# 富士通 高並列スーパーコンピュータ VPP の開発思想

高村　守幸（富士通株式会社）
*4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-88, JAPAN*
*email : takamura@ayame.mfd.cs.fujitsu.co.jp*

## 1　はじめに

ある設計目標を達成するためのコンピュータの設計においては、幅広い範囲のテクノロジー、アーキテクチャ、及びインプリメンテーションが吟味選択される。これ故にコンピュータ設計は多種多様であり、「ユニーク・ソリューション」ではない。よって、良いコンピュータの構成とは何なのかはだれにも答えられない。コンピュータシステムは、半導体、実装、アーキテクチャ、システムソフトウエア、各種アプリケーション、及び運用機能から構成されていることは明白である。これらの六つの要素がそれぞれ持っている特性と先端性、及びその組み合わせがコンピュータの「フレーバ」を決定する。

これを踏まえ、大規模科学計算分野の高速処理において富士通が考える最適なアプローチとは何であるかを以下に述べ、VPP シリーズの半導体テクノロジーとアーキテクチャ、性能測定結果、及び今後の展望について述べる。

## 2　半導体テクノロジー

ここ 20 年の間、スーパーコンピュータの高性能は、高速な半導体素子及びベクトル処理に依存してきた。高性能の追求は、クロックサイクル時間の高速化のために、ECL 素子の独占的使用を決定づけた。しかし、スーパースカラ方式のマイクロプロセッサの高速化のモーメンタムにより、CMOS テクノロジーは急速な進歩を遂げた。CMOS テクノロジーを用いたスーパーコンピュータは、ECL を用いた従来機に対して単体性能で劣るが、コスト／パフォーマンス、最大システム性能、筐体サイズ、及び消費電力においては大きな優位性を発揮することが可能となった。現状の微細加工技術ではゲート長が 0.35 μm まで実用化され、実験室レベルでは既に 0.07 μm 技術が可能なことが確認されている。

これから 10～15 年間の CMOS テクノロジーの進歩は殆ど約束されているといっても過言ではない。1995、96 年 に富士通は 0.35 μm CMOS-LSI を全面的に使用したスーパーコンピュータ VX/VPP300/VPP700 を出荷した。少なくともこれから 10 年間は最先端 CMOS テクノロジーが富士通のスーパーコンピュータを実現する半導体素子であり続けると予測する。

## 3　アーキテクチャ

### *3.1　キャッシュ対ベクトルレジスタ*

計算機の性能を実質的に決定する要因は主記憶性能にある。ベクトルプロセッサは主記憶を高速にアクセスする手段を持っている点で高性能マイクロプロセッサを差別化している。

マイクロプロセッサの性能は、キャッシュに依存する。キャッシュはデータがメモリ空間に局所的に存在するときは高速アクセス性を発揮するが、大規模科学計算にみられるような局所性がない広域データアクセスにおいては、マイクロプロセッサはピーク性能に対して僅かな性能(10%程度)しか発揮出来ない。

ベクトルプロセッサでは、プログラマブルな大容量ベクトルレジスタを駆使して、メモリからロードを先行して実行するなどの高度なスケジューリング技術が確立されている。これによりスーパーコンピュータで頻発する局所性を利用できない広域データの処理においても高い実行性能を引き出すことが可能である。

図1に、ベクトル処理とスカラ処理の例として富士通VPP500の単一ノード及びIBM RS6000で図示のループを実行した実効性能(効率)の測定結果を示す。図1より、スカラ処理はベクトル処理に比べて堅固性がないことが判る。半導体技術が進歩してキャッシュが大きくなろうが性能が急激に低下するポイントが多少右に移動するだけで本質的な解決にはならない。これに対し、VPP500ベクトルユニットではデータ量が増大するにつれて限りなくピーク性能に近づく。

表1にアプリケーションコードから抜粋した典型的なDOループを、最新5種のマイクロプロセッサを用いたワークステーションとVPP500ベクトルユニットで動作させた際の性能を示す。VPP500ベクトルユニットはマイクロプロセッサと比較して、効率で5倍以上、絶対性能で20倍強の高性能が発揮できることが検証された。マイクロプロセッサは、4年で3倍の性能向上を遂げている。ベクトルプロセッサの20倍強の性能アドバンテージは、マイクロプロセッサの10年先の性能を今日実現していると言える。

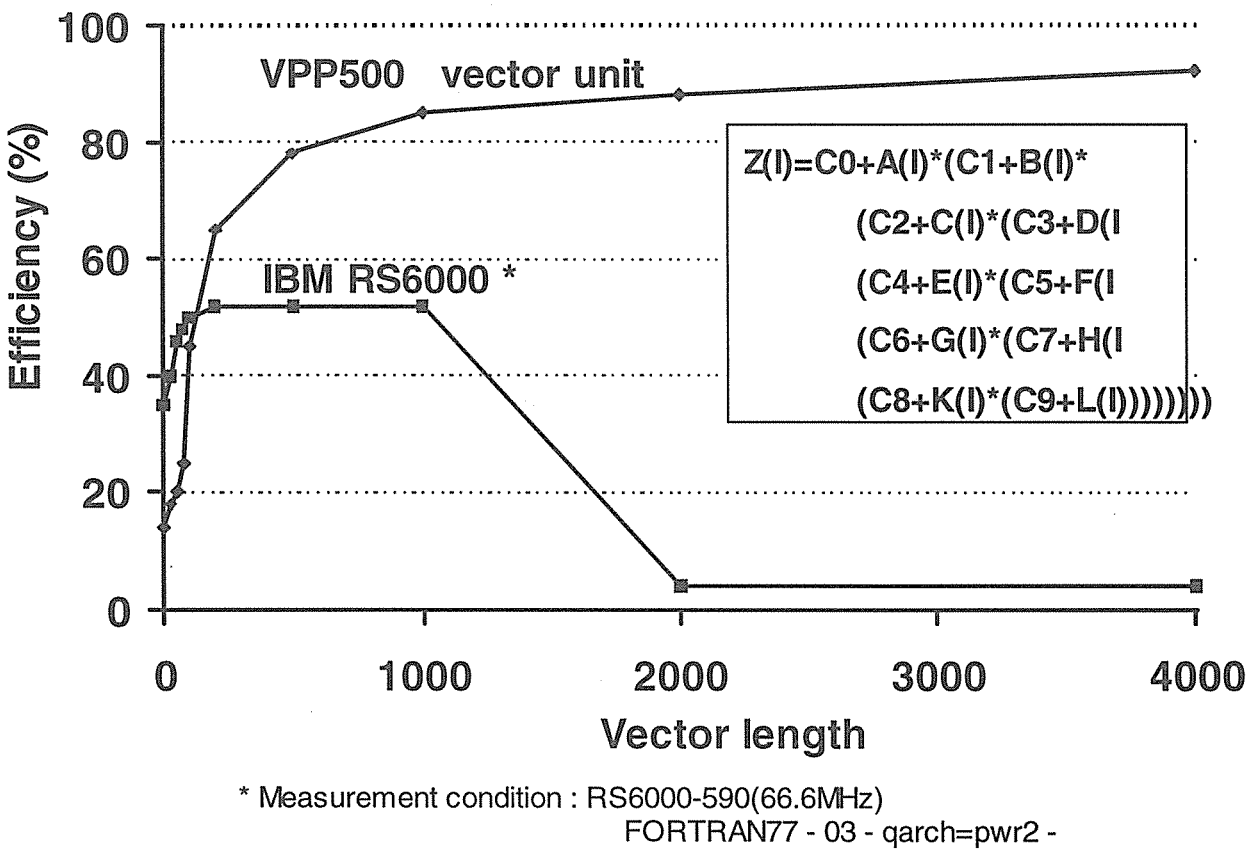


図1. VPP500ベクトルユニットとRS6000の性能比較

表 1. VPP500 ベクトルユニットとマイクロプロセッサの性能比較

| | | RISC Microprocessor | | | | | | Vector | Ratio |
|---|---|---|---|---|---|---|---|---|---|
| | | PA7200 | R10000 | Power2 | Ultra | Alpha | Ave. | VPP500 | |
| Peak MFLOPS | | 240 | 400 | 266 | 333 | 600 | 368 | 1600 | 4.3 |
| All DO Loops (64) | Efficiency (%) | 7.94 | 5.25 | 10.1 | 7.83 | 5.63 | 7.3 | 38 | 5.1 |
| | Sustained MFLOPS | 19.0 | 21.0 | 26.8 | 26.0 | 33.7 | 25 | 603 | 23 |
| DO Loops without IF's (40) | Efficiency (%) | 8.69 | 6.31 | 11.9 | 9.07 | 6.34 | 8.4 | 47 | 5.5 |
| | Sustained MFLOPS | 20.8 | 25.2 | 31.6 | 30.2 | 38.0 | 29 | 755 | 25 |

### *3.2 ベクトルプロセッサの高速演算制御方式*

図 2 にベクトル処理方式で確立・実用されている高速化方式技術を示す。これらの中には、複数の演算パイプラインの並列実行や、浮動小数点数演算及び主記憶アクセスの非同期実行などが含まれる。

マイクロプロセッサの処理能力向上のため、各世代にわたって機能追加が行われてきた。マイクロプロセッサの高速化が図 2 に示すベクトル高速化技術を漸次「導入」することにより実現されている事実に注目されたい。即ち、マイクロプロセッサとベクトルプロセッサの技術動向は以下のように法則化できる。

> The pursuit of higher performance scalar processors leads asymptotically to vector processing.

マイクロプロセッサの進化　　ベクトル処理の高速化方式

| 過去 | 現在 | 未来 | |
|---|---|---|---|
| no | yes | yes | 1. 演算器のパイプライン化 |
| no | yes/no | yes | 2. 同時に複数の演算／ロードストアの実行 |
| no | yes | yes | 3. パイプラインの多重化 |
| no | yes/no | yes | 4. 非同期演算／演算順序の自動変更 |
| no | yes | yes | 5. マスク演算 |
| no | no | yes | 6. プリロードによるスケジューリング |
| no | no | ?? | 7. 大容量レジスタ |
| no | no | ?? | 8. 多様なメモリアクセスパターン |
| no | no | no | 9. 単一命令による複数演算 |
| no | no | ?? | 10. ユーザによるコンパイラ指示 |

図 2. マイクロプロセッサの進化

### 3.3 ベクトルパラレルアーキテクチャ

主記憶を共有して複数のベクトルプロセッサを接続する共有メモリ結合並列方式では、現在および近い将来のメモリ技術においては、数ギガフロップスのベクトルプロセッサを数 10 台程度結合するのに必要なメモリデータを供給する主記憶の実現は極めて困難である。この方式では数 100 ギガフロップスを越えることは不可能である。仮に実現できたとしても共有メモリ上でのアクセス競合、データ供給能力不足、及び複数の CPU キャッシュのデータ一致制御により効率が低下し演算性能を引き出すことは困難になってくる(図 4 の CRAY 社 C90、T90 を参照)。すなわち、メモリ実現の壁により性能のスケーラビリティが頭打ちになる。

共有メモリのデータ供給能力の壁を打破するため、適正規模のベクトルプロセッサと主記憶装置からなる要素計算機(PE)を複数台結合網で接続する分散メモリ型の並列方式を採用することが合理的である。

一方、マイクロプロセッサを数百、数千台結合した超並列計算機(MPP)が多数研究・開発・商用化されてきた。しかし、その実効性能は、ピーク性能の 2%から 30%程度に留まっている[3][6]。この低効率は MPP のアーキテクチャと実現方法に原因がある。

- 1000 台規模のプロセッサを結合する必要上採用しているネットワークが低性能、低機能であり、隣接以外のプロセッサ間通信に対して高い転送効率および使用率を実現できない。
- **3.1** で述べた通り、マイクロプロセッサ性能はキャッシュ動作に依存しており、スーパーコンピュータで頻発する広域アクセスデータに対してキャッシュミスにより強固な性能を発揮できない。
- 主記憶スループットの実現が貧弱であり、性能を発揮できない。

さらに問題は、MPP は従来のプログラムをそのまま走行させることができず、1000 台のプロセッサとその結合を意識した並列プログラミングが必要となることである。MPP の並列プログラムの作成、デバッグおよびチューニングの作業は、想像を越える負担としてユーザにのしかかることになる。今日 MPP は、その高性能に期待していた人々に対して失望以外の何物も与えていないといえる[1][2]。

我々は、以上のような共有メモリ型結合並列方式の限界及び MPP の問題点を解決するため、独自のアプローチ、すなわち分散メモリ型ベクトル並列アーキテクチャを創出した[4][5]。

これが、ベクトル並列処理方式(VPP)であり、その概念の中心は以下にある。

1. 要素プロセッサには、ベクトルプロセッサを配置する。
2. 1. により、要素プロセッサの性能を MPP における要素プロセッサ(マイクロプロセッサ)の 20 倍強とすることができる。
3. 2. により、相互結合する要素プロセッサ台数を MPP の 1/20 に抑えることができ、相互結合ネットワークをクロスバで構築することができる。
4. 3. により、多様な要素プロセッサ間通信形態に対応でき、並列プログラミングが容易になるとともに高い実効性能を発揮できる。また、複数ユーザが同時に使用するようなシステム運用にも効率よく適応できる。これらの優位性はすべてクロスバより発せられる。

   クロスバは、相手のプロセッサが通信中でない限り必ず通信ができ、任意のプロセッサ間の距離がすべて等しく、任意の PE グループに分割が可能であり、オペレーティングシステムによる

動的なPE割付が容易にできる。各PEグループ内のPE間結合は、他グループとは独立なので他グループにおけるジョブの実行に妨害されることはない。また、任意のプロセッサを選択してグループ化してもクロスバの特性は保持されるため、物理PEの割付とは関係なく、実行性能が再現可能である。

さらに、クロスバネットワークはフォルト・トレラント性を有する。データ転送中のいかなる故障も、その影響を及ぼす範囲は送信及び受信中の2台のPEとクロスバパスのみである。

5. 1. により、プログラマブルな大容量ベクトルレジスタを駆使して、メモリアクセスの高度なスケジューリングを行うベクトル技術が確立されており、これを活用できる。ベクトル処理方式により、スーパーコンピュータで頻発する広域アクセスを扱うプログラムに対しても高い実効性能を発揮できる。
6. 1. により、従来のスーパーコンピュータで走行していた逐次プログラムがPE上でそのまま走行することができるというメリットが得られ、並列化プログラム作成の前段階の立ち上がり促進を果たす。さらに、従来の逐次プログラムがベクトルプログラムであれば、アルゴリズムおよびプログラム構造の延長線上で並列化を考えることが可能な場合が多い。

富士通VPPにおいて開発した分散メモリ型ベクトル並列方式は、共有メモリ結合方式とMPPの問題点を解決した実践的な並列計算機であると言える。

## 4 性能

スーパーコンピュータのアーキテクチャは、富士通VPP500において創成した分散メモリベクトル並列型、伝統的共有メモリベクトル並列型、およびマイクロプロセッサベースのMPP/SMP型の3種に大別されることは既に述べた。図3に3種のアーキテクチャの性能特性を示す。MPP/SMPは広い性能レンジをカバーするが、そもそも単体プロセッサの性能が低い為、システムとしての性能も極めて低い。共有メモリベクトル並列型はMPPに対して効率は高いがプロセッサ台数の増加に伴って、メモリ実現上の物理限界から結合台数が頭打ちとなる。富士通分散メモリベクトル並列型は単体性能の効率が高く、プロセッサ台数の増加に伴う効率低下が緩やかであり、多数台のプロセッサを結合可能である為、より高い性能レンジもカバーできるというスケーラビリティの優位性がある。図4に並列性能のベンチマークプログラムであるNAS Parallel Benchmarkテストの結果を示す。富士通のVPP500及びVPP700は分散メモリベクトル並列型、Cray社のC90, T90、及びNECのSX-4は共有メモリベクトル並列型、SGIのPower Challenge XL、IBM社のSP-2、及びCray社のT3D, T3Eは、MPP/SMPである。図4に示す実測データは図3の性能特性を良く表している。

## 5 VPP開発の歴史

富士通のVX/VPP300/VPP700は全面的にCMOSテクノロジーを採用し、VPP500と互換性のある商用のスーパーコンピュータである。VPP500は、航空宇宙技術研究所と富士通の間で1989年に開始された実現検討及び共同研究の成果を取り入れて商用化したシステムであり、この共同研究は数値風洞(NWT)の開発につながった。この研究の中でNALはベクトル処理ノードの概念、性能設定の合理性において、また富士通はクロスバネットワークとPE間通信やそのハードウエア／ソフトウエアのインプリメンテーションにおいて貢献した。

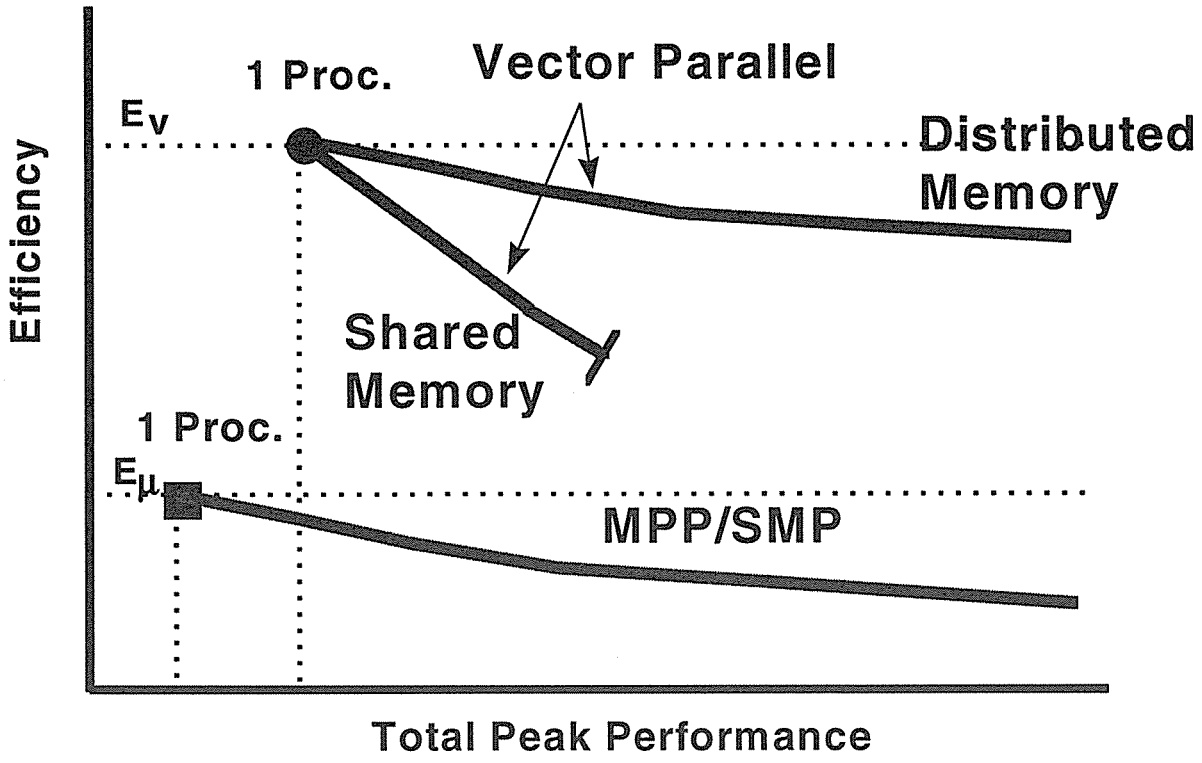


図３．アーキテクチャによる性能比較

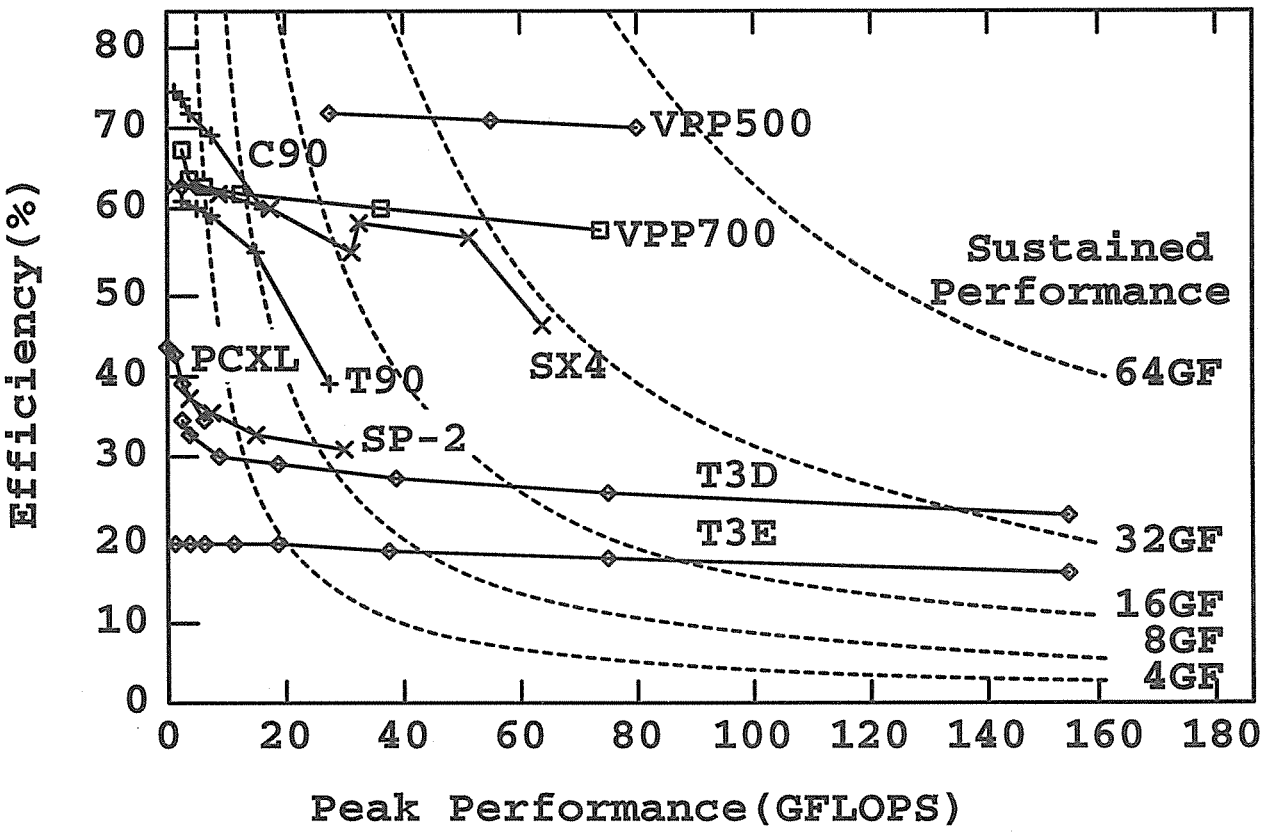


図４．NAS Parallel Benchmark 1.0 -- BT Class B の結果

## 6 今後の展望

富士通はVPP500およびVX/VPP300/VPP700で確立したベクトルパラレルアーキテクチャを今後とも維持・発展させ、かつ上位互換を維持する。次期VPPは64ビットのアドレッシングと64ビット整数演算機能をサポートする。その他多くの設計改良を行いスカラ及びベクトル処理の両面で実効性能の向上を図る。向上した演算処理能力に釣り合うように、主記憶容量とその性能、及びPE間通信スループットとその立ち上がり時間の向上を図る。また、限りなく増大するI/O処理要求に応えるため、より高速なI/Oプロセシングエレメント(IOPE)、及び並列分散ファイルシステムによりI/O性能の向上を図る。

オペレーティングシステムや言語処理システムは、さらなる機能追加と性能向上を図る。

## 7 おわりに

富士通VPP500において創成された分散メモリベクトル並列処理方式(VPP)は、従来のアーキテクチャの欠点を克服する鮮明な開発思想に基づく「フレーバ」を有している。VPPシステムを活用することによりユーザにおいてこれまで実現できなかった大規模科学技術計算を強力に推進する事を大いに期待する。

## 8 参考文献


1. J. Worlton, "The MPP Bandwagon," HPCwire report, May 5, 1992.
2. Gordon Bell, "Ultracomputer: A Teraflop Before Its Time," *Communications of the ACM*, Vol. 35, No. 8 Aug. 1992, pp. 27-47.
3. Subhash Saini and David H. Bailey, "NAS Parallel Benchmark Results 12-95," Report NAS-95-021, NASA Ames Research Center, Moffett Field, CA 94035, December, 1995.
4. Moriyuki Takamura and Teruo Utsumi, "Why Vector Parallel?," *The Proceedings of the HPC Conference '94*, Singapore, Sept. 1994, pp. 394-398.
5. Akira Nodomi, Masayuki Ikeda, Moriyuki Takamura, and Kenichi Miura, "Hardware Performance of the VPP500 Parallel Supercomputer," *High Performance Computing : Technology, Methods and Applications* (North-Holland, Amsterdam, 1995), pp. 103-120.
6. David H. Baily, Elic Barszcz, Leonard Dagum and Horst D. Simon: "NAS Parallel Benchmark Results", Tech. Report RNR-94-006, NASA Ames Research Center, Moffett Field, CA94035, March 21,1994.