

## P I E E E の

3C-10

## アクセス衝突を低減させるデータ配置機構

小池汎平, 田中英彦

(東大 工学部)

1. はじめに

現在、我々は、複数台の推論ユニットを自動負荷分散ネットワークで結合した、高並列推論エンジンP I E E の並列処理向け実験環境P I E E E (Parallel Inference Engine Experimental Environment) [1~3]の開発を進めている。P I E E E は、ハードウェアによる並列処理の評価をはじめ、これまでに提案してきた種々の高速化技法の実験、大規模並列プログラムの特性解析、実用化システムに向けての検討など、広い範囲の実験に用いることを目的として設計されている。P I E E E は、複数の推論ユニットが、負荷分散ネットワークと共有メモリネットワークの2種類のネットワークで結合されている。P I E E E の個々の推論ユニットは、68020 をメインプロセッサとして持ち、推論用コプロセッサと通信用DMA コプロセッサにより、推論処理、並列処理の支援がなされる。現在、P I E E E 上のシステムソフトウェアF L O N G の開発も進められている[4]。

P I E E E を用いた実験項目のうち基本的なものとして、コミットド・チョイス言語F L E N G [5]の効率的な実装方法を検討することが挙げられる。そこで今回は、F L E N G の実装するにあたって重要な共有データのメモリ上への配置法について検討する。

2. P I E E E のハードウェア構成

P I E E E は、4台の推論ユニットを負荷分散用と共有メモリアクセス用の2種類のネットワークで結合したマルチマイクロプロセッサシステムである。基本処理要素である推論ユニットは、ゴールの書き換え等の基本処理を行なう。推論ユニットは、68020 プロセッサ、ローカルメモリ、VECTOR-UNIRED、DMA コントローラ、の4つの要素から成っている。VECTOR-UNIRED とDMA コントローラは、68020 のコプロセッサとしてP I E E E の推論処理と並列処理を支援する。VECTOR-UNIRED は、ゴールフレームの単一化と縮退をパイプライン処理する。DMA コントローラは、負荷分散用と分散メモリアクセス用の2系統が用意されている。後者の基本機能は、他の推論ユニットのローカルメモリと自分のローカルメモリの間で、任意の大きさのデータを転送することである。この機能により、各推論ユニットは分散型の共有メモリを

持つことになる。

負荷分散ネットワークを通して推論ユニット間でゴールフレーム(GF)のやりとりが行なわれる。負荷分散ネットワークには、負荷情報をフィードバックし、接続可能なプロセッサのうちで、負荷最少のものへの経路を自動的に設定する、自動負荷分散機能が付いたスイッチングユニットを用いている[6]。共有メモリネットワークは、各推論ユニットが持つ分散共有メモリ間でデータのやりとりを行なうためのものである。共有メモリ用ネットワークも負荷分散ネットワークと同様のスイッチングユニットを用いており、推論ユニットの設定した値により自動的に経路を設定するモードで動作可能となっている。

3. P I E E E 上のF L E N G 実行モデル

ゴール書き換えモデルでF L E N G を実行する場合、ゴールフレームは、処理単位である1つのリテラルと、そのリテラルから指される構造データとからなり、このゴールフレームを書き換え、新しいゴールフレームを生成することを繰り返すことが、推論ユニットでの基本的な処理となる。

推論ユニットにおけるF L E N G 実行の基本的なサイクルは、以下のようなものである。

スケジューリング:

ゴールメモリからゴールフレームをひとつ取り出す。

単一化:

定義節とGFのヘッド部の単一化を行なう。

縮退:

単一化に成功したひとつの定義節のボディ部の各リテラルについて、縮退処理を施し、新たなゴールフレームを作成する。そして、縮退の結果生成されたゴールフレームを適切なプロセッサに負荷分散させると同時に、複数のリテラル間で共有関係が生じたデータ(共有変数セル、共有構造データ)を分散共有メモリ上に配置する。

システム術語の実行:

ボディ部にシステム術語があった場合は、縮退処理を施す代わりにこれを実行する。

4. 負荷分散と共有データ配置の戦略

一般に、P I E E E 上で共有データを並列処理する場

合、負荷分散と分散メモリ上への共有データ配置に関して、望ましい場合は、

- ①処理負荷が各プロセッサに均等に分散されていること、
- ②各プロセッサが分散共有メモリ上の共有データに衝突なく並列にアクセスできる（可能性が高い）こと。
- ③なるべく負荷分散、共有データの配置とも自推論ユニットに対して行なえること。

である。

①は、プロセッサの処理能力を最大限有効に利用するため、②は、共有メモリネットワークの能力を最大限有効に利用するため、③は、負荷分散/共有データ配置時のオーバヘッドと共有データアクセスのオーバヘッドを最小化するために重要である。この様子を図1に示す。しかし、一般に、これらの条件は相反する関係にあり、トレードオフにより最適な状況を作り上げるような、各プロセッサへの負荷分散及び各分散メモリ上への共有デ

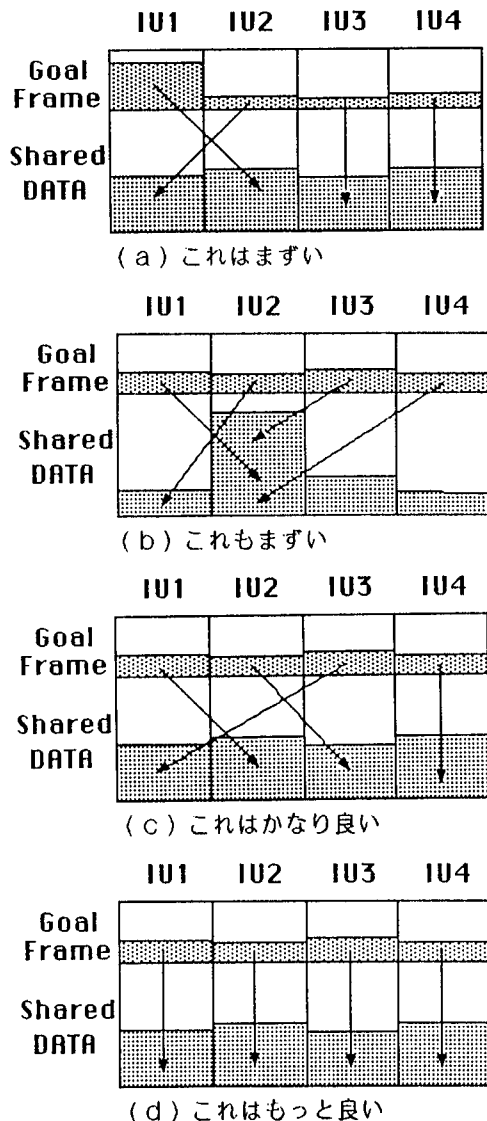


図1 負荷分散と共有データ配置

ータ配置に関する戦略が必要となる。

PIEEEで負荷分散と共有データの転送に用いる2種類のネットワークは、共に負荷分散機構を持っている。このため、得られる動的情報としては、

- ①負荷分散ネットワークを通して現在接続可能なプロセッサの出力する負荷情報（例えば、実行可能なゴールフレームのサイズの合計）
- ②共有メモリネットワークを通して現在接続可能な分散メモリの出力する情報（例えば、個々の共有メモリ内の全共有データを外部から指すリファレンス数の合計）
- ③自プロセッサの各種情報

が利用できる。例えば、①と③の情報より、他のどのプロセッサの負荷も自プロセッサの負荷よりも高いことがわかった場合は、縮退によって生成されたゴールフレームは自プロセッサに分散し、共有データも自プロセッサに配置し、各プロセッサで独立に処理を行なわせるのが最良であろう。しかし、逆に、処理の立ち上がりなど負荷が不均一な状況では、ゴールフレームを積極的に分散し、これに伴ない最適な共有データ配置を行なうのが、処理の分散化、共有データアクセスの分散化の点で有利であろう。これらの議論をもとに、上記の状況を実現するためのどのような戦略が最適か、及び、更に良い戦略を行なうためにはどのような機能の追加が必要かの検討を行なう必要がある。

## 5. おわりに

PIEEE上にFLENGを実装する際に重要な、負荷分散に伴ない共有データ配置について、PIEEEの持つ負荷分散機構を利用した方法を検討した。PIEEEの稼動を待って実験を行なう予定である。

### <参考文献>

- [1] 小池, 山内, 野田, 田中, “高並列推論エンジン実験環境PIEEE—全体構成—”, 第34回情処全大, 4P-3, 1987.
- [2] 野田, 小池, 山内, 田中, “高並列推論エンジン実験環境PIEEE—推論ユニット—”, 第34回情処全大, 4P-4, 1987.
- [3] 山内, 小池, 野田, 田中, “高並列推論エンジン実験環境PIEEE—自動負荷分散ネットワーク—”, 第34回情処全大, 4P-5, 1987.
- [4] Nilsson M. and Tanaka H., “Flong - A Portable Software System for the Inference Machine PIEEE”.
- [5] Nilsson M. and Tanaka H., “FLENG Prolog - Turning supercomputers into Prolog machines”, Proc. Logic Prog. Conf. '86, Tokyo, June 1986.
- [6] 坂井, 小池, 田中, 元岡, “動的負荷分散を行う相互結合網の構成”, 情処論, Vol. 27, No. 5, 1986.