U.D.C. 681.322-181.2-185.4:681.32.066

スーパーコンピュータHITAC S-820のオペレーティングシステム

Operating Systems for Supercomputer HITAC S-820

大房義隆*　*Yoshitaka Ôfusa*
片田　久**　*Hisashi Katada*
吉澤康文***　*Yasufumi Yoshizawa*
上　政之****　*Masayuki Kami*

スーパーコンピュータ用のオペレーティングシステムは，利用者の持っているプログラム財産と高速なベクトル演算機能を持つハードウェアとを結び付け，それぞれの資源を効率よく利用できるシステム運用機能を提供する必要がある。

日立製作所は，従来からスーパーコンピュータHITAC S-810を提供していたが，今回このS-810の上位のプロセッサグループとしてHITAC S-820を開発し，S-810で培った技術を基に新しいアーキテクチャを採用し，演算性能の強化に加えて，システムの運用性についても大幅な機能強化を図った。

これにより，31ビットアドレッシング方式の利用や，対話環境での技術計算プログラムの実行によるプログラム開発の生産性向上，多様な運用環境でのシステム資源の効率のよい利用，更に，拡張チャネルシステムや拡張記憶装置などの利用によるシステム性能の強化を図ることができた。

1 緒　　言

スーパーコンピュータは，ハードウェアの高速演算能力と，それを有効に活用するソフトウェア，特に，FORTRANに代表される技術計算向け言語による自動ベクトル化などの高度な最適化技術によって，従来のはん(汎)用プロセッサでは成し得なかった大規模かつ複雑化する技術計算ニーズにこたえてきた。

一方，スーパーコンピュータの運用方式は，従来，技術計算業務を主体として運用するケースが多かったが，最近は，技術計算業務に加えて，一般のバッチ業務や対話処理によるプログラム開発業務，更に画像処理への適用といった多様化する利用形態を同時に処理し，より効率よく運用できるスーパーコンピュータが必要となっている。

本稿では，スーパーコンピュータHITAC S-820(以下，S-820と略す。)の開発に伴い，S-820を制御するオペレーティングシステムについても大幅な機能強化を図ったので，その制御方式の特長，及び代表的な機能の概要について紹介する。

2 オペレーティングシステムの概要

2.1 S-820システムのハードウェア構成

S-820システムのハードウェア構成を**図1**に示す。

命令プロセッサは，HITAC Mシリーズのはん用プロセッサの持つ命令を実行するスカラープロセッサと，スーパーコンピュータの能力を最大限に発揮するベクトルプロセッサで構成され，それぞれが並列に動作可能である。

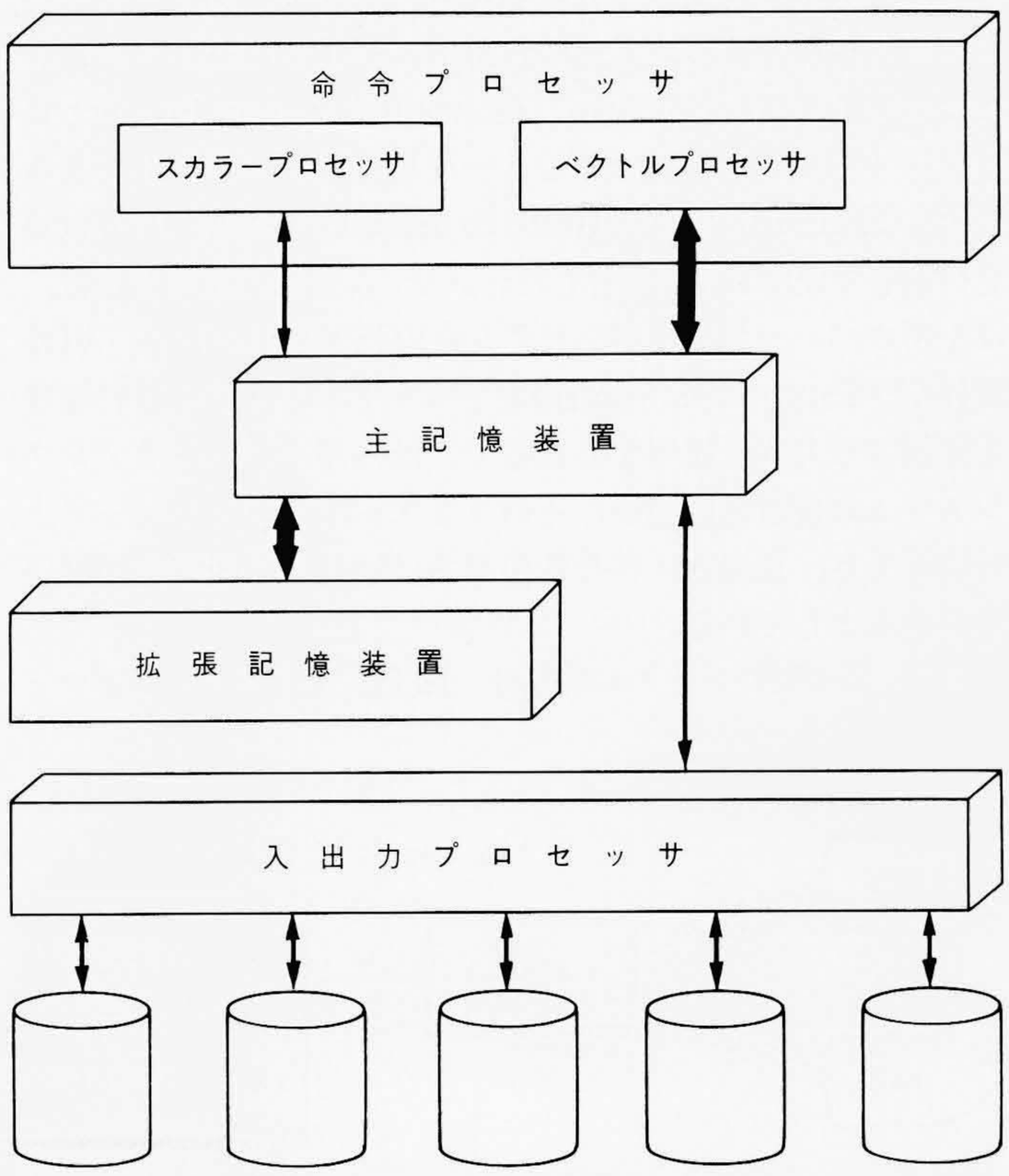


図1　S-820システムのハードウェア構成　　高速なベクトル演算を実現するベクトルプロセッサ，高速なデータ転送能力を持つ拡張記憶装置，また，入出力プロセッサは拡張チャネルシステムを搭載している。

* 日立製作所ソフトウェア工場　** 日立製作所システム開発研究所　*** 日立製作所システム開発研究所 工学博士
**** 日立ソフトウェアエンジニアリング株式会社

主記憶装置は，高速・高集積のスタティックメモリ素子によって構成され，最大512Mバイトの大容量化を実現している。

拡張記憶装置は，1Mビットのダイナミックメモリ素子によって構成され，最大12Gバイトの大容量化を実現した。プログラムの一時的入出力データセット，及びベクトルジョブのスワッピング用データセットとして使用する。

入出力プロセッサは，拡張チャネルシステムによって最大64チャネルまで接続でき，ブロックマルチプレクサチャネルは，3Mバイト/秒及び6Mバイト/秒のチャネルが使用可能である。

S-820の特長は，はん用プロセッサに装備されていないベクトルプロセッサと拡張記憶装置にある。オペレーティングシステムは，これらの資源をセンタ管理者の運用方針に応じて最適に配分し，ベクトルジョブを効率よく実行させる。

2.2 オペレーティングシステムの構成

スーパーコンピュータ用のオペレーティングシステムは，以下に示す方針に基づいて開発した。

(1) ベクトルプロセッサ，拡張記憶装置などを，ユーザープログラムで容易に利用できる。

(2) スーパーコンピュータシステムの高速演算能力を最大限に発揮させる。

(3) HITAC Mシリーズのはん用システムが持つ機能はそのまま利用可能とし，センタ運用が容易かつ柔軟にできる。

上記の方針に基づき，日立製作所のHITAC Mシリーズのはん用オペレーティングシステムであるVOS3/ES1(Virtual-storage Operating System 3/Extended System Product 1)をベースに，スーパーコンピュータ専用のプロダクトであるVOS3/HAP/ES(VOS3/HAP/Extended System Product)を付加して使用する形態とした。システム全体の運用管理は，はん用オペレーティングシステムのVOS3/ES1が行い，VOS3/HAP/ESは，VOS3/ES1の下でベクトルジョブの実行管理を分担する方式を採用することで，はん用オペレーティングシステムの豊富な機能が，そのままスーパーコンピュータ上で利用でき，互換性，移行性の高さに加えて，親和性の高いシステムとしている。

また，S-820のシステム構成は，**図2**に示すスタンドアロン構成や，**図3**に示すHITAC Mシリーズはん用プロセッサとの疎結合マルチプロセッサ構成など，システム規模や運用形態に合わせた柔軟なシステム構成を構築できる。

VOS3/HAP/ESの特長的な機能については**3**章で述べる。

2.3 関連プロダクトの概要

2.3.1 SAR/D/ES

SAR/D/ES(System Activity Report/Display/Extended System Product)では，従来からシステム資源の使用状況，ジョブの動作情報を表示し，VOS3システムの効率のよい運転を支援する機能を提供していたが，以下に示すベクトルジョブ用資源の動作状況を表示する機能を追加した。

(1) ベクトルプロセッサの使用状況

(2) 主記憶装置の使用状況

(3) 拡張記憶装置の使用状況

2.3.2 SAMH E2

大規模な配列処理などでは，処理データを主記憶装置にすべて格納できない場合，そのデータは外部記憶装置と主記憶装置間の入出力処理によるデータ転送が必要となる。この入出力時間を大幅に短縮する機能として，パラレル入出力(複数ボリューム並列入出力)機能をFORTRAN言語の機能として提供していた。

SAMH E2(SAM High-speed Performance option Extended Version 2)は，このパラレル入出力機能を，アクセス法QSAM(Queued Sequential Access Method)，BSAM(Basic Sequential Access Method)の高速オプションとしてはん用化し，かつ性能，操作性を向上させるものである。主な特長を以下に示す。

(1) パラレル入出力による入出力時間の短縮

(a) データセットを複数のディスクボリュームに分散割当てし，並行入出力を行う。

(b) 入出力時間はボリューム数(パラレル度)をNとすると，最高$\frac{1}{N}$に短縮可能である。

(c) 大容量バッファを使用するため，バッファを31ビットアドレッシング領域に確保する。

(2) 操作性の向上

プログラム変更が不要で，1枚のジョブ制御文(DD文)の指

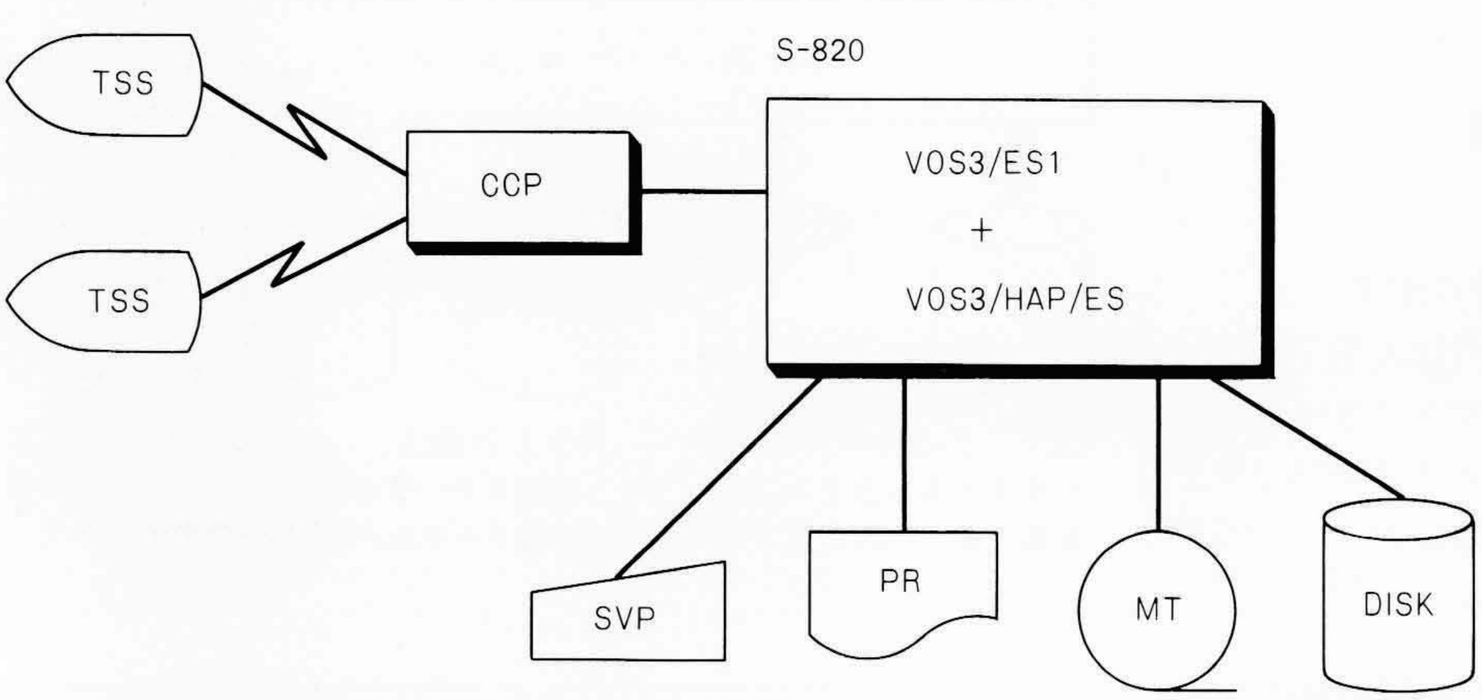
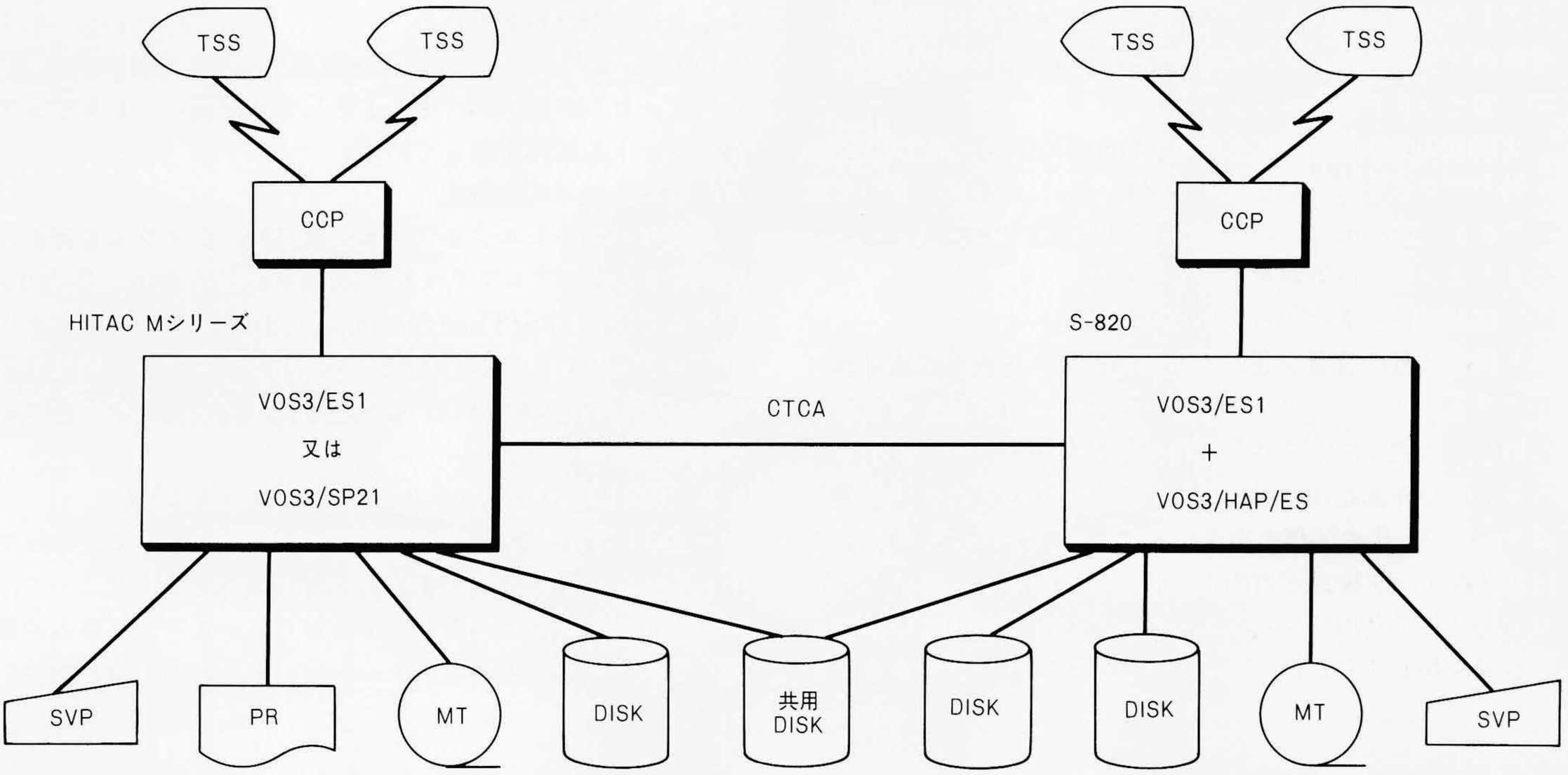


図2 S-820のシステム構成例(スタンドアロンの場合) 1台のS-820で，技術計算業務，一般のバッチ業務，更にTSSによるプログラム開発業務などを同時に運用できる。



注：略語説明　CTCA(Channel to Channel Adapter)，VOS3/SP21(VOS3/System Product 21)

図3　S-820のシステム構成例(疎結合マルチプロセッサの場合)　技術計算業務と，それ以外の業務に負荷分散を図る。S-820は技術計算業務の専用プロセッサとし，JSS 4(Job Spooling Subsystem 4)を利用し，共用DISK経由でジョブを受け渡しする。

定だけで，データセットの複数のボリュームへの自動割当てが可能である。

3　VOS 3/HAP/ESの機能

この章では，はん用オペレーティングシステムであるVOS 3/ES1の下でベクトルプログラムの実行を管理するVOS3/HAP/ESの特長を**3.1**節に，S-820システムで強化した機能を**3.2**節以降で述べる。

3.1　VOS3/HAP/ESの特長

VOS3/HAP/ESの特長を**図4**に示す。

(1)　ベクトルジョブの実行制御機能

VOS3/ES1では，ジョブクラスと呼ぶジョブのグループごとに使用可能な資源の割当て量や多重度を，システムの管理者が決定できる。VOS3/HAP/ESでは，主記憶装置，拡張記憶装置の割当て量及びCPU(Central Processing Unit)の割当て方式などをジョブクラスごとに設定できる。これによって，

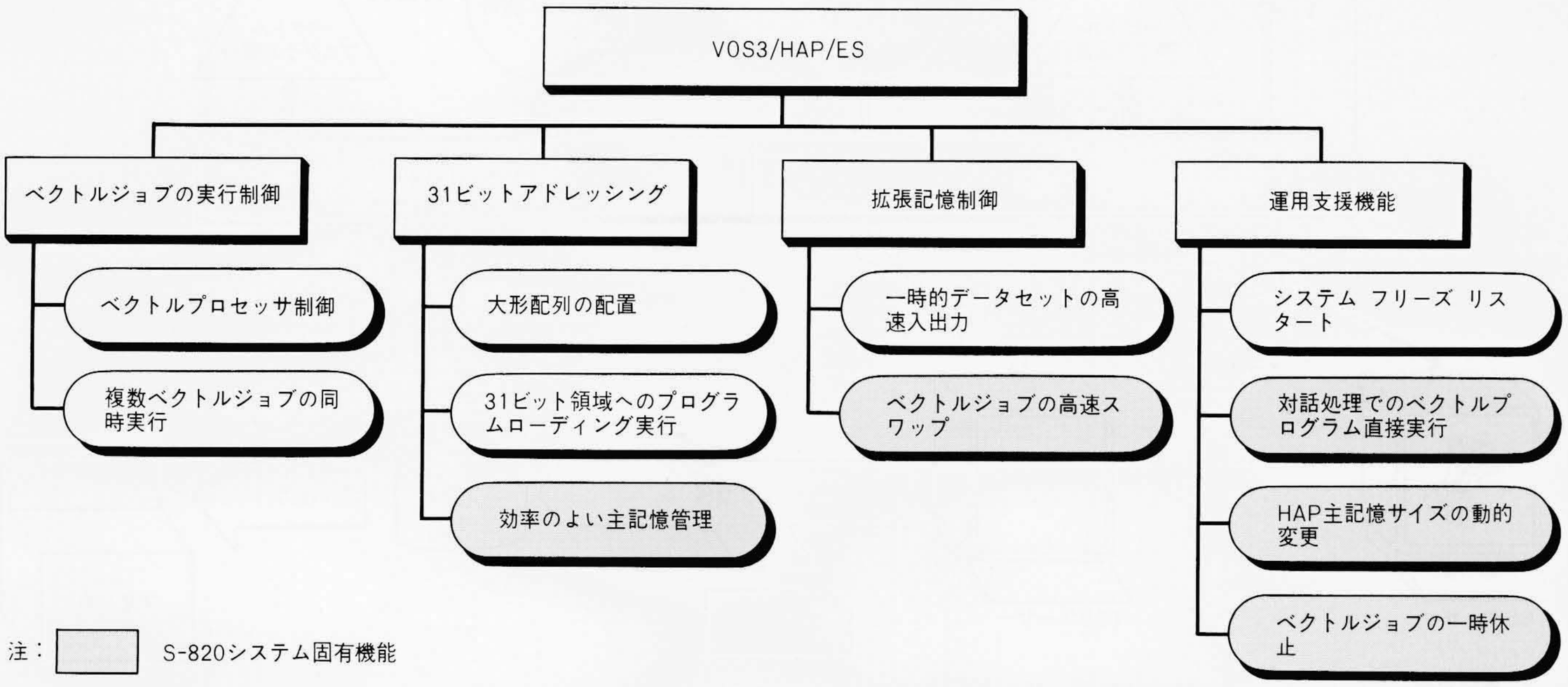


図4　VOS 3/HAP/ESの特長　VOS 3/HAP/ESは，ベクトルプロセッサ，主記憶装置及び拡張記憶装置のハードウェア資源を効率よく制御する。また，システムを柔軟に運用するための豊富な機能を提供する。

ベクトルジョブのスケジューリング方式を柔軟に設定でき，複数のベクトルジョブの同時実行を効率よく実現できる。

(2) 31ビットアドレッシング

大規模な科学技術計算を高速に行うために，FORTRANの配列やプログラムを31ビットアドレッシング領域に置いて実行できる。S-820システムでは，仮想記憶方式の利用により，効率のよい主記憶管理を実現している。

(3) 拡張記憶装置による入出力の高速化

主記憶装置の容量を超える大規模データを扱う技術計算プログラムでは，データを磁気ディスクなどの外部記憶装置に格納し，必要に応じて外部記憶装置からデータの一部を取り込み，再び外部記憶装置に格納するという入出力処理を繰り返す。この入出力処理を高速化するため，VOS3/HAP/ESでは高速なデータ転送が可能な拡張記憶装置を，一時的データセットとして利用できる。ユーザーはジョブ制御文に拡張記憶装置を指定するだけでよく，ユーザープログラムを変更することなく使用可能である。

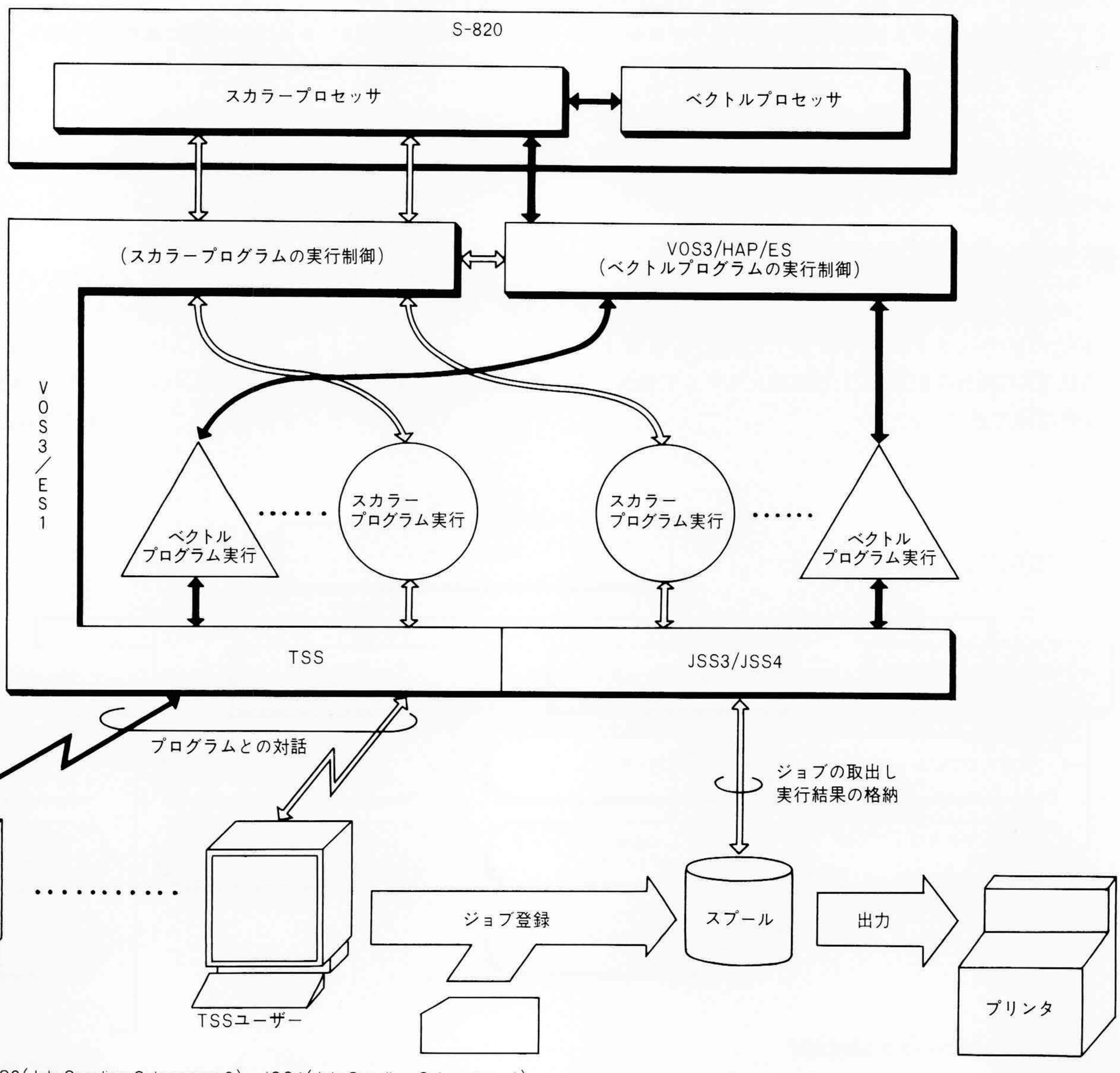
S-820システムでは，拡張記憶装置をベクトルジョブの高速なスワッピング用データセットとしても使用し，システムスループットの強化を図っている。

(4) 豊富な運用支援機能

実行中のベクトルジョブを残したまま，システムを停止・再開できるシステムフリーズリスタート機能に加え，S-820システムでは，TSS(Time Sharing System)ジョブでのベクトルプログラムの直接実行機能，ベクトルジョブの一時休止機能及びHAP主記憶サイズの動的変更機能などにより，柔軟なシステム運用が図れる。

3.2 TSSにおけるベクトルプログラムの直接実行

S-820システムでは，仮想記憶方式を前提とするTSSジョブからもベクトルプログラムを直接実行できる。これによって，ベクトルプログラムの開発，すなわちソースプログラムの編集，コンパイル，実行及びチューニングなどの一連の操作を



注：略語説明 JSS3(Job Spooling Subsystem 3)，JSS4(Job Spooling Subsystem 4)

図5 S-820システムの利用形態 S-820システムでは，バッチジョブとしても，対話処理としてもベクトルプログラムを実行できる。

TSS端末から対話的に実行でき，プログラム開発の生産性向上が図れる。

また，グラフィック端末などからもベクトルプログラムを実行できるため，画像処理などの分野にもスーパーコンピュータを適用できる(**図5**)。

3.3　主記憶装置の高効率制御

S-820システムの主記憶装置及び仮想記憶装置の対応関係を**図6**に示す。

主記憶装置は，VOS3/ES1のスカラー処理用として使用するスカラー主記憶装置と,ベクトル処理用として使用するHAP主記憶装置から成る。

すなわち，従来のユーザープログラム，JSS3(Job Spooling Subsystem 3)/JSS4などのサブシステム，及びオペレーティングシステムに対しては，スカラー主記憶装置を割り当てる。

HAP拡張リージョン(ベクトルプログラム用領域)に対しては，HAP主記憶装置を1Mバイト単位に割り当てる。主記憶の割当て単位を1Mバイトとすることによって，大容量の主記憶割当て及び解放に要するオーバヘッドの削減を図っている。

3.4　拡張記憶装置への高速スワッピング

ベクトルジョブは大規模な配列を扱うため，1ジョブ当たりの主記憶負荷が高く，ベクトルジョブの多重度は主記憶装置の容量に制限される可能性がある。

VOS3/HAP/ESでは，バッチ及びTSSから実行中のベクトルプログラムを主記憶装置から補助記憶装置に追い出すことによって，主記憶の負荷を自動的に調整する。この追い出し処理をスワッピングという。しかし，ベクトルジョブは大規模な主記憶装置を使用することが多く，ベクトルジョブを補助記憶装置に転送すると，チャネルのデータ転送ネックなどによってスワッピング時間が大幅に延びてしまう。また，チャネルを占有することによって，システム全体の入出力時間に影響を及ぼすことが予想される。

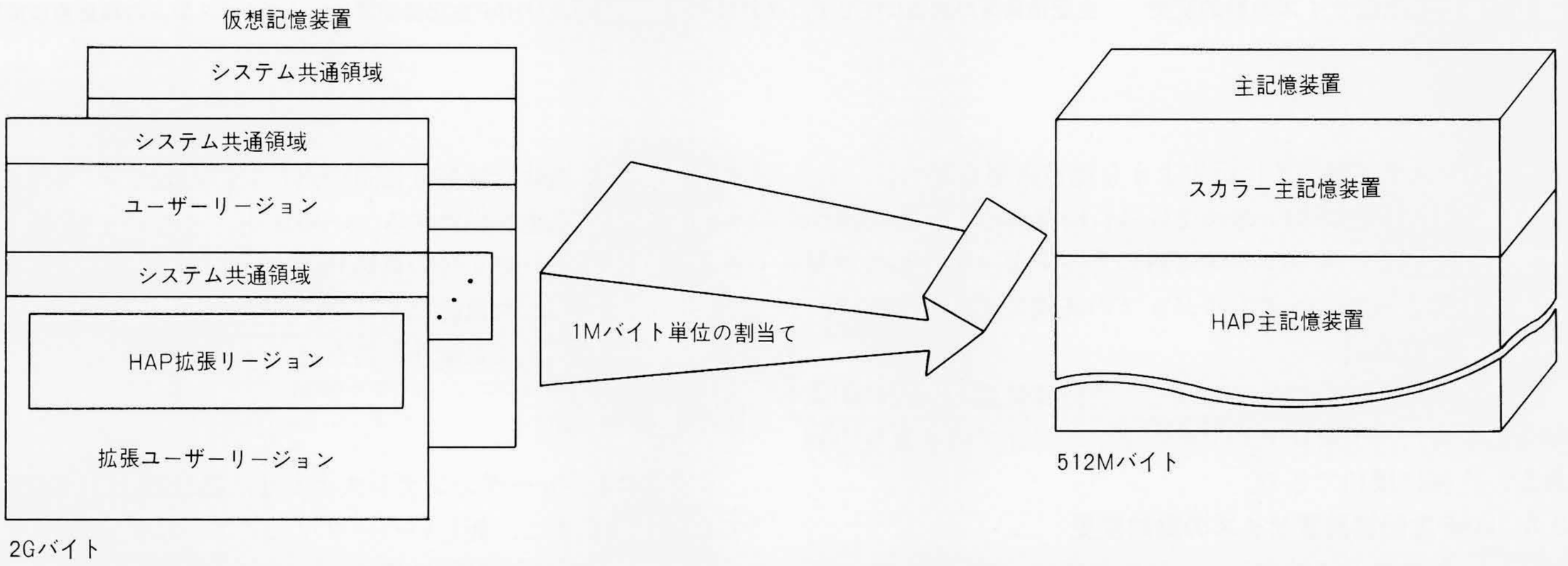


図6　主記憶装置の管理方式　主記憶装置は，VOS3/ES1のスカラー処理用のスカラー主記憶装置と，1Mバイト単位に割り当てるHAP主記憶装置(ベクトルプログラム用領域)に分割し，大容量な領域の割当てオーバヘッドを削減している。

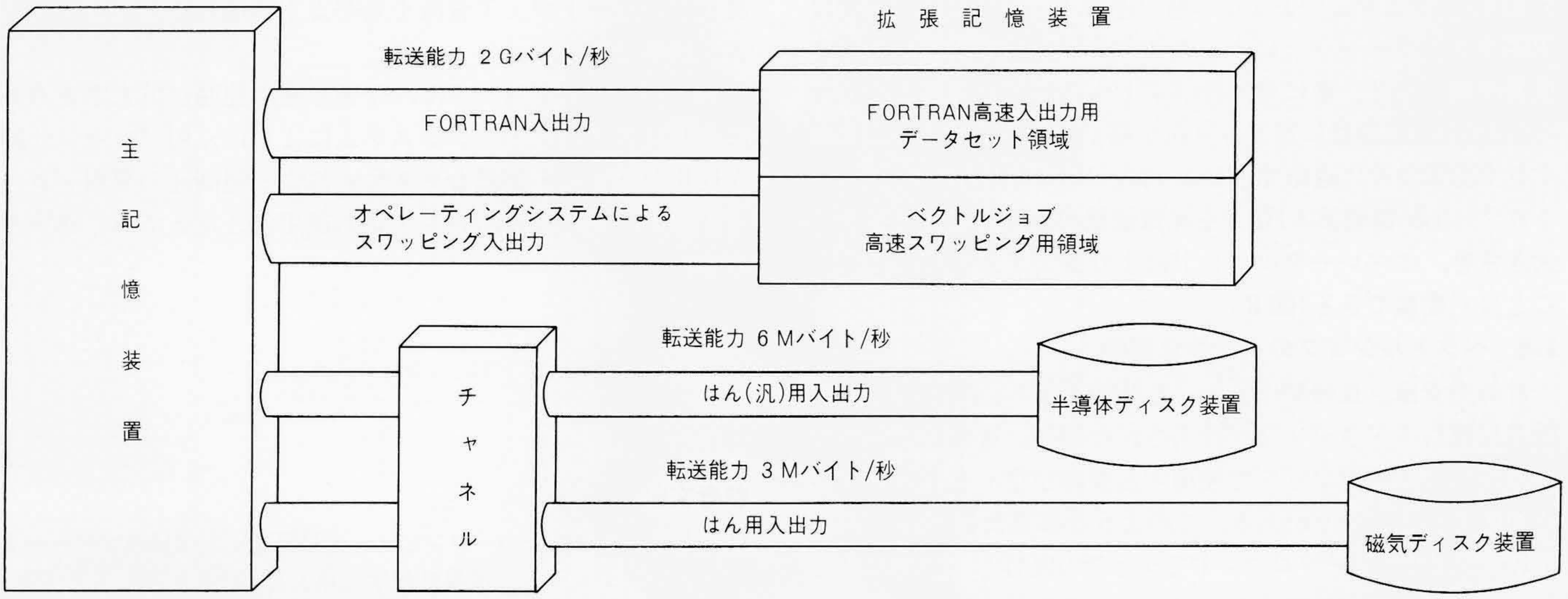


図7　S-820システムにおける記憶階層　S-820システムには，大容量・超高速の拡張記憶装置を接続できる。VOS3/HAP/ESでは,拡張記憶装置にFORTRANユーザーの一時的データセットとオペレーティングシステムのスワッピング用データセットを配置可能とし，入出力時間を大幅に短縮する。

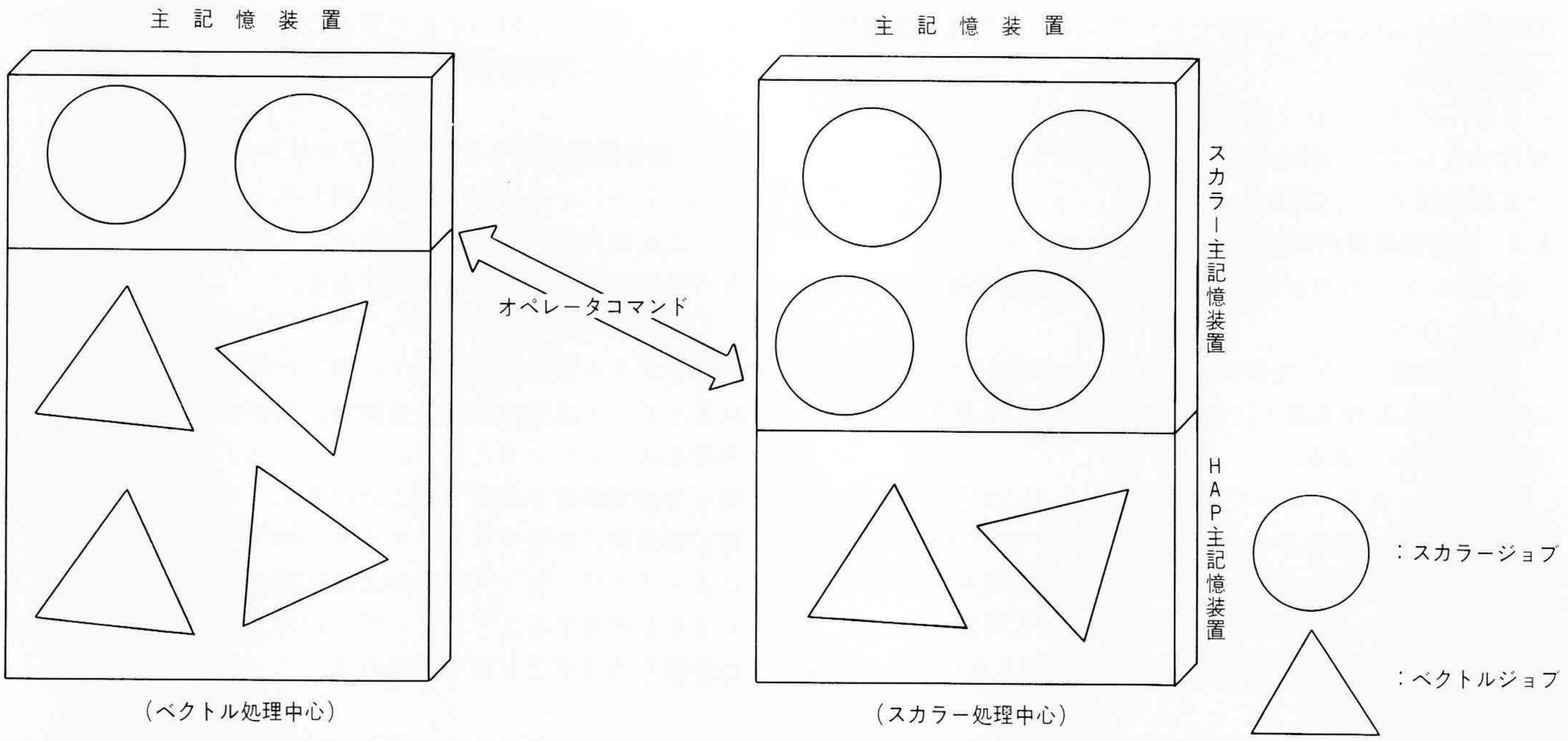


図8　主記憶装置サイズの動的変更　主記憶負荷の変動に応じて，スカラー主記憶装置及びHAP主記憶装置のサイズをシステムの稼動中に変更できる。

S-820システムは**図7**に示すような記憶階層を持つ。

VOS3/HAP/ESでは，最大2Gバイト/秒の転送能力を持つ拡張記憶装置を，スワッピング用のデータセットとして利用することによって，ベクトルジョブの高速なスワッピング入出力を実現する。

このスワッピング方式によって，主記憶装置の容量の制限から生ずるバッチ及びTSSでのベクトルジョブの多重度の制約が，大幅に緩和できる。

3.5　HAP主記憶装置サイズの動的変更

計算機システムに対するジョブの負荷は一定ではなく，その運用形態によって，1日の中でも大きく変動する。例えば，計算センタでの典型的な運用形態は，昼間はTSSや小規模ジョブを主体とし，夜間は大規模ジョブの実行が主体となる。S-820システムをこのような形態で運用する場合には，昼間はTSSやスカラージョブによってスカラー主記憶装置に対する負荷が高くなり，逆に夜間はベクトルジョブが主体となるためHAP主記憶装置に対する負荷が高くなる。このように変動する主記憶負荷に適応できるように，VOS3/HAP/ESでは，スカラー主記憶装置とHAP主記憶装置の二つの主記憶装置の大きさを，オペレータコマンドによりシステムを停止させることなく変更できる(**図8**)。

3.6　ベクトルジョブの一時休止機能

計算機資源を長時間にわたって占有するジョブの実行中に，緊急に実行すべきジョブが投入された場合，資源に余裕がない状況では，緊急ジョブで使用する資源が空くまで実行が待たされてしまうことがある。このような事態に備えて，VOS3/ES1ではオペレータコマンドでジョブの実行を一時的に休止させる機能を提供していた。

VOS3/HAP/ESでも，実行中のベクトルジョブを一時休止させ，ベクトルジョブが使用している主記憶及びHAP拡張記憶を強制的に補助記憶装置に追い出して，緊急ジョブの実行を優先させる機能を提供する。一時休止させるジョブの選択及び一時休止したジョブの再実行は，オペレータコマンドにより指定する。この機能によってシステムの運用，ジョブのスケジューリングに柔軟性を持たせている。

4　結　　言

S-820のオペレーティングシステムは，従来のS-810で培ってきた技術を基に，新しいハードウェアアーキテクチャを取り入れて開発した。対話処理でのベクトルプログラムの直接実行機能や，従来からの31ビットアドレッシング方式によって，ベクトルプログラム開発の生産性向上が図れる。また，高速な命令プロセッサ，拡張チャネルシステム，拡張記憶装置などのハードウェア資源を効率よく利用した高性能システムを構築できる。

スーパーコンピュータは，今後も適用範囲が拡大する方向にあり，スタンドアロンシステムによるはん用プロセッサ的な運用形態から，疎結合システムによる超高速の専用プロセッサまで，更に幅の広いシステム運用にこたえていく必要がある。

参考文献

1)　大房，外：スーパーコンピュータHITAC S-810のオペレーティングシステム，情報処理学会第33回全国大会　講演論文集(I)，319～320，(昭61-10)
2)　日立製作所：スーパーコンピュータHITAC S-820機能説明書，6020-2-002(昭62-6)