

ベクトル並列スーパーコンピュータ VPP500 の並列処理*

6K-2

石坂 賢一†
富士通株式会社

1 はじめに

VPP500 は、それぞれが高性能なベクトルプロセッサである要素計算機 (PE) を高速クロスバネットワークで結合した並列計算機である。VPP500 は、最大 222 台の PE を接続することにより 355GFLOPS の演算性能と 222GB の主記憶容量とを実現し、超大規模な科学技術計算を高速に実行することを可能とした。

本稿では、VPP500 の概要について述べた後、クロスバネットワークなど VPP500 で採用した並列処理アーキテクチャの特徴について説明し、最後に並列処理性能について触れる。

2 VPP500 の概要

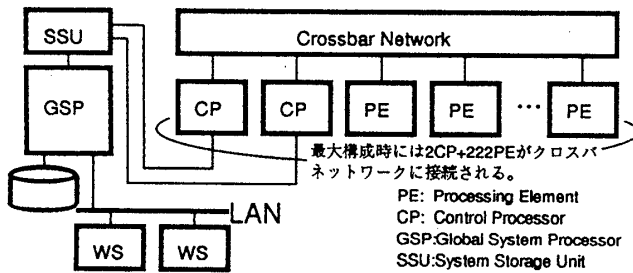


図 1: VPP500 システム構成

VPP500 は、図 1 に示すように、複数の PE をクロスバネットワークで結合した並列コンピュータシステムである。

各 PE は、LIW 型 RISC スカラ演算ユニット [1]、1.6GFLOPS のベクトル演算ユニット [2]、最大容量 1GB の主記憶、および、演算ユニットの動作とは独立に PE 間のデータ転送を実行するデータ転送ユニットを 1 ボードに実装したベクトルプロセッサである。

VPP500 は、UNIX SVR4 をベースとして分散メモリ型高並列ベクトル計算機のための機能を拡張/強化したオペレーティングシステム [3] によって制御される。また、ベクトル処理と並列処理とを実現することのできる FORTRAN 言語処理系 [3] を中心としたプログラミング環境を提供している。

3 VPP500 の並列処理アーキテクチャ

並列計算機において高い並列処理効率を得るための重要なポイントは、PE 単体の性能を出来る限り高くすることで PE 台数を比較的少なくすること、および、転送性能の強力な結合網を使用して PE 間通信によるオーバーヘッドを低減することの 2 つである。

多くの分野のプログラムに適応しようとする、トラスのような隣接結合やオメガネットのような多段網では複数の PE 間通信が干渉して性能が低下するため、完全結合やクロスバ網のような結合網とする必要がある。

また、同時に走行する多数の並列プログラムをそれぞれ自由に PE に割り付けても高い性能が安定して得られるような、柔軟な運用が可能な並列処理アーキテクチャでなくてはならない。

このような観点から VPP500 で採用した並列処理機構について以下に述べる。

3.1 クロスバネットワーク

クロスバネットワークではどの PE 間の通信でも等距離で他の通信の影響を受けずに同時実行可能であるため、PE 間の通信パターンによらず広い分野のプログラムを効率よく実行することができる。また、クロスバネットワークの自己同型¹という性質により、PE のパーティションにかかわらず常に高いネットワーク性能が得られるため、複数の並列プログラムを高効率で運用することが可能である。

VPP500 のクロスバネットワークは、任意の PE 間で 400MB/s のスループットのデータ転送性能をもつ。PE と CP を合計 224 台接続する最大構成のクロスバネットワーク全体では 89.6GB/s の巨大なバンド幅となる。また、モジュラリティの高い実装方法によって、最大構成時においても図 2 に示すように基本クロスバボード (XB1 および XB2) の組合せで完全クロスバを実現している。

3.2 データ転送ユニット

高い並列処理効率を実現するために、演算とは独立非同期に通信を実行するデータ転送ユニット (DTU) が各 PE に備えられている。DTU は、ソフトウェアによる通信オーバーヘッドを削減するための高機能通信プロセッサで、以下に示すような特徴をもつ。

* Parallel Processing Architecture of VPP500
† Ken'ichi ISHIZAKA, Fujitsu Limited

¹ 任意の PE の組合せ (パーティション) を選択しても PE 間のトポロジはまたクロスバであること

- ・ 送受各 400MB/s の通信を同時実行
- ・ OS を介せずにユーザモードでの転送指示が可能
- ・ 転送指示のキューイング
- ・ PE 間 WRITE/READ アクセス
- ・ 多様なメモリアクセスパターン
連続/ストライド/サブアレイ/間接アドレス
- ・ アドレス変換機構
PE 番号およびメモリアドレスの仮想化

3.3 PE 間同期機構

複数の PE が協調してプログラムを実行するために、VPP500 ではバリア同期を実行するハードウェア機構を持つ。これは図 3 に示すように、各 PE のプログラム進行状態情報を全ての PE に放送する機構と PE 内でこの情報を受信して同期を検出する機構とからなる。PE 内に同期をとる PE グループを示すマスクレジスタを持つことによって、プログラムを任意の PE からなるパーティションで走行することを可能としている。

4 性能評価

VPP500 の性能の例として、LINPACK ベンチマークプログラム (次元数は 5000 元固定とした) の実行性能を測定した。PE 台数を 2 台から 32 台まで増加させた場合の実行性能測定結果を図 4 に示す。このグラフより、多数の PE を使用したシステムにおいても高い並列化効率を得られていることがわかる。

5 おわりに

VPP500 は、高い PE 演算性能と強力な並列処理アーキテクチャの採用により、高い並列処理性能を実現した。また、クロスバネットワークの採用により、柔軟で効率のよい運用を可能とした。

今後は、並列化オーバーヘッドをさらに低減するための PE 間グローバル演算機構などのハードウェアを開発するとともに、全てのユーザが容易に並列計算機の高い性能を引き出せるような並列化コンパイラを開発してゆく所存である。

参考文献

- [1] 中島, 北村, 田村, 滝内: VPP500 スカラプロセッサの特徴, 情報処理学会計算機アーキテクチャ研究会報告, 104-17, pp.129-136, Jan. 1994.
- [2] 持山: ベクトル並列スーパーコンピュータ VPP500 のベクトル処理, 情報処理学会第 49 回全国大会, 6K-01, Sep. 1994.
- [3] Kenichi Miura, Moriyuki Takamura, Yoshinori Sakamoto, and Shin Okada, "Overview of the Fujitsu VPP500 Supercomputer", *Digest of Papers, COMPCON Spring 93*, San Francisco, Feb. 1993, pp. 128-130.

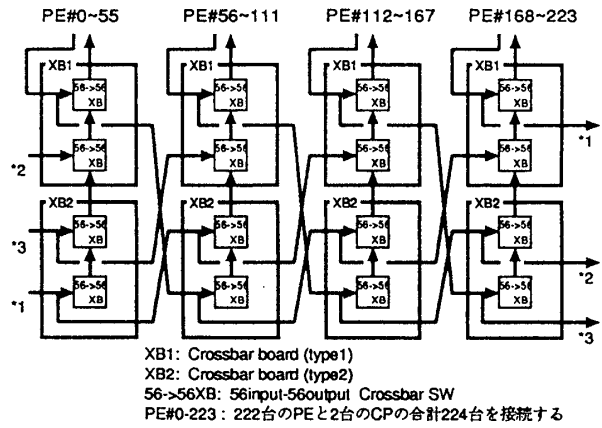


図 2: 最大構成時のクロスバネットワーク

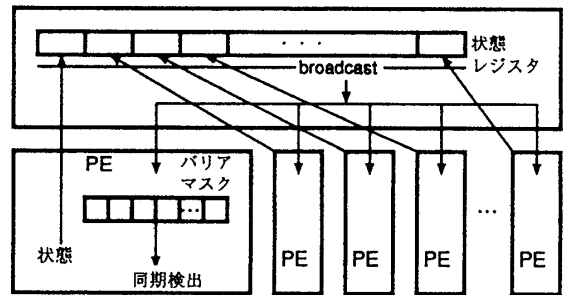


図 3: バリア同期機構

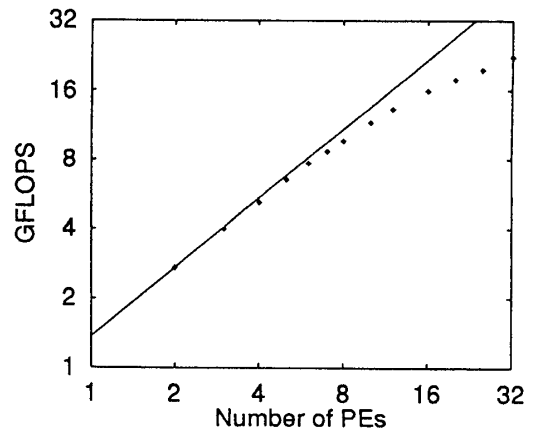


図 4: VPP500 の性能 (LINPACK5000 元)