移行する。これを P E 資源の要求に用いることができる。

また凍結制御ワードの状態に関わらず凍結例外を引き起こす命令として FRZE 命令を用意する。この命令の即値オペランドは K P E への DMND の引き数となる。この命令は並列プログラムの一番最後で、プログラムの終了命令として使用される。すなわちこの命令のオペランドがプログラムの終了コードになる。

このほかに発注先で例外を引き起こす命令として、ADMND、FDMND 命令を用意する。これは発注先がイン

スタンス生成直後にそれぞれ凍結、解析例外を引き起こす。これはデバッグ用に用意された命令だが、この命令を特権命令とするかユーザに解放するのは検討の余地がある。ユーザが使えれば、インスタンス内のユーザ定義の環境初期化に用いることができよう。

解析、凍結例外に移行するタイミングは次表に示すとおりである。

表3 例外状態への移行時期

命令	移行時期
DMND	受信点確保後、発注を D Q に積む前
TASS	命令実行後 ( TASS 成功後)
AOF, FRZE, TRAP	命令実行後
FDMND, ADMND	インスタンス生成直後、発注先で例外その他
	フェッチ直後、実行前

DMND が受信点確保後に例外を起こすのは勝手にユーザが受信点を用意できないため、TASS の場合は履歴をとる都合を考えたためである。

## 7: デバッグ支援機構を用いたデバッグング

並列プログラムのデバッグでは並列プログラムにデバッグ用の処理 (プローブ) を付け加えることで、プログラムの振舞いが変化してしまうプローブ効果が問題となる。無論、プログラムの逐次部分のデバッグには従来の手法も有効である。問題は並列特有の共有資源に対しての競合にかかわるバグで、ブレイクポイントを用いた従来の反復手法ではデバッグできない。並列計算機のデバッグではこれに対処する機構を提供しなければならない。

プローブ効果を避ける 1 つの明快な方法は、ブレイクポイントで同時に全 M P E を凍結し、その間に K P E がデバッグ処理を行うことである。発注／受領型並列計算機は 5. で述べたように凍結例外が生じた瞬間に、厳密には数命令のずれの範囲で全 M P E を停止できる。そのずれに影響されるバグも考えられるが、発注／受領型並列計算機が制御する中粒度のインスタンスに対しては十分小さいと考えることができる。

デバッグ支援用の機構として用意した解析／凍結制御ワード、AOF 命令などを利用して、特定の並列制御命令ごとやプログラム中任意の場所で凍結および解析例外を引き起こすことができる。これから凍結例外を発注／受領型並列計算機用のブレイクポイントとして使

用でき、従来の繰返し式のデバッグ手法も有効に利用できる。また DMND 命令など並列リンク命令で発生させる解析例外を利用し、スタブモジュールで他 P E の動作をシミュレートすれば、P E 単体での逐次部分のデバッグも容易になる。

しかし従来型のデバッグ手法が使えるだけでは不十分である。並列プログラムのデバッグが困難なもう一つの理由が、一般的に逐次プログラム以上に並列プログラムが複雑であることが挙げられる。

これを解決する方法には、実行時にイベントヒストリーを取得して後で解析したり、またはプログラムとは別にその振舞いについて記述し、実行時のイベント発生列とチェックする方法などが挙げられる。

イベントヒストリーを取得する単純な方法は、プログラム自身が必要なイベントを取得するプログラムを書くことである。そのためのライブラリを提供したり、また処理系が自動的にコード中に埋め込んでしまう方法も考えられる。発注／受領型並列計算機では解析例外を利用してソフトウェアプローブを構築し、イベントヒストリーの取得に用いることができる。

ソフトウェアプローブの問題点はやはりプローブ効果を引き起こすことにある。プローブの実行負荷が小さければ、常にプログラムに挿入したままにしておくが、理想的には実行への影響が小さいハードウェアプローブを用意すべきであろう。発注／受領型計算機ではまだ検討中であるが、最低限の機能として共有メモリ中の特定番地へのアクセスが凍結例外を引き起こす機構を考えている。

さらにヒストリーを取得した場合、そのヒストリーをどう処理するかが問題である。このヒストリーは非常に膨大になる。このためデバッガ側は、最低限として膨大なヒストリーを解析し、その中から重要な部分を拾い出す機能は持たなければならない。

発注／受領型並列計算機のデバッガは K P E の K モード上におかれる。そして各 P E の S モードにデバッグ用のプローブやさらに高機能なリモートデバッガをおく構成が考えられる。このデバッガがこれらの機構を用い、並列計算機用のブレークポイントや並列制御構造(3.)の状態表示、イベントヒストリーの取得とそれを解析する機能を構築することができる。

## 8. おわりに

本稿では発注／受領型並列計算機では例外処理機構を構築するために、3つの動作モードが必要であることを述べた。特権レベルはこのモード境界上に設定することで整合する。

そしてこのモデルの上で例外処理機構を構築した。並列計算機の例外としては、他 P E に影響せず例外発生 P E だけで処理できる解析例外と、計算機全体で処理する必要があり、全 P E を凍結する凍結例外に分けて考え、それぞれの処理手続きを述べた。また、ソフトウェア的にこれらの例外を発生させるデバッグ支援機構を構築し、どのように並列計算機のデバッグに応用するかを論じた。

これらの機構は、現在は MC68010上にソフトウェアでエミュレートしている。またこの機構上でのデバッガの実装が今後の課題となる。本稿では触れなかったが、実現に際してはデバッガの表示機構も問題であり、複雑な並列プログラムの状態を理解しやすいように表示するための方法も検討されなければならない。それにかみ、従来のソースレベルよりもっとアブストラクトレベルのデバッガが必要になる。

## 謝辞

本研究について日頃ご指導いただいている本学齋藤延男教授、阿刀田央一教授、五十嵐智助手に感謝する。また発注／受領型並列計算機の実行系の検討を行った星野浩志氏に感謝する。

## 参考文献

- [1] 富澤ほか：手続きフロー型並列計算機における分散型組込み制御機構，信学論(D)，J71-D，10，pp.1921-1930(1988)。
- [2] 富澤ほか：手続きフロー型並列計算機の制御機構と並列記述言語，情報処理学会計算機アーキテクチャ研究会資料，79-15(1989)。
- [3] E.F.Gehring, D.P.Siewiorek, Z.Segall:Parallel Processing, Digital Press, pp.80-84(1987)
- [4] C.E.Mcdowell, D.P.Helmbold:Debugging Concurrent Programs, ACM Computing Surveys, Vol.21, No.4, pp.593-622(1989)。