TM-1289

LSI配線プログラムを用いた並列推論 マシンPIM／cの負荷分散方式の評価

朝家　真知子（日立）、中川　貴之（日立）、
垂井　俊明（日立）、井門　徳安（日立）、
杉江　衛（日立）

ICOT
Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan
(03)3456-3191～5

Institute for New Generation Computer Technology

LSI配線プログラムを用いた並列推論マシンPIM/cの負荷分散方式の評価

概　要

第五世代コンピュータプロジェクトの一環として開発した並列推論マシンPIM/c（256プロセッサ構成）上に、LSI配線プログラムを実装し、システム性能を評価した。PIM/cは、8プロセッサエレメントで構成する密結合クラスタを32クラスタ疎結合する階層構成を採用している。PIM/cのクラスタ構成に適した、プロセッサ間通信量の少ないプロセッサマッピング方法を考案し、960ネットの配線問題において、1プロセッサの実行に比べて256プロセッサで94.8倍の台数効果を観測することができた。また、LSI配線プログラムに動的負荷分散支援ハードウエアを活用し、ソフトウエアで動的負荷分散制御を行った場合に比べて、1.8倍に高速化した。

1．　はじめに

（財）新世代コンピュータ技術開発機構（ICOT）では、通産省の第五世代コンピュータプロジェクト（昭和５７年度～平成４年度）の一環として、大規模知識処理を目指した、並列推論マシンの研究開発を進めてきた[1]。我々は、256プロセッサで構成する並列推論マシンPIM/c (Parallel Inference Machine model c)[2]を開発した。PIM/cは、8プロセッサエレメントでメモリを共有するクラスタを32台接続し、クラスタ間はメモリを分散する階層構造を採用している。

一方、並列推論マシンが対象とする問題は、推論を行なう大規模な知識処理プログラムである。我々は、ICOTで開発された並列知識処理応用プログラム（ICOT無償公開ソフトウェア[3]）の中から、並列化版「LSI配線プログラム[4]」を選び、PIM/c上に実装してシステム評価した。「LSI配線プログラム」は、階層構造を持たないフラットな分散メモリ型並列計算機上で開発されたプログラムであり、オブジェクト指向プログラミング手法を用いて並列化を行なっている。並列計算機を効率良く実行させるには、仕事（負荷）を各プロセッサに均等に配り、プロセッサ間の通信を極力削減することが必須であるが、階層構造のPIM/cにプログラム移植するにあたっては、クラスタ内の通信の局所性を生かし、さらに通信量を削減するための静的負荷割付の検討が必要となった。

また、推論過程では、処理の過程で動的に次に実行すべき処理が選ばれる。そのため、あらかじめ負荷量を推測し、固定的な負荷割当を行うことは困難である。動的に負荷量が変化する場合、各プロセッサの負荷量を判断して均等に負荷割当を行うことをソフトウエアで制御することは可能であるが、ソフトウエアが各プロセッサに負荷量を問い合わせるため、処理に時間がかかる。この課題に対応するため、PIM/cでは、負荷量を動的に判断して負荷割当てを行なうためのハードウエア支援機構[5]を搭載している。本稿では、その負荷分散支援ハードウエア支援機構の実用性を検証した。

本稿では、我々の開発した並列推論マシンPIM/c上で、並列化版「LSI配線プログラム」を動作させて、次の項目を評価する。

（１）　クラスタ構成向け静的負荷割り付け方式
（２）　PIM/cの動的負荷分散ハードウエア支援機構

2．　PIM/cの構成

図１に、PIM/cのプロセッサ構成を示す。プロセッサエレメント(PE)8台がスヌーピングキャッシュを介して主記憶を共有し、1クラスタ(CL)を構成する。複数のプロセッサ間の通信のために、クラスタに１つのクラスタコントローラ(CC)を設け、クラスタ内のプロセッサを代表してクラスタ間通信処理を担当させる。クラスタコントローラをクラスタに１つずつ設けることにより、プロセッサがそれぞれ独立に通信路を設けてクラスタ外のプロセッサと直接交信する場合に比べて、プロセッサ間ネットワークのハード量を１／８にしている。PIM/cシステムは、クラスタ32台をクロスバネットワークで接続して、全256プロセッサを構成している。

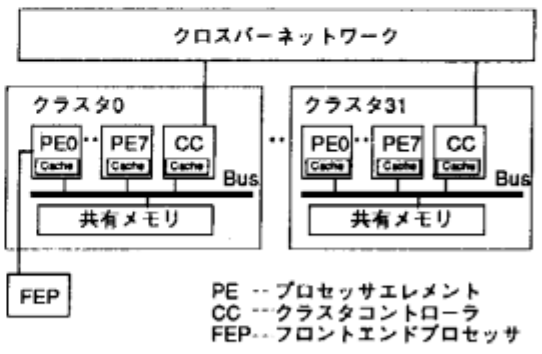

図１．PIM/c のプロセッサ構成

プログラム実行時の負荷分散方法は、クラスタ内の主記憶共有プロセッサにおいては言語処理系による自動負荷分散、クラスタ間においてはユーザがプログラム中に負荷分散先を指定する述語を記述する明示的な負荷分散である。さらに、PIM/cではクラスタ間における動的負荷分散を効率良く実現するために、クラスタ間負荷分散のための通信を高速化するハードウエア機構を備えている。クラスタ間の処理では、共有メモリがないので、大域的な変数であるクラスタ負荷値を参照する場合には他クラスタにメッセージを発行して負荷値を問い合わせ、結果のメッセージを待つ。通常メッセージはクラスタとネットワークの間で一旦バッファリングされ、それ以前のメッセージが全て処理されるまで待ち合わせる必要がある。PIM/cでは、負荷値通信の際のこのようなバッファの待時間を削減するために、緊急メッセージ専用の短絡路を設けた。さらに、クラスタの負荷値をネットワーク中の簡易レジスタに格納することで、クラスタ間の負荷値報告の通信処理を高速化している。プログラム中の記述方法は、負荷分散指定箇所に1語の述語を追加するだけで、処理系が自動的に負荷分散先クラスタを選択することができる仕様とした。負荷分散先クラスタはスマートランダム方式[1][6]によって選択する。

3．LSI配線プログラムとPIM/c上の負荷分散方式

3．1　LSI配線プログラム概要

ICOTで開発されたLSI配線プログラムは、配置図上のLSI部品の端子間を結ぶプログラム（ソース量KL1言語5000行）である。縦方向・横方向の格子で配線層を使い分ける2層配線を扱う。回路情報として、格子サイズ・配線禁止領域・スルーホール禁止点・ネットリストの情報を与えて、配線を計算する。

このプログラムの並列化手法は、並列オブジェクトモデルに基づく。配線格子上における全ての配線線分をオブジェクト＝プロセスに対応させ、これをラインプロセスと呼ぶ。格子の端から端までの線分をマスタラインプロセスと呼ぶ。マスタラインプロセスは、それぞれ直行するラインプロセスとメッセージを交換して、並列に配線線分を探索して行く。このマスタラインプロセス単位に実行ノード（クラスタ）に配置し、並列計算機にマッピングしている。

3．2　静的負荷割り付けによる性能チューニング

LSI配線プログラムについて、PIM/c上の負荷の静的ノード割当方法変更を検討して性能向上を図った。

ICOT開発のオリジナルLSI配線プログラムは、階層構造を持たないフラットな分散メモリ型並列計算機上で開発され、各ノードへの負荷の均等化を重点課題として設計された。負荷均等化を実現するため、配線図の縦横の格子（グリッド）上のプロセスを順番に（サイクリックに）ノードに割り当てている。これを負荷均等化マッピングと呼ぶ（図2.プロセスの負荷均等化マッピング方法）。このプログラムをPIM/cに実装すると、各プロセッサの稼働率は均等化するが、図2中の(1)から(2)を結ぶ短い配線でも、複数のノード（クラスタ）の間に配線処理が生じる。

PIM/cのプロセッサ構成を考えると、クラスタ内はメモリを共有しているのでクラスタ内処理は高速に扱うことができ、

クラスタ内プロセッサ間処理時間　＜　クラスタ間処理時間

の式が成り立つ。一般に、LSI配線の問題を扱うとき、配線しなければならない端子間の距離が遠く離れていることは考えにくい。従って、配線データの局所性を活かしたプロセス割当を行い、クラスタ内プロセッサ間処理を増やすことによって、プログラムの実行高速化を図ることが出来る。

そこで、クラスタ内プロセッサ間処理を増やし、ネットワーク経由のクラスタ間通信を減らすために、図3.のようにプロセスの通信局所化マッピングによるプロセス割当て方式を検討した。本方式では、縦・横の格子数、すなわちマスタラインプロセス数を実行可能な最大ノード数に分割して、複数のマスタラインプロセスの単位で連続的に割り当てる。その結果、図の例で示す配線では始点(1)と終点(2)が同じクラスタCL0内に存在する。

さらに、通信を局所化するマッピング方式を採用することによって、クラスタ間に発生した無駄な配線処理を削減する効果もある。並列処理の過程では、必ずしも最短経路のみを探索するわけではないため、配線領域が空いていれば、複数のプロセスが探索を行う。複数のラインプロセスが配線終了要求メッセー

[1] スマートランダム方式：
負荷を分配しようとするクラスタが、ランダムに選んだ他クラスタの負荷値と値比較を行い、相手側の負荷値が小さい場合にのみ、負荷を分配する方法。

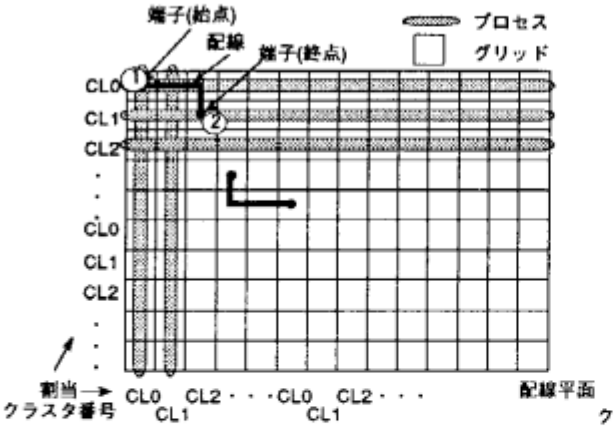

図2.プロセスの負荷均等化マッピング方法

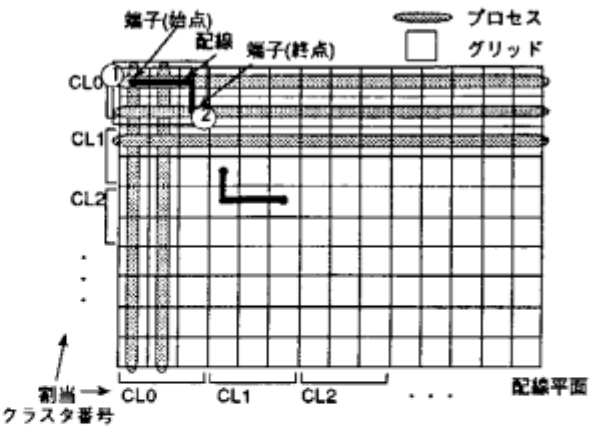

図3.プロセスの通信局所化マッピング方法

ジを1つのラインプロセスに返信すると、最初に配線終了要求メッセージを受理されたラインプロセスだけが選ばれ、他のプロセスはキャンセルされる。したがって、クラスタ内とクラスタ間の通信レイテンシ差によって、クラスタ内のプロセス間で配線処理を終了することができ、クラスタ間に関与してしまうような配線処理はキャンセルできる。

3．3　動的自動負荷分散支援ハードウエアを用いた負荷分散方式

上述のように、「静的な負荷分散（マッピング方法）の変更」によって性能改善を図ることができたが、動的な負荷分散の観点からも、PIM/cの性能評価を試みた。プログラム実行中に動的負荷分配をする場合には、負荷分配をする時点で各プロセッサの負荷値を明らかにする必要がある。以下の条件で、ソフトウェアで動的負荷分散制御を行なった場合と、PIM/cの独自機構である動的負荷分散ハード支援機構を使った負荷分散方式について、LSI配線プログラムに応用して実験および性能比較した。

（1）ソフトウェア制御による動的負荷分散

［負荷分散先選定方法］負荷分散元クラスタが全クラスタに問い合わせ通信を行ない、処理待ちプロセスを持たないクラスタに対して、負荷分散する。

（2）ハードウェア制御による動的負荷分散

［負荷分散先選定方法］処理待ちプロセス数を負荷値と定義し、負荷値レジスタ（動的負荷分散ハード支援機構）の値をクラスタ間で比較、スマートランダム方式で選択したクラスタに負荷分散する。

4．　評価結果と考察

4．1　静的負荷割り付け方式

「3．2　PIM/c上での性能チューニング」で述べた方法に従って、

（1）負荷均等化マッピングの場合（図2）

（2）通信局所化マッピングの場合（図3）

について、性能を計測した。

評価に使用したデータは配線数960NET（グリッド数640×320）で、プロセッサ台数1,8,16,32,64,128,256台について、実行時間を計測した。いずれも、ほぼ97%程度以上の配線率を得た。台数効果の結果を図4に示す。256PE構成の1PEに対する相対性能は、負荷均等化マッピングの場合に82.3倍、通信局所化マッピングの場合に94.8倍である。クラスタ内における言語処理系の自動負荷分散制御を考慮しても、良好な値と言える。また、256PE構成の実行時の（1）と（2）の方式を比較した結果、実行時間にして95秒、および81秒であり、プロセスの割り付けを変更した結果、約1.2倍に性能を向上させることに成功した。

各クラスタ内の処理内容は、ほとんどが配線を探索している部分で、100%近いプロセッサ稼働率が観察できた。その間のクラスタコントローラの稼働率は、50%程度である。クラスタコントローラの通信処理が全体のボトルネックにならずに、クラスタ内の処理が十分行なえており、並列処理の理想的な稼働状況を示している。

クラスタ構造の計算機にプログラムチューニングするときは、クラスタ間の通信を極力抑え、クラスタ

内の処理を増やすことによって、プログラム実行効率が上がることがわかる。クラスタ構造の計算機の実行効率は、データの局所性に依存するので、クラスタへのデータ割り付けを慎重に行う必要がある。

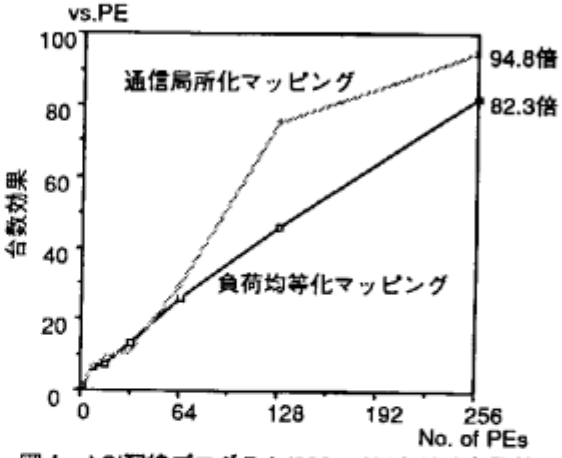

図4　LSI配線プログラム(960net)における台数効果

4．2　動的自動負荷分散支援ハードウエア効果

動的負荷分散支援ハードウエアを使用して8クラスタ（64プロセッサ）で実行した結果は、配線数136NET規模の場合、実行時間が32秒で、ハードウエア機構を使用せずにソフトウェア負荷分散制御によって同条件で実行した58秒と比較して、約1.8倍に性能が向上した。

また、ハードウエア支援動的負荷分散制御では、負荷分散先を判断・指定するためのソフトウェア制御文（約50行）を、動的負荷分散を指定するための述語1語に置き換えることができ、複雑なプログラミングが不要である。

クラスタ間の負荷が不均一な問題に対して、簡易なプログラミングで、より高性能な動的負荷分散が実現できることを実証した。

5．　まとめ

プロセッサ8台で構成するクラスタを32クラスタ接続した、合計256プロセッサ構成の階層型並列推論マシンPIM/cを開発した。PIM/cには、動的な負荷分散処理支援のために、各クラスタの負荷値を高速に判断する機構を設けた。我々は、PIM/c上に、ICOTの開発した並列化版LSI配線プログラムを実装して、クラスタ構成、および、動的負荷分散支援ハードウエアの効果を評価し、以下の結論を得た。

（1）　階層型並列計算機指向のプロセス割当て方法を導いた。

LSI配線プログラム移植に際して、クラスタ内プロセッサ間通信時間とクラスタ間通信時間のレーテンシ差を考慮し、クラスタ内の局所性を生かしたプロセスのマッピング方法に変更した。本評価では、クラスタに割り当てる格子の単位を大きくすることによって、クラスタ内の局所性を保つことを可能とした。その結果、960ネットの配線問題において、プロセス割当方法変更前に比べて、1.2倍に高速化した。さらに、256PE構成の実行で94.8倍の台数効果をあげることができた。

（2）　動的負荷分散ハード支援機構の効果をアプリケーションプログラムで実証した。

LSI配線プログラムにおいて、動的に負荷が変化する部分に、PIM/c独自機構である動的負荷分散ハード支援機構を用いて動的負荷分散を行った。その結果、ソフトウエア制御による動的負荷割当処理に比べて、1.8倍に高速化した。また、ハード支援機構を用いた動的負荷分散方式は、負荷分散のための制御文が不要である。以上により、応用プログラムにおける動的負荷分散ハード支援機構の有効性を実証した。

なお、本研究はICOTの再委託研究として行なわれた。

参考文献

[1]　（財）新世代コンピュータ技術開発機構：第五世代コンピュータ国際会議（第五世代コンピュータの研究開発成果）、1992.6

[2]　後藤厚宏、瀧和男、中川貴之、杉江衛：「並列推論マシンPIM/c」、情報処理学会第４０回全国大会講演論文集(III), pp.1177-1178

[3]　（財）新世代コンピュータ技術開発機構：「無償公開プログラムリスト」,1992.6

[4]　伊達博、大嶽能久、瀧和男：「並列オブジェクトモデルに基づくLSI配線プログラム」、情報処理学会論文誌、Vol.33,No.3,pp.378-385(Mar.1992)

[5]　井門徳安、前田浩光、垂井俊明、中川貴之、杉江衛：「並列推論マシンPIM/c－負荷分散支援機構－」情報処理学会第４０回全国大会講演論文集(III), pp.1181-1182

[6]　M.Sugie,M.Yoneyama,N.Ido and T.Tarui：" LOAD-DISPATCHING STRATEGY ON PARALLEL INFERENCE MACHINES ", Proc. of FGCS'88 Vol.3, pp.987-993