

システム記憶と複合計算機システムSCMP について

金子 雅道

富士通 情報システム事業本部ソフトウェア事業部

近年の高度情報化社会の進展により、大型計算機システムに対する性能、信頼性の向上へのニーズはますます高まっている。本論文では、このようなニーズに対応するために、富士通が新しく開発した半導体記憶装置『システム記憶』、および、このシステム記憶を介して結合された新しい概念の複合計算機システム『SCMP™』について述べる。データベースやSORT, FORTRAN の一時的作業ファイルなどをシステム記憶へ展開することにより、データ処理や障害発生後の回復処理をより高速に行うことができる。また、SCMPを利用することによって、秒オーダーの高速ホットスタンバイを実行する高信頼性システムや高速処理システムを構築することが出来る。

SYSTEM STORAGE AND SYSTEM STORAGE COUPLED MULTIPLE PROCESSOR: SCMP

Masamichi Kaneko

Software division, Computer Systems Group, FUJITSU

3-3-1 Shinyokohama Kouhoku-ku, Yokohama-shi, 222 Japan

A large computer system is needed high-performance and high-reliability. This report describes "System Storage" and the "System storage Coupled Multiple Processor :SCMP™" that are developed by FUJITSU for resolved these problem.

System storage is new type semi-conductor storage unit. Loading into system storage with DATABASE, temporary file of SORT, FORTRAN, etc, the computer system can achieve more high speed data processing and recovery processing.

And using SCMP, it is possible to build a high performance system and a high reliable system such as a hot-stand-by system that exchangeing in a few seconds.

1 まえがき

近年の高度情報化社会の進展により、「より大量に」、「より高速に」、「より安全に」データ処理を行うことが、大型計算機システムに対してより強く求められている。

富士通では、このような社会的ニーズに応えるために各種の製品群を顧客殿へ提供しているが、より一層のサービスを図るため、新しい半導体記憶装置『システム記憶(SS: System Storage)』を開発し、大型計算機Mシリーズ用の新OS『MSP-EX』でサポートした。このシステム記憶上へデータベースなどを展開することにより、計算機システムの処理能力、信頼性が一段と向上する。

また、システム記憶を介した新しい概念の複合計算機システムSCMP™(System storage Coupled Multiple Processor)を利用することによって、高信頼性、高処理能力のシステムを構築することができる。

2 背景

従来、計算機システムの処理能力や信頼性を向上させるために、CPU、主記憶装置、外部記憶装置などの処理能力／信頼性向上や、容量の増設、などの計算機システムを構成する各種機器や素子の進歩は当然のこととして、次のようなシステムの的なアプローチが行われてきた。

- 処理能力向上
 - ・ 仮想記憶空間の大きさの拡大
 - ・ 計算機システムの複合化による全体的な処理性能向上
 - ・ 外部記憶装置内のデータを主記憶へ常駐することなどによるI/O処理の改善
- 高信頼性化
 - ・ 計算機システムの複合化による全体的な信頼性向上

このようなアプローチによって、ユーザニーズの多くをカバーしてきたが、しかしながら、各技術とも、次のような課題が残されている。

- TCMP構成
 - ・ CPU同士が主記憶を介して密に接続されているため、ソフトウェア障害やメモリ障害に対するシステム保護を完全には行うことができない。
 - ・ 接続可能なCPU台数には限界がある。
- I/O処理の改善
 - ・ 外部記憶装置内のデータを主記憶へ常駐化するのみでは、主記憶の揮発性という制約から、DASD性能とCPU性能のアンバランスを完全に解消することはできない。
- 複合システム
 - ・ DUALシステムにおいては、他系計算機システムのダウン検出、ダウン後の業務引き継ぎなどを共用DASDや通信回線を經由しているため、高速化に限界がある。
 - ・ DUPLEXシステムにおいては、共用DASDに対する競合が発生し、複合システム全体としての性能は頭うちになる。

上記のような技術的問題点を根本的に解決したのが、システム記憶とSCMPである。

3 システム記憶とSCMP

システム全体の処理能力と信頼性を向上させるため、「システム記憶」という複数の計算機システム間で共用可能な半導体記憶装置を開発した。システム記憶の特徴は、

- ・ システム記憶上のデータに対する128テラバイトの多重仮想空間
- ・ 半導体ディスク装置等の外部記憶装置に比べて、より高性能なアクセス
- ・ 独立した複数の匡体(SSU)によるデータの二重化
- ・ バッテリバックアップによる不揮発性の実現
- ・ 複数の計算機システムからの共用、および、システム記憶を介した通信機構である。

以下に、各特徴の詳細を述べる。

(1) システム記憶の空間

システム記憶には、主記憶空間とは独立したSS実空間およびSS仮想空間の2種類の空間が存在する。これらの空間は、いずれも47ビットでアドレス表現され、128テラバイトの大きさを持っている。

SS実空間は、システム記憶の物理的な記憶アドレスを示したものであり、システム記憶装置に接続されているいずれの計算機システムからも同一のものである。

また、SS仮想空間は、各計算機システムの主記憶上に構築されているSSアドレス変換テーブルによって、SS実空間に位置づけられるものである。OSは、主記憶の仮想空間毎にSSアドレス変換テーブルを用意することにより、主記憶の仮想空間と1対1でSS仮想空間を生成している。そして、主記憶の仮想空間を切り替える際に、使用するSSアドレス変換テーブルもOSが自動的に切り替えることによって、主記憶の仮想空間とSS仮想空間との対応関係を維持している。このため、システム記憶を使用する処理プログラム群は、OSを介することなくシステム記憶へ直接アクセスすることができる。また、主記憶上の仮想空間同士がシステム記憶上の同一データ域を共用要求した場合は、各SS仮想空間から同一のSS実アドレスをポイントするようにSSアドレス変換テーブルをOSが作成することによって、SS上のデータ共用を実現している。さらに、OS自身が保持しているシステム記憶上の管理データをアクセスするために、OS専用のSS記憶空間も創成している。

このように、システム記憶を利用するプログラムがシステム記憶上のデータを直接アクセスすることによって、OSの介在によるオーバーヘッドを削減し、また、SS仮想空間を完全に分離することによってシステム記憶上のデータの破壊防止と機密保護を実現している。

図1にシステム記憶の多重仮想空間を示す。

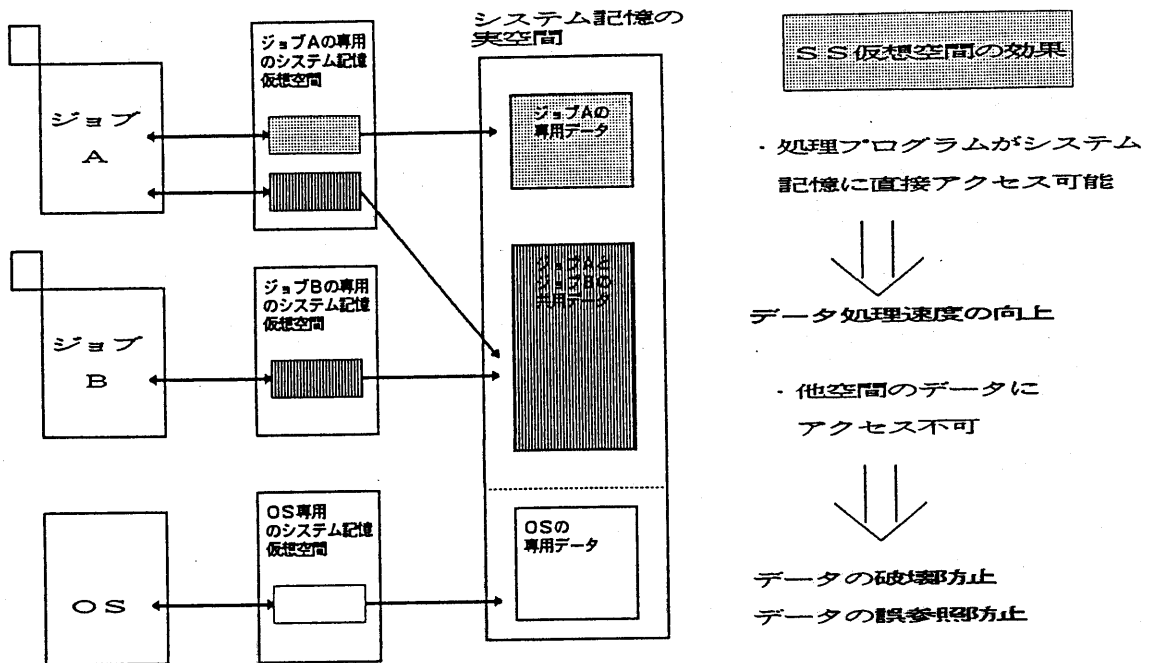


図1 システム記憶の多重仮想空間

(2) システム記憶へのアクセス

システム記憶へのアクセスは、専用のCPU命令によって主記憶の仮想空間とSS仮想空間との間を高速に行われる。

SS仮想空間上のデータを更新/参照する際には、主記憶と同様の保護機構が働き、プログラムミスによるデータの破壊や悪意のユーザプログラムによる機密データの読み込みを防止している。

各プログラムからのシステム記憶へのアクセスは、当社のオンラインデータベースシステムAIMやFORTRANライブラリなどのサービスプログラムが自動的に行うので、ユーザのアプリケーションプログラムがシステム記憶を考慮する必要はない。

(3) 二重化機構

ハードウェアオプションによって、複数のシステム記憶装置を二重化して使用することができる。

システム記憶へデータを格納する際には、二重化されている両方のシステム記憶へハードウェアが自動的にデータを書き込むため、アクセスするプログラムが意識する必要はない。また、二重化されているシステム記憶の何れからデータを読み込むかの決定もOSとハードウェア機構とが自動的に制御している。

二重化されているシステム記憶装置の一方に何らかの障害が発生した場合、OSは当該システム記憶装置を自動的に切り離して、片肺運転で業務を続行する。計算機システム、システム記憶、およびDASDをそれぞれ二重化することによって、ノンストップシステムを構築することができる。

図2に、システム記憶の二重化によるノンストップシステムの構築を示す。

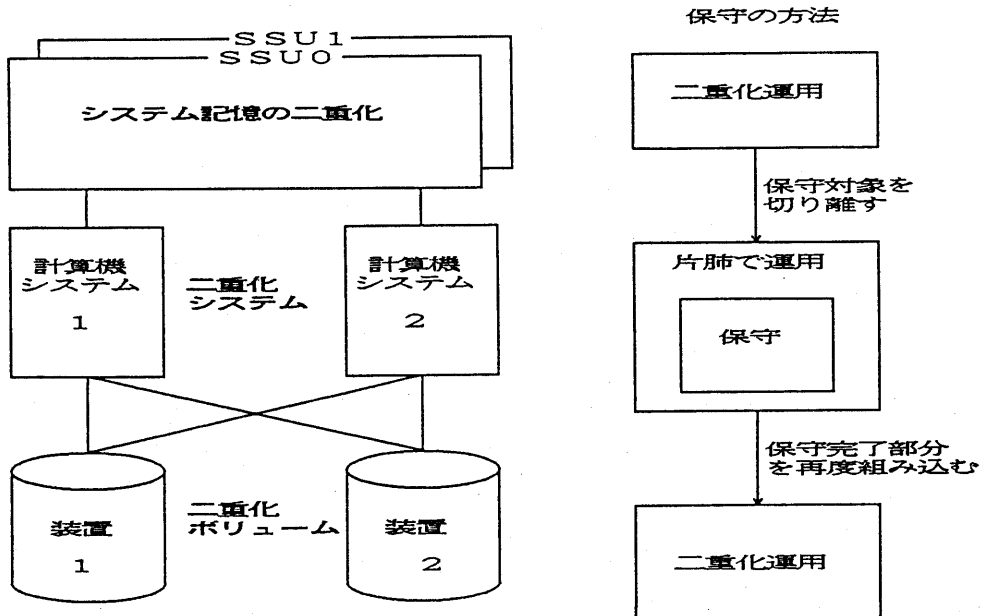


図2 二重化機構によるノンストップシステムの構築

(4) バッテリバックアップ機構

システム記憶装置にバッテリバックアップ機構を装備することにより、電源停止後最大24時間の不揮発性が保証される。停電などの電源系異常が発生しても、システム記憶上のデータは保存・再使用することができるので、揮発性の主記憶には置けなかったオンライントランザクション処理のログデータなどをシステム記憶へ展開することができる。

(5) システム間共用

複数の計算機システム同士が、同一のシステム記憶を共用することができる。また、システム記憶を介して、複数の計算機システム間で通信を行うことができる。各計算機システムは、このシステム間共用機構、システム間通信機構を介して、データ処理の負荷分散、他計算機の状況把握、業務依頼を行うことができる。

図3、および、表1にシステム記憶と従来の記憶装置との比較を示す。

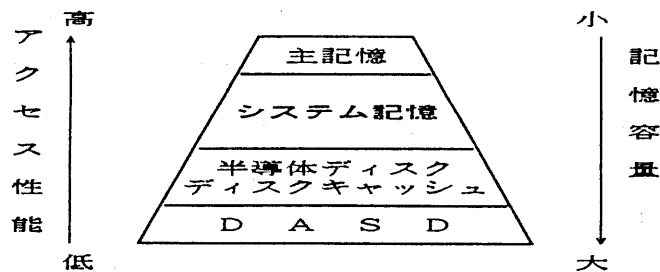


図3 システム記憶による新たな記憶階層の構築

記憶装置種別	アクセス時の CPU オーバーヘッド	システム間 の データ共用	IPLに跨 がった データ保全	電源障害時 の データ保全	二重化
主記憶	小	×	×	×	×
システム記憶	小	○	○	○ (オプション)	○ (オプション)
半導体ディスク ディスクキャッシュ	大	○	○	○ (オプション)	○
DASD	大	○	○	○	○

表1 システム記憶と従来の記憶装置との比較

「SCMP」とは、システム記憶を共用することによって構築された新しい形態の複合計算機システムのことである。システム記憶に接続されている各計算機システムは、共用DASDで接続された従来の複合システムと同様に独立しており、また、システム記憶を共用することによって従来の共用DASDによる複合システムを越えた機能を発揮することができる。例えば、計算機システム間の共用データをシステム記憶上に常駐することにより、計算機システム間のデータの引き継ぎを高速に行うことができる。また、システム間通信機構を用いることにより、計算機システム間の同期処理や他系計算機システムへの処理依頼を能動的かつ高速に行うことができる。

各計算機システムがシステム記憶へアクセスする際は、アクセス対象のデータ領域のみを逐次化すればよいので、共用DASDのような装置全体の逐次化による競合は起きない。また、アクセスそのものもDASDに比べて高速なので、システム記憶上の同一領域の逐次化の競合もほとんど発生せず、SCMP構成を採ったことによる他計算機システムからの悪影響によるオーバーヘッドは起こり得ない。

図4にシステム記憶のシステム構成を、また、表2にシステム記憶導入の効果を示す。

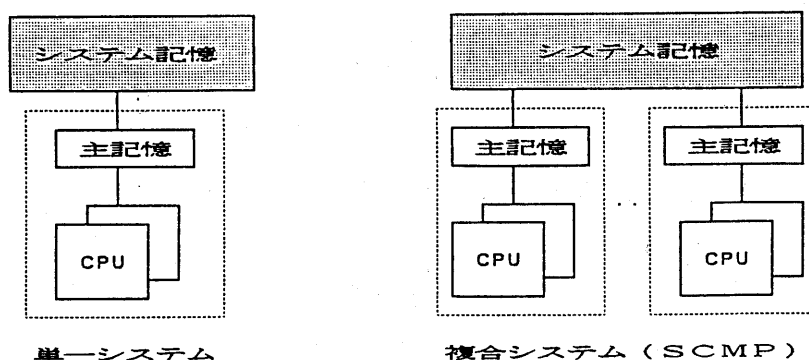


図4 システム記憶によるシステム構成

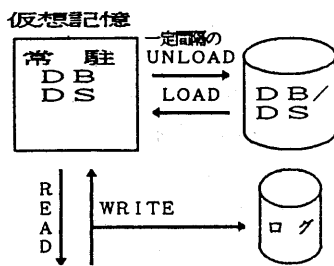
単一システム	システム記憶導入の効果	SCMPシステム
○	入出力性能の向上による ・トランザクション処理性能向上 ・レスポンスタイム短縮 ・スループットの向上 ・ターンアラウンドの短縮	○
○	システムダウン後の高速再立ち上げ	○
—	クラスタ多重化によるスループットの向上	○
—	業務の高速引き継ぎ	○
—	システム多重化によるノーダウン化	○

表2 システム記憶導入の効果

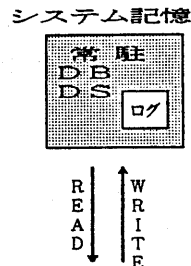
4 システム記憶の適用例

システム記憶を利用するプログラムは、システム記憶上のアドレス、主記憶上のアドレスおよび転送長を指定して、専用のCPU命令を発行することにより、システム記憶を自由にアクセスすることができる。すなわち、システム記憶は、任意のバイトに直接位置づけをすることができる新しい概念のデータセットを格納するための媒体として考えることができる。また、システム記憶と主記憶との間のデータ転送は、スーパーバイザを介することなく処理プログラムが直接行うため、DASDを使用した場合に比べて、性能が大幅に向上する。図5に、データベースと更新ログデータをシステム記憶上に展開した例を示す。本例では、DASD上に展開されていたデータベースの内容を、システム記憶上に常駐してデータを参照／更新するものである。また、ログデータもシステム記憶上に展開し、DASDへの書き込みは行わず、ダウン後のリカバリ処理は、システム記憶上のログデータをそのまま用いる。このような適用により、①I/O時間の削減 ②ダウン後のデータリカバリの高速化を図ることができる。

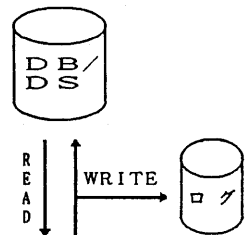
仮想記憶常駐化（従来）



システム記憶常駐化



DASD直接利用



仮想記憶	常驻化の効果	システム記憶	DASD
◎	CPUオーバーヘッド	◎	×
○	入出力の削減	◎	×
△	ダウン時のリカバリ	◎	○
×	システム間共用	◎	◎

図5 データベースのシステム記憶常驻化

5 SCMP構成の適用例

SCMPとは独立した計算機システム同士が、システム記憶という通信機能を備えた共用記憶装置によって結合されたものである。各計算機システムは、システム記憶を介して、他系の計算機システムの状態把握や業務依頼を能動的に行うことができるので、業務の指示や引き継ぎを迅速に行うことができる。図6に、SCMP構成による高速ホットスタンバイシステムの例を示す。

本例では、システム記憶上に「システム監視表」、「データベース」、および「ログデータ」を展開し、各計算機システムが共用したものである。各計算機システムは、システム監視表に自システムの状況を定期的に書き込み、かつ、他システムの状況を把握する。もし、他計算機システムからの書き込みが途絶えていれば、当該計算機システムはダウンしているのを、業務の引き継ぎをする。また、自計算機システム内にトラブルがあって、業務の続行が不可能になったことを認識した場合は、システム間通信機構を用いて、他計算機システムへ業務の引き継ぎを依頼する。業務の引き継ぎ処理においては、システム記憶上のログデータを使用するため、高速かつ確実にデータベースを復旧することができる。

このように、システムダウンの認識およびデータリカバリをすべてシステム記憶上で行えるので、秒オーダでのホットスタンバイシステムを構築することができる。

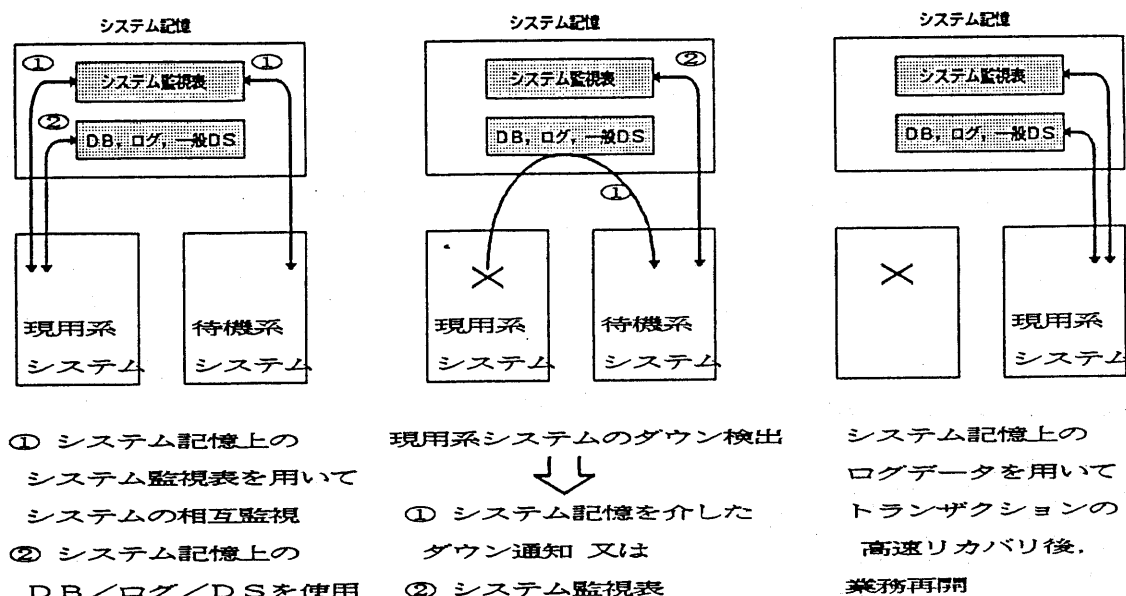


図6 SCMP構成によるホットスタンバイシステム

6. 終わりに

ワークステーションの高性能/高機能化、顧客の情報処理量の激増とオンライン化による社会的影響など、情報処理システムを取り巻く環境は急激に変化している。こうした中でホストコンピュータは大量のデータを一括処理するデータサーバの役割を主に担うようになり、より大きな処理能力を持ち、しかもより高い信頼性を持つシステムが求められるであろう。今回当社が開発したシステム記憶及びそれを介して接続される複合計算機システムSCMPはこうしたニーズに応える有力な方法である。今後は更に操作、監視等の複合システム制御機能の充実やシステム記憶の適用範囲の拡大に努めていきたい。