計算機アーキテクチャ 73－6
（1988. 10. 31, 11. 1）

スーパコンピュータにおける記憶階層について

阿部 仁　和田 英夫　石井 幸一　河辺 峻
（日立製作所）

スーパコンピュータの内部計算能力は飛躍的に向上しており、これに伴い一段と高い入出力処理能力が必要である。しかし現在もその差は大きく、この問題を技術的に解決する必要がある。

スーパコンピュータ HITAC S－820では、主記憶と外部記憶の間に「拡張記憶」と呼ぶ記憶階層を設けている。これは主記憶と外部記憶の中間に位置付けられる、高速・大容量の内部半導体記憶である。これによって、特にベクトルジョブの実行時間の短縮、OSの多重ベクトルジョブ実行環境の改善を行っている。

本文では、S－820拡張記憶の位置付けおよび性能改善の効果について報告する。

AN EFFECT OF THE EXTENDED STORAGE FOR SUPERCOMPUTER

Hitoshi ABE, Hideo WADA, Kouichi ISHII and Shun KAWABE

Kanagawa Works, HITACHI, Ltd.
1 Horiyamashita, Hadano, Kanagawa, 259-13, Japan

In recent years a vector performance of a supercomputer is highly increasing, and according to it more higher I/O process performance is also nesesary. But now a gap between them is apparently large. Therefore some technical improvements must be done.

We desighned the extended storage for the supercomputer HITAC S-820, which is a highspeed and large capacity memory of semiconductor between a main storage and auxiriary magnetic disks. By an effective use of the extended storage, an execution time of vector job is shortened, and a direct time sharing execution of vector jobs is possible.

We report an effect of the extended storage for the supercomputer HITAC S-820.

１．はじめに

スーパコンピュータの内部計算能力は飛躍的に向上しており、これに伴い一段と高い入出力処理能力が必要である。しかしその差は大きく、この問題を技術的に解決しなければならない。

スーパコンピュータＨＩＴＡＣ Ｓ－８２０では、主記憶と外部記憶の間に拡張記憶と呼ぶ高速・大容量半導体記憶を設け、この問題を改善しようとしており、ベクトルジョブの実行時間の短縮、ＯＳの多重ジョブ実行環境の改善に効果を発揮している。

本文では、Ｓ－８２０拡張記憶の位置付けおよび性能改善の効果について報告する。

２．問題点

（１）一般にスーパコンピュータは、汎用のコンピュータに比して大容量の主記憶を実装できるが、それでも大規模ジョブでは必要なデータ及び途中結果が全て主記憶に入り切らないことがあり、そのためジョブ実行中にディスクへのＩ／Ｏ処理が必要となる。するとスーパコンピュータの内部計算の速度向上に対応してＣＰＵ時間が飛躍的に短縮されるにもかかわらず、総実行時間はあまり短縮されないことになり、Ｉ／Ｏの多いジョブについてはスーパコンピュータ利用の効果が減少してしまう。

（２）また上記のようにスーパコンピュータのジョブは多量の主記憶を占有するため、多重ジョブ環境において、ジョブを一旦スワップアウト／インしようとするとディスクへのＩ／Ｏ処理が非常に多くなり、各ジョブへの影響とシステム全体の効率低下が生じる。したがって、スーパコンピュータの大規模ジョブのスワップアウト／インは、多くの制約があり困難である。

ＨＩＴＡＣＳ－８１０／Ｓ－８２０では、この点を改善すべく、主記憶と外部記憶の間に位置する高速大容量の半導体記憶（拡張記憶）を設置している。特にＳ－８２０では、後者の問題にも対応できる拡張記憶を開発した。

３．Ｓ－８２０概要

ＨＩＴＡＣ Ｓ－８２０は、最大ピーク性能３ＧＦＬＯＰＳを実現した、ユニプロセッサでは世界最高速のスーパコンピュータである。Ｓ－８２０はパイプライン方式のプロセッサであり、ベクトルプロセッサ（ＶＰ）、スカラプロセッサ（ＳＰ）、主記憶（ＭＳ）、拡張記憶（ＥＳ）、ＩＯプロセッサ（ＩＯＰ）、サービスプロセッサ（ＳＶＰ）よりなるシステムである。図１にＳ－８２０モデル８０の構成を示す。

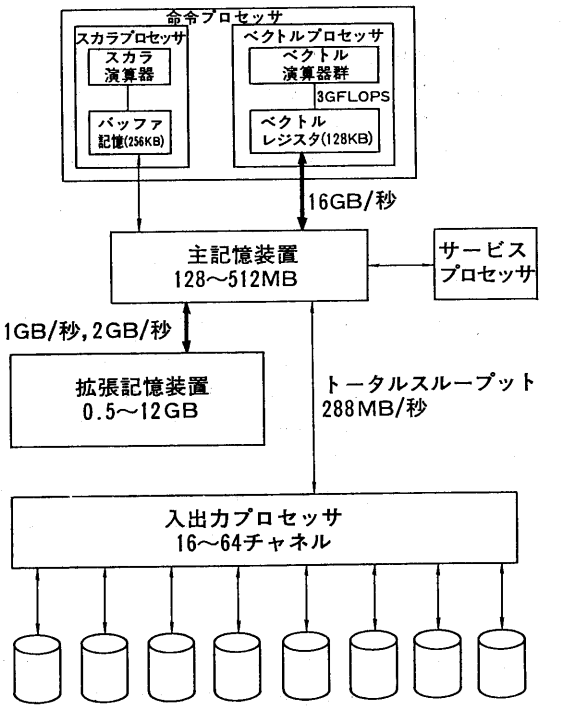


図１　Ｓ－８２０／８０システム構成

スカラプロセッサは、通常のＭシリーズの汎用命令とベクトルプロセッサ制御命令を実行できる。一方、ベクトルプロセッサは、専用の加算パイプライン、乗算パイプライン、除算パイプラインを持ち、ベクトル演算を高速にパイプライン処理できる。演算に必要なデータおよび演算結果のデー

タは超高速のベクトルレジスタに格納される。ベクトルレジスタのデータは、ロード／ストアパイプラインを通じて主記憶と転送される。（スカラプロセッサのようにキャッシュメモリは介さない。）主記憶とベクトルレジスタの転送を司るロード／ストアパイプラインは、合計して１６Ｇバイト／秒の転送能力を持っている。

一方主記憶と入出力デバイスとのデータ転送は、ＩＯプロセッサが行う。ＩＯＰは最大６４チャネルまで制御可能であり、チャネルの下に各種のＩ／Ｏデバイスを接続する。その総転送能力は最大２８８Ｍバイト／秒あるが、チャネルの転送能力は、各々最大３～６Ｍバイト／秒である。

拡張記憶（ＥＳ）は、主記憶（ＭＳ）と接続される半導体記憶装置であり、主記憶と外部記憶の間に位置する階層記憶をなしている。ＭＳとデータ転送を行うことができ、その最大転送能力は２Ｇバイト／秒である。また、ＭＳが最大容量５１２ＭＢ実装可能なのに対し、ＥＳは最大容量１２Ｇバイトまで実装可能である。このようにＥＳは、ＭＳに対して２４倍の容量を持つことができ、チャネルの転送能力に対し３桁近く高速である。

４．拡張記憶のハード

半導体メモリ素子の集積度は年々進歩しており、多量の半導体記憶装置を安価に使用することが可能になってきている。Ｓ－８２０の拡張記憶も１ＭビットＤＲＡＭを使い、基板実装の技術水準を最高度に生かし、スペース効率及びコストパフォーマンスを高めている。

表１に拡張記憶の実装諸元を示す。１ＭＤＲＡＭ ５７６個を、多層プリント基板の表裏両面に実装し、１枚当り６４ＭＢの記憶容量を実現している。本プリント基板を、９６枚、電源装置などとともに90×86×172 cm³の筐体に実装する。この筐体１個で容量６Ｇバイト、２個で最大容量の１２Ｇバイトを実現している。

磁気ディスク装置に対してスペース効率も良く、コスト対性能比も良い。

ＥＳはメカニカル部分を含まず、信頼性も高い。使用している半導体記憶素子の記憶障害に対しては、ＥＣＣコードによるメモリ障害の検出／訂正を行っており、２ビットの誤り訂正も可能である。

表１ Ｓ－８２０ＥＳハード諸元

使用素子	１ＭビットＤＲＡＭ
パッケージ	６４ＭＢ／パッケージ 両面実装
転送速度	最大２Ｇバイト／秒
容量	最大１２Ｇバイト
ｴﾗｰﾁｪｯｸ	２ビットエラー訂正

５．ベクトルジョブ性能の改善

前述のようにスーパコンピュータのベクトル性能が増すにつれて、ＣＰＵ時間は短縮されるが、Ｉ／Ｏ時間も短縮されないと、総実行時間は余り短縮されない。Ｉ／Ｏが多くベクトル化率が低いものは、スーパコンピュータで実行する効果がないという結果になる。すなわち、ベクトルジョブのＩ／Ｏ時間の短縮が重要になってくる。Ｓ－８２０では、ディスク上ではなく、ＥＳ上にファイルを作成し、主記憶とのデータ転送を行えば、この処理時間が飛躍的に短縮。できる。

Ｓ－８１０／Ｓ－８２０用ＯＳは、ＥＳのファイルをＦＯＲＴＲＡＮの標準入出力としてサポートしている。ＯＳは、ジョブのＥＳ使用宣言をＪＣＬ（ジョブ制御言語）から読み取り、自動的にこのファイルをＥＳ上に割り当てる。例えばＪＣＬのファイルのＤＤ文で、ＵＮＩＴ名＝ＤＡＳＤならディスクを、ＵＮＩＴ名＝ＥＳならＥＳを、そのファイルの実体として割当てる。したがって、ＦＯＲＴＲＡＮソースプログラムを書き換えて、ＲＥＡＤ／ＷＲＩＴＥ文の

機器指定等を変更する必要もなく、非常に簡単にＥＳを使用することができる。

Ｉ／Ｏファイルが多数あり全てＥＳ上に割当てると、ＥＳの割当て上限値を越えてしまうような時は、そのうちＩ／Ｏ時間および回数が多く、ＥＳ化により最も効果が上がるファイルを選んで、ディスクからＥＳ使用に切り替える事でも良い。

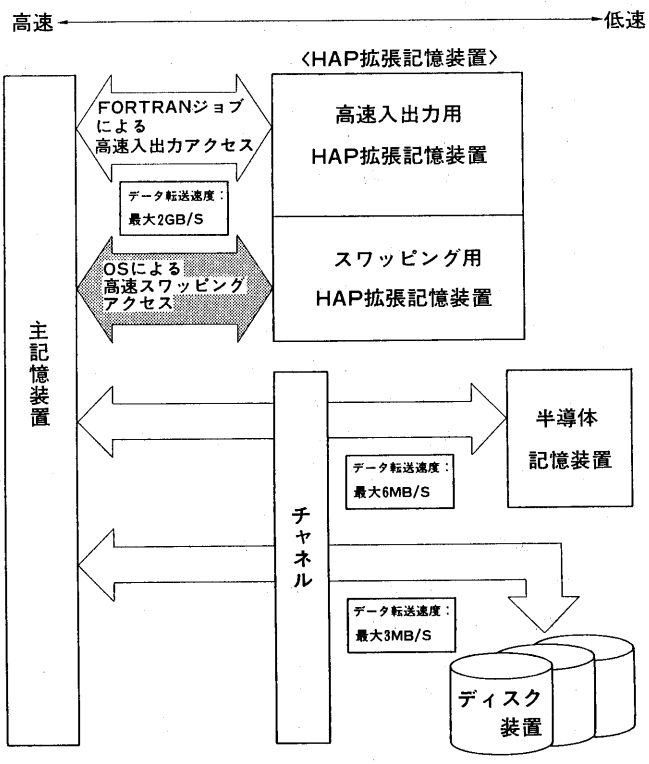
６．ベクトル多重ジョブ環境の改善

前述のように多量のＭＳ領域を使用しているベクトルジョブは、実行途中でディスク装置にスワップアウト／インすることはＩ／Ｏネックを招き、性能上好ましくない。特に、複数のベクトルジョブの同時実行環境、例えばベクトルジョブのＴＳＳ下での直接実行などはその応答速度に問題が生じる。そこで、Ｓ－８２０では、ＥＳをスワッピングファイルとしても用いることとした。これによりＭＳ中の領域をＥＳに高速にスワップアウト／インするため、マルチジョブ実行時のジョブ多重度の制約を軽減でき、ベクトルジョブのＴＳＳ直接実行環境の提供が可能となる。

このためにＥＳをベクトルジョブのユーザファイル用とスワッピング用に分割できる。この分割は、各センタの運用状況に応じて、ＩＰＬ時のシステムパラメータとして設定することが出来る。

７．転送制御方式

データの転送方式は、２種類あり、それぞれを使いわけている。ベクトルジョブの中間ファイルとしてのＭＳとＥＳとのデータ転送には、同期型転送が用いられる。一方、スワッピングのためのＭＳとＥＳのデータ転送には非同期型転送が用いられる。
同期型転送では、スカラプロセッサの命令制御部がこの転送を制御し、終了まで命令制御部は次の命令を実行しない。



効果　(1)ベクトルジョブスワッピングの高速化
　　(2)チャネルビジー等の入出力ネックの緩和

図２　ＥＳによる性能改善

非同期型転送では、スカラプロセッサの命令制御部はＥＳに対して非同期転送の起動のみ行い、起動後は次の命令実行に移る。ＥＳは命令制御部が与えた起動情報に従い、まずコマンド語（転送指令）をＭＳからフェッチし解読したうえで、この指令によるデータ転送動作を開始する。即ち、Ｉ／Ｏ動作でのチャネルの動作と同様のことを行う。Ｓ－８２０では複数のコマンド語があり、データ転送以外にコマンド語アドレスのジャンプ等も指定でき、コマンド語のチェインができる。一連のコマンド語の実行が終了するとＥＳは、スカラプロセッサに対して割り込みを起こす。これにより、非同期転送の終了と終了状態とを検知できる。

上記の形態を使いわける理由は、ベクトルジョブ中のファイルアクセスでは一回当りのデータ転送量が必ずしも多くないので、同期型転送方式で転送終了まで待つ時間が非常に短く、非同期型転送方式により割り込み処理ルーチンを走ってからジョブに復帰するより、ジョブ性能に有利な場合が多いためである。これに対して、非同期型転送は、多量のデータ転送を伴うことが多いので、Ｉ／Ｏ処理と同様な使い方を採用している。

８．Ｓ－８１０拡張記憶との相違

Ｓ－８２０と前期機種Ｓ－８１０の拡張記憶の相違を表２に示す。

表２　Ｓ－８１０／８２０ＥＳの相違点

	Ｓ－８１０	Ｓ－８２０
転送形態	同期転送	同期転送 非同期転送
転送速度 (GB/s)	１or０．５ max	２or１ max
容量	３ＧＢmax	１２ＧＢmax

Ｓ－８２０では、ＥＳの転送速度の向上およびジョブスワッピングのための非同期転送が追加されている。

９．拡張記憶の効果

以下に拡張記憶の効果についての例を示す。

例１

（Ｓ－８２０使用、テストプログラム、シングルジョブ環境下、ＥＳを６４ＭＢ使用して測定）

単位：秒	ベクトル処理 ＋ディスク	ベクトル処理 ＋ＥＳ
ＣＰＵ時間	０．８２	０．０８
Ｉ／Ｏ時間	１３９．０７	－
総実行時間	１３９．８９	０．０８

このプログラムは、数値をＥＳ（or ディスク）へ書いて、それを読み出す動作を繰り返すもので、合計６４ＭＢの書き込みと６４ＭＢの読み出しを行っている。ＥＳを使用すると、ＣＰＵ時間＝総実行時間（経過時間）となっており、Ｉ／Ｏ時間は０になった。

例２

（Ｓ－８２０使用、数値計算実ジョブ、シングルジョブ環境下、ＥＳを３１ＭＢ使用して測定）

	ベクトル処理 ＋ディスク	ベクトル処理 ＋ＥＳ
ＣＰＵ時間	１１１１秒	１０９５秒
Ｉ／Ｏ時間	１９８１秒	３０６秒
総実行時間	３０９２秒	１４０１秒

上記の例では、全てのＩ／ＯがＥＳ化されてはいないが、Ｉ／Ｏ時間は相当に短縮されている。また、拡張記憶の制御はＩ／Ｏ制御に比べて簡単なので、Ｉ／ＯをＥＳに切り換えると、ＣＰＵ時間として計測されている時間も多少減少している。

例３

（Ｓ－８２０、マルチジョブでのスワッピング改善例）

図３に測定結果を示す。各ジョブは、コンパイル、リンケージ、実行の３ステップからなる。ジョブ多重度を３とした時の必要な総記憶容量は、実主記憶容量の１．５倍である。

ディスク使用時は、各ジョブの実行時間は、シングルラン時の約８倍を要した。これに対し、ＥＳ使用時は３ジョブの終了時までに、シングルラン時の２．５倍で済んだ。

また、ディスク使用時のＣＰＵ使用率の低下に比べて、ＥＳ使用時にはＣＰＵをほ

ぼ１００％使い切っている。
（図３ではジョブのシングルラン時の実行時間を１として、相対値で実行時間を示している。）

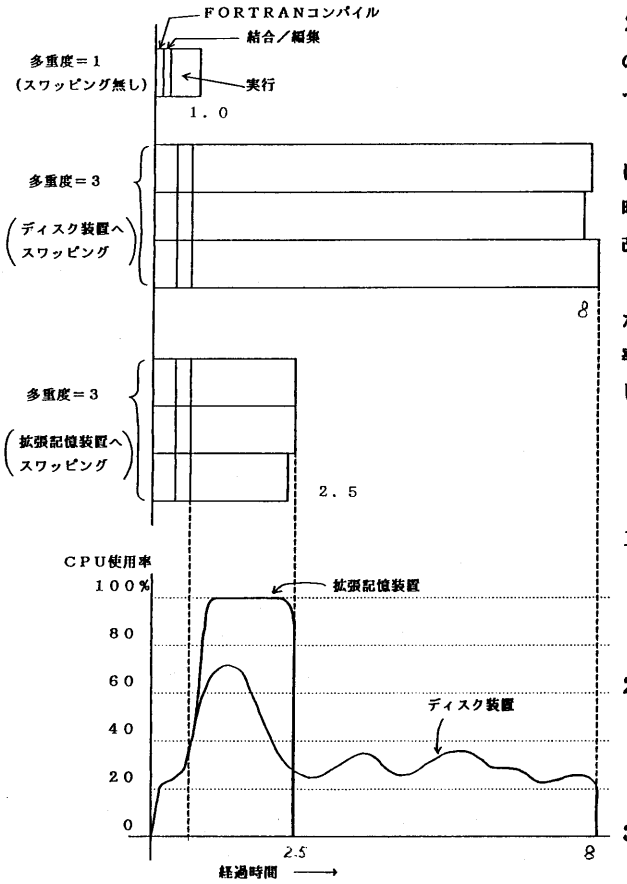


図３　ＥＳによるスワッピング効果例

１０．拡張記憶装置の今後

拡張記憶はスーパコンピュータの階層記憶の一つとして、今後とも重要な意味を持つ。その主記憶に対する大容量性とＩ／Ｏに対する高速性という特長から、Ｉ／Ｏとは異なる高速インタフェースを外部に対して持つことを考えることができる。これにより、大量のデータを高速に転送する。

また今後増えるマルチプロセッサ型のスーパコンピュータの高速の共用記憶装置として、現在のＥＳの技術をさらに発展させることが考えられ、今後の検討課題である。

１１．おわりに

スーパコンピュータＨＩＴＡＣ　Ｓ－８２０の記憶階層、特に拡張記憶装置（ＥＳ）の位置付けおよびその性能に対する効果について報告した。

Ｓ－８２０では、主記憶と外部記憶の間に拡張記憶を設け、ベクトルジョブの実行時間の短縮、ＯＳの多重ジョブ実行環境の改善を計っている。

スーパコンピュータは、今後も適用範囲がますます拡大する方向にあり、さらに効率的なコンピュータシステムの開発を推進していきたいと考えている。

参考文献

1）河辺、他「シングルプロセッサで最大性能２ＧＦＬＯＰＳのＳ－８２０」日経エレクトロニクス、１９８７年１２月２７日号、NO437, pp111-125

2）大房、他、「スーパコンピュータ　ＨＩＴＡＣ Ｓ－８２０のオペレーティングシステム」、日立評論、１９８７年１２月号、Vol69、pp21-26

3）長堀、他、「ＦＯＲＴＲＡＮベクトルコンパイラとチューニングソフトウェア」、日立評論、１９８７年１２月号、Vol69、pp27-34