

スーパーコンピュータ SX-7 ハードウェア

日本電気株式会社 コンピュータ事業部 西川 岳

1. まえがき

1998年1月に東北大学に納入されたSX-4の後継機であるSX-7は、SXシリーズの最新機種で、SX-4で開発されたソフトウェア資産の移行性を容易にする32CPUの共有メモリ方式のアーキテクチャを継承したスケーラビリティと使いやすさに優れたスーパーコンピュータです。

2. SX-7システムのハードウェア構成と特長

2.1 ハードウェア特徴

SX-7の主な特徴は以下の通りです。

CPUの最大性能が8.83GFLOPSの世界最速の1チップベクトルプロセッサを使用しています。

記憶素子に高速大容量の256Mビット・DDR-SDRAMを採用し、高いメモリスループットと大容量化を実現しています。

CPUのスカラユニットにおける命令処理方式は、4way スーパースカラを採用しており、スカラ最大性能値は1.1GFLOPSを実現しています。

SX-4同様の32個のベクトルプロセッサ(CPU)による共有メモリ方式のアーキテクチャの継承をしており、大規模な共有メモリシステムを実現しています。線幅0.15ミクロン(μm)デザインルールを用いた超高速高集積CMOS LSI技術により、CPUの1チップ化を実現しています。

大規模演算に不可欠なメモリからの高速データ供給能力を超高速クロスバスイッチにより実現し、実プログラムでの超高速演算を可能にしています。

以下に、上記特長を備えたSX-7システムのアーキテクチャ、システム構成、そしてテクノロジーの概要について紹介します。

2.2 ハードウェア構成

SX-7システムは、1ノード32CPUで構成されており、これらはGigabit Ethernetスイッチ(日本電気製 Catalyst6509、1000BASE-SX)により接続されています。システム全体の理論最大性能値は2048.56GFLOPSとなり、SX-4の性能値である256GFLOPSと比較すると、約8倍の理論最大性能となります。1CPUの性能におい

ても、SX-7の1CPUにおける理論最大性能値は8.83GFLOPSで、SX-4の性能値である2GFLOPSと比較すると約4倍の性能を達成しています。

表1にSX-7の主要諸元についてSX-4との比較を示します。

表 1 SX-7システムとSX 4システムの主要諸元比較

項目		SX-7	SX-4
理論最大演算性能		2048.56GFLOPS	256GFLOPS
CPU数		232	128
主記憶装置	容量	1792GB	32GB
	最大総合データ供給能力	8189.6GB/s	2048GB/s
	使用記憶素子	256Mb SDRAM	4Mb SSRAM

次に、SX-7システムを構成しているノードのブロック図を図1に、主要諸元についてのSX-4との比較を表2に示します。

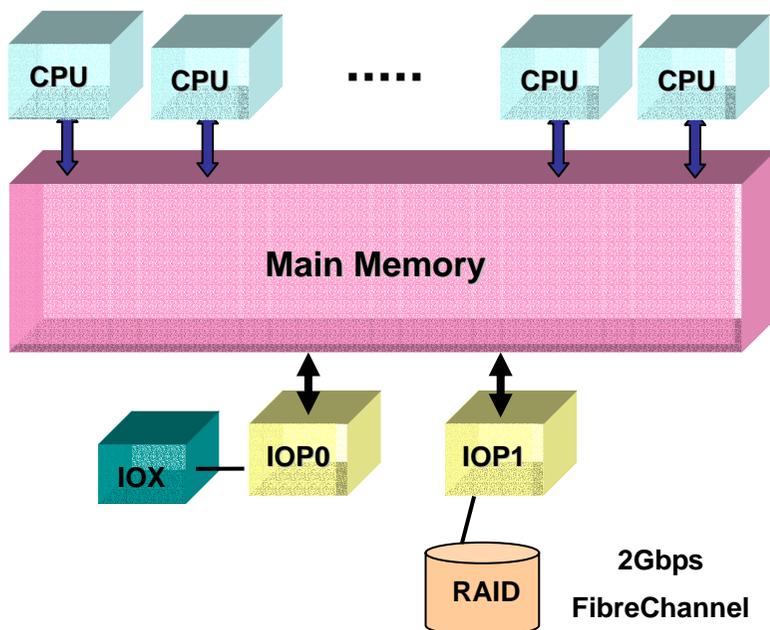


図1 SX-7システムのノード基本構成

ノードは、32個のCPUを持つことが出来、256GB(1ノード)の主記憶装置で構成されています。これらのノード内ではSX-4同様に共有メモリの特性を生かした並列処理が可能であり、クロックサイクル当たり512個の浮動小数点演算(256個の浮動小数点加算と256個の浮動小数点乗算)を実行することにより、282.56GFLOPSの理論最大性能を実現しています。

また、SX-7の1ノードにおける設置面積(保守エリアを含む)と消費電力をSX-4と比較すると、設置面積は1ノード44.5m²のSX-4に対して、SX-7は約17.3m²、消費電力も1ノード当たりSX-4は78.0KVAであるのに対し、SX-7では28.0KVAとなっています。設置面積では約1/2.5、消費電力では約1/2.7と両者とも大きく改善されました。これらの改善は、1CPUのチップ数を37個から1個へ導いた、最先端のCMOS LSI設計技術と実装技術によって実現しました。

表2 SX-7とSX-4を構成するノードの主要緒元比較

項目		SX-7	SX-4
理論最大演算性能		282.56GFLOPS	64GFLOPS
CPU数		32	32
CPU	レジスタ	ベクトルレジスタ	144KB
		ベクトルマスクレジスタ	256bit×16
		スカラレジスタ	64bit×128
	データ形式	固定小数点	32/64bit
		浮動小数点*	32/64/128**bit
		論理	IEEE / IBM / CRAY
	ベクトル演算パイプライン	64bit	64bit
	ベクトル演算パイプライン	5種類×8セット	5種類×8セット
	スカラ演算パイプライン	1セット	1セット
	キャッシュメモリ	命令：64KB オペランド：64KB	命令：64KB オペランド：64KB
主記憶装置	容量	256GB	8GB
	最大データ転送能力	1130.24GB/s	512GB/s

*) SX-7では国際標準であるIEEE754規格のみサポートします。

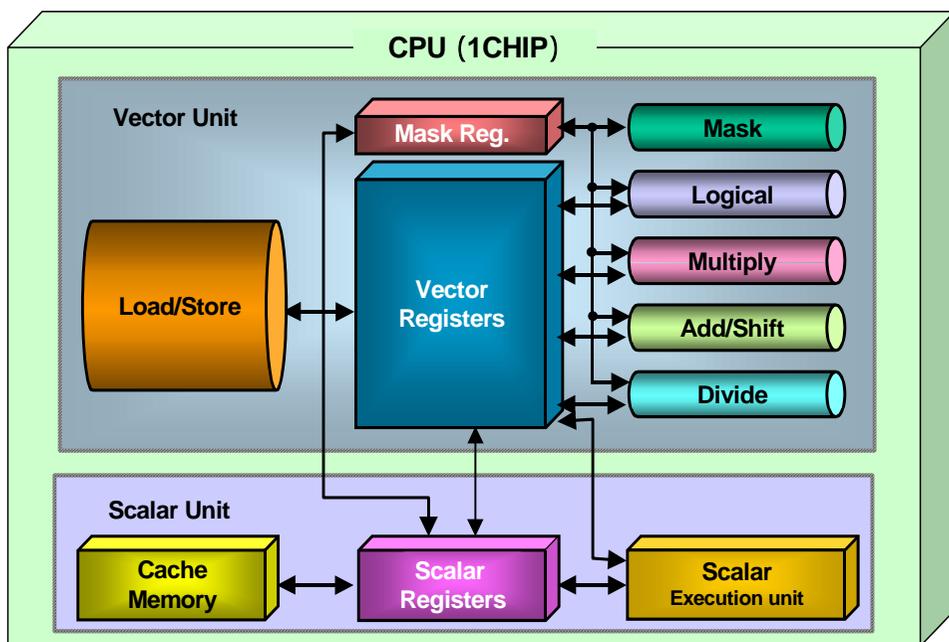
**) 128bitは、スカラ命令のみサポートします。

次節以降に、SX-4から強化された点を中心にSX-7のシステムのノードのハードウェアについて説明します。

2.3 CPU

SX-7 のシステムの心臓部である CPU の内部構造を図 2 に示します。

CPU は機能的にスカラユニットとベクトルユニットに分けられます。スカラユニットは、命令の読み出し、解釈、実行制御を行うと共に、スカラ演算パイプラインを有しており、ベクトル演算パイプラインと同時並行にスカラ演算を実行します。ベクトルユニットは CPU あたり 8 セットのベクトル演算パイプラインが単一命令ストリーム / 複数データストリーム (SIMD) 型の並列処理を実行しており、ベクトル演算を高速実行します。

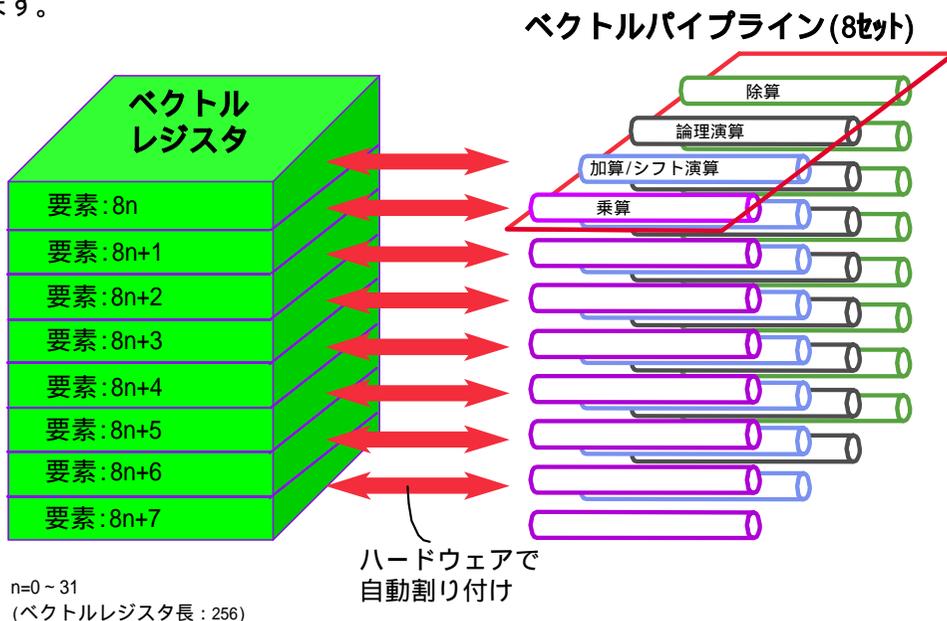


CPUの内部構造

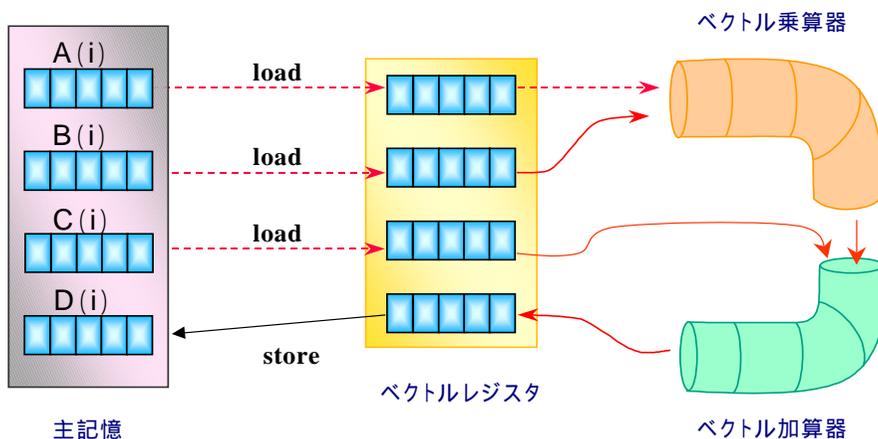
SX-7 システムの CPU の第 1 の特長は、5 多重 8 並列のベクトル演算パイプラインにより構成されているベクトルユニットです。SX-7 の CPU は、ベクトルパイプラインとして加減算 / シフト、乗算、除算、論理演算、マスク演算の 5 種 (5 多重) を有しています。5 種類のベクトル演算パイプラインは独立して動作が可能であり、データの依存関係が無ければ、5 種の演算が同時に実行できます。さらに、各演算用のパイプラインは 8 個 (8 並列) が用意され、1 つのベクトル演算命令により 8 個のベ

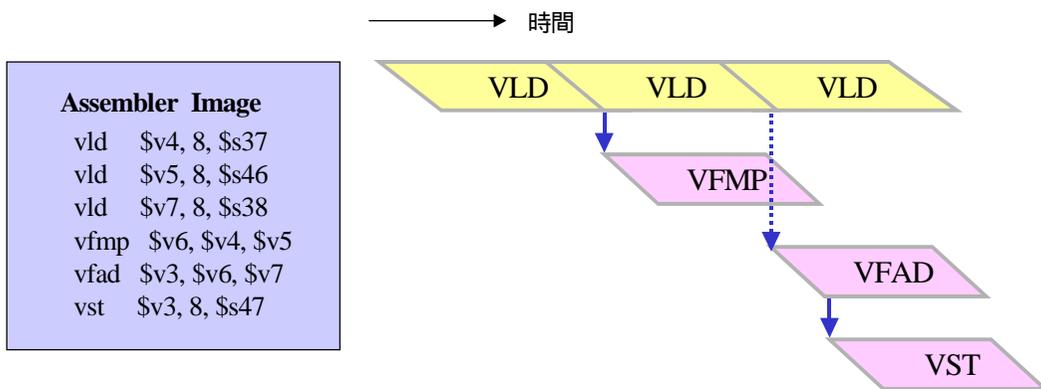
ベクトル演算のパイプラインが SIMD 型の並列処理を行います。図 3 にベクトルレジスタに格納されるデータ要素番号と 8 並列のベクトル演算パイプラインの関係を、図 4 に簡単な演算のベクトル処理の動作（チェイニング動作）、アセンブライメージ、そしてタイムチャートを示します。

第 2 の特長は最先端の超高速・高集積の CMOS テクノロジと最先端 LSI 設計技術により、SX-4 では 37 個の LSI で構成されていたベクトルプロセッサ(スカラーユニット+ベクトルユニット)をたった 1 つのチップで実現できたことです。この LSI 技術とパッケージング技術により、1 チップで 8.83GFLOPS という演算性能を実現しています。



SX-7のベクトルユニットの構成





$D(I) = A(I) * B(I) + C(I)$ におけるベクトルパイプラインチェイニング動作

第3の特長は、4way スーパースカラ方式を採用しているスカラユニットの性能向上です。SX-7 のスカラユニットの命令パイプラインは、最大4個の命令を同時に解読し、命令の処理に使用する演算器、レジスタなどのハードウェア資源の使用状況を監視することにより、最大4個の命令の処理を同時に起動することが可能です。SX-7 のスカラピーク性能は 1.1GFLOPS で、スカラピーク性能が 250MFLOPS である SX-4 の約4倍に向上しているため、実アプリケーションでの高い実効性能が期待できます。

第4のCPU（及び共有メモリノード）の特長として、高いメモリスループットが挙げられます。ベクトル機はキャッシュに依存しない分、メモリからの強力なデータ供給能力が性能を握る鍵となります。超高速クロスバススイッチにより、SX-4 においては 16GB/s であった1CPU におけるスループットは、SX-7 においては 35.3GB/s にまで向上しました。図5は、CPU台数と主記憶のデータ供給能力を評価するベンチマークプログラム”Stream”における性能比較グラフです。

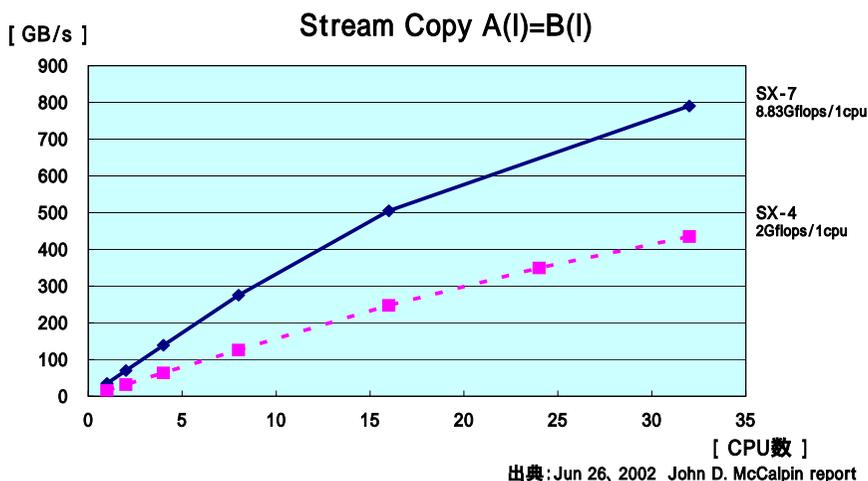


図5 データ供給能力 (Stream ベンチマーク)

これらの特長を備えた SX-7 は、ベンチマークプログラムである Linpack の HPC (次元数に制限を設けず、システムの持つ CPU / メモリなどの資源を有効に使用した場合の並列処理性能を評価することを目的とするプログラム) において高い性能を出しています。図 6 は HPC における性能比較グラフです。

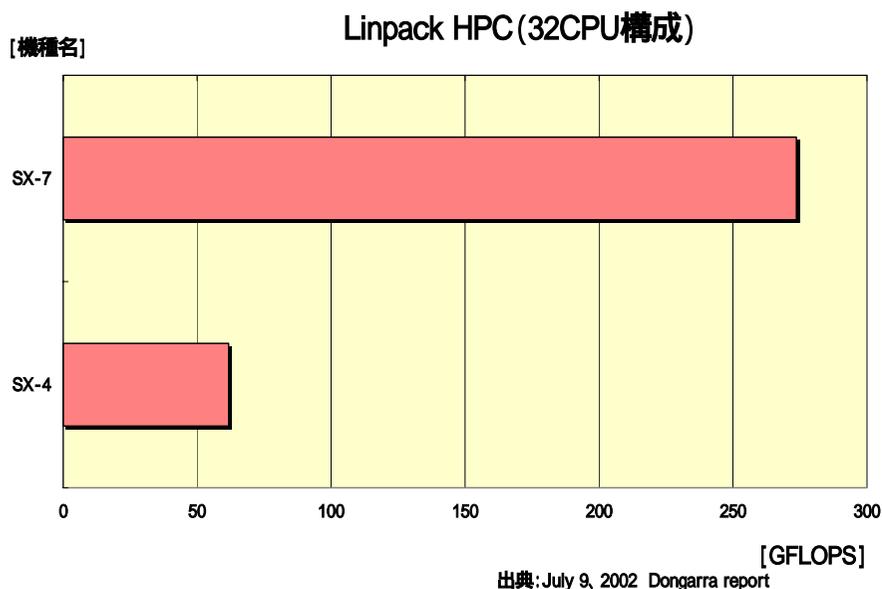


図 6 ベンチマークプログラム Linpack の HPC における性能比較

以上、SX-7 システムの CPU の主な特長について述べました。この他にも、SX-4 システムで採用されていた高速演算手法 (例えば、大容量スカラレジスタ、RISC アーキテクチャの命令体系、パイプラインの連結 (チェイニング) 機能など) は SX-7 でも引き続き採用され、高速演算に貢献しています。

2.4 主記憶装置

スーパーコンピュータにおいて演算能力が高くなければならないことは当然ですが、演算器へのデータ供給能力も演算能力に見合った強力な能力が必要となります。

SX-7 システムの主記憶装置は、この要求を満足させるために SX-4 と同様に CPU と主記憶装置間のデータバス制御部に分散制御方式を採用し、データを格納するデータ記憶部に独立動作可能な複数のバンクからなる多バンク構成を採用しています。表 3 に SX-7 を構成するノードの主記憶装置が有するデータ供給能力について SX-4 との比較を示します。

表 3 主記憶装置のデータ供給能力

項目	SX-7	SX-4
主記憶装置容量	256 G バイト	8 G バイト
インタリーブ数	16384	1024
CPU 当たりのデータ供給能力	35.32 G バイト / 秒	16 G バイト / 秒
最大データ供給能力	1130.24 G バイト / 秒	512 G バイト / 秒

CPU と主記憶装置間のデータバス制御部は分割され、CPU と主記憶装置に分散しています。この結果、全ての CPU は専用のデータバス制御部を有し、CPU あたり 35.32 G バイト / 秒のデータ供給能力を実現しています。CPU の数が増えると、データバス制御の数も増えることにより、32 台の CPU から構成されるノードでは、1130.24 G バイト / 秒の総合データ供給能力となります。すなわち、共有メモリ型マルチプロセッサ構成であるノードにおいて、マシンサイクルあたり 256 要素の同時アクセスを可能とするデータ供給能力を実現しています。

CPU はマシンサイクルあたり最大 8 要素の同時アクセスを可能としています。アクセスしたデータは、次に説明する主記憶装置内のバンク競合などにより、アクセス要求した時と異なる順序で主記憶装置から送られてきます。CPU 側のメモリアクセス制御部は、アクセスしたデータを 1 つのベクトル命令で処理する単位 (最大 256 要素) 毎に、マシンサイクルあたり 8 要素のデータ列として、アクセス要求した時の順番に整列し、ベクトル演算パイプラインにデータが連続して供給できるようにしています。

32 個の CPU が同時に主記憶をアクセスするケースにおけるデータ記憶部のアクセス競合対策として、データ記憶部は 16384 個の独立動作可能なバンクからなる多バンク構成としています。さらに、DDR (Double-Data-Rate) 型メモリ素子の採用によ

り、メモリのスループット性能を向上させており、マルチプロセッサ構成時でも主記憶装置からのデータ供給能力を十分に確保しています。

演算能力が高くなると単位時間に処理するデータ量が増加するため、主記憶容量に対する要求も増大します。この要求に応えるため、SX-7 システムではノードあたり 256 G バイト、システムとしては 1792 G バイトの総合主記憶容量を提供しています。この主記憶容量の拡大により、大規模配列を主記憶上に展開することが可能となり、プログラムの実行時間を大幅に短縮することを可能としています。

また、主記憶装置は入出力装置とも接続され、システムの CPU 間の通信・ノード間のデータ通信・入出力装置データ転送における高速処理を実現しています。

2.5 入出力装置

システムの演算性能の向上に伴い、大容量のストレージや高速なネットワークが必要となってきます。また、多様な周辺機器を接続するために標準インターフェースが求められます。

SX-7 では、ストレージ・インターフェースとして Ultra3-SCSI および ANSI 準拠の高速インターフェース FibreChannel を、ネットワークインターフェースとして Ethernet および GigabitEthernet を提供し、これらの要求に応えます。特に SX-7 では、2Gbps の FibreChannel をサポートする入出力装置を新規に開発し、同インターフェースを有するディスクアレイ装置を接続しています。高速ディスクアレイ装置は、装置あたり 572.8G バイトの容量を持ち、大容量ファイルの高速アクセスを可能とするとともに、演算能力が高くなったことにより、ますます増加することが予想されるデータの格納容量を拡大しています。

3. SX-7 のテクノロジー

3.1 LSI 技術

表 4 に SX-7 及び SX-4 に使用している LSI の諸元比較を、図 7 にスーパーコンピュータのクロックサイクルを示します。LSI 設計においては、バイポーラテクノロジーによる SX-3 の動作周波数を凌駕する 552.5MHz の動作周波数を達成すると共に、6000 万トランジスタを集積し、SX-4 と比較して動作周波数で 4.4 倍、集積度で 15 倍を実現しました。又 SX-4 では 37 個の LSI で構成されていたプロセッサ機能を、SX-7 では写真に示すように 1 チップで実現しました。

今回開発した LSI では、NEC が独自に開発した最先端 CMOS デバイステクノロジである 0.15 μm 、8 層銅配線プロセスの採用と低電源電圧化を進めることにより、高

集積化と高性能化を実現しました。

表4 フルカスタム LSI の諸元比較

項目	SX-7	SX-4
設計ルール	0.15 μm	0.35 μm
搭載トランジスタ数	6000 万	400 万
動作周波数	552.5MHz	125MHz
配線層構成	銅 8 層	アルミ 4 層
電源電圧	2.5V, 1.8V	3.3V, 1.2 V

CPU の LSI には 128KB のキャッシュメモリを内蔵していますが、これらはデバイス性能を最大限引き出すためにすべて専用設計をしています。メモリセルに高密度セルを採用し収容性向上を図ると共に、製造ばらつきの影響をキャンセルするタイミング生成回路を搭載し、高速かつ安定動作を実現しました。又、テスト時間を短縮するために、BIST (Built In Self Test) を採用しており、基本的な RAM 動作のほか複数のパターン依存テストが可能であり、様々な製造不良をリジェクトする信頼度の高いテストを実現しました。

チップ内へのクロック分配方式は、まずチップ全面へグローバルクロックを分配し、次にローカルエリアへクロック分配する 2 段階のクロック分配方式を採用しています。グローバルクロックには、クロック分配専用の低抵抗の厚膜配線層により配線して抵抗による波形なまりを低減すると共に、電源・グランド配線によりシールドすることによりクロストークノイズの影響を小さくしています。又クロックドライバ用電源を分離することによりクロックジッタを低減しています。

LSI の PAD 配置にはフリップチップ方式を採用しており、LSI パッドから内部電源配線までの線長を短縮することにより抵抗およびインダクタンス成分を低減し、電源電圧変動を抑制しています。又、オンチップの大容量デカップリングコンデンサを搭載することにより、高速スイッチング動作によるノイズを低減しています。

システムの性能向上のためには、LSI の高速化に加えて LSI 間的高速信号転送が必要になります。この実現のために、LSI 間の遅延ばらつきの影響を緩和するデータ転送方式、ノイズや信号波形歪みの低減を実現する回路技術を開発しました。

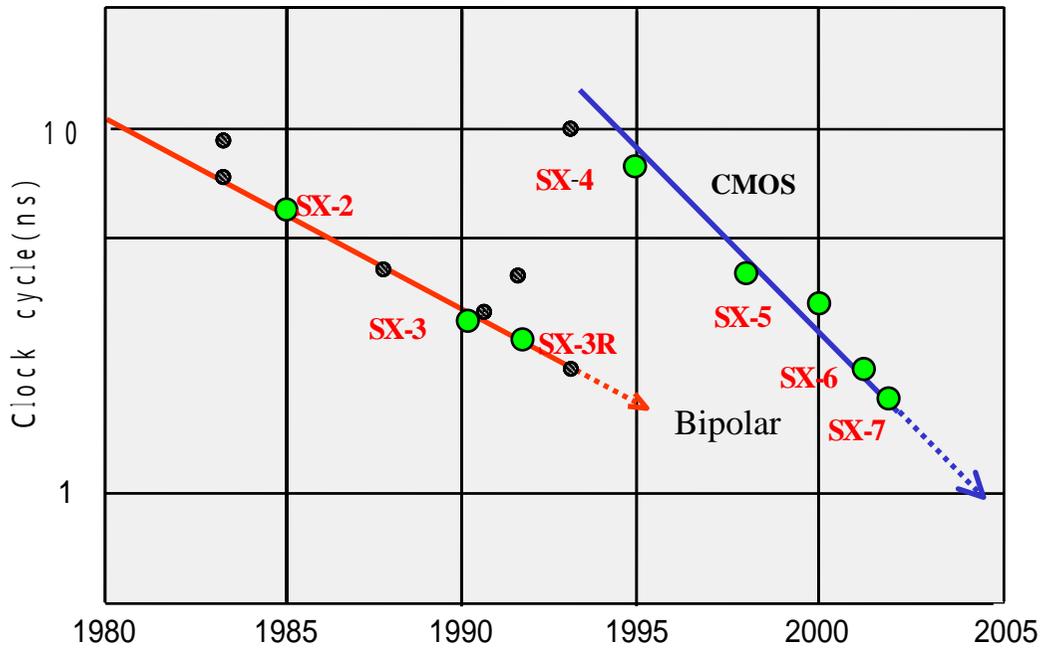


図7 スーパーコンピュータのクロックサイクル

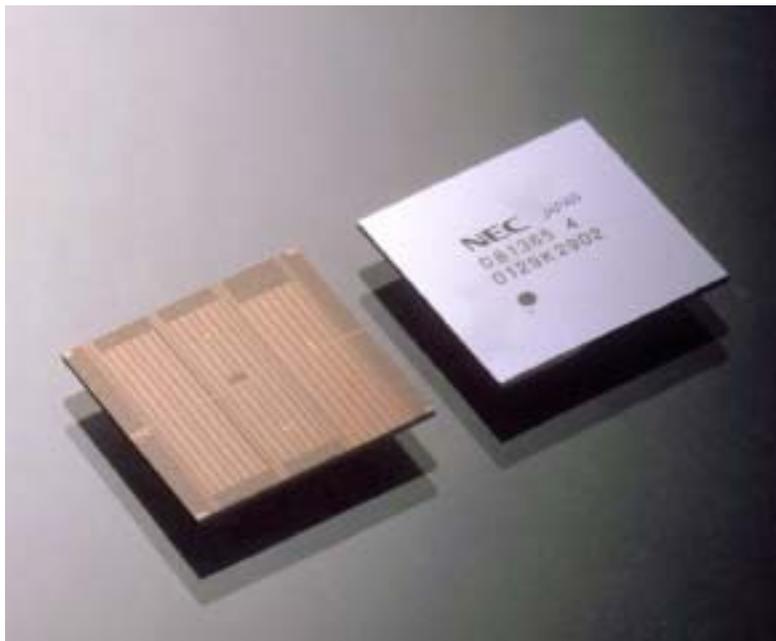


写真1 CPUチップ外観

3.2 実装技術

SX-7 システムでは、1つのノード内においてすべての CPU が巨大なメモリ空間を共有する共有メモリ方式に対応するため、1つのケーブル筐体の4面に CPU パッケージと MMU (Main Memory Unit) パッケージを実装し、1つのノードを構成しています。

筐体の1面に最大 16 個の CPU パッケージを搭載し、隣接する1面に最大 64 個の MMU パッケージを搭載することによって、1つのノードを構成する1つの筐体に 32CPU/128MMU パッケージという高密度実装を実現し、従来機種と比較して大幅な省スペース化を実現しています。そして、CPU/MMU パッケージ間を高速ケーブルで接続しています。図 8 に SX-7 の外観と筐体構造を示します。

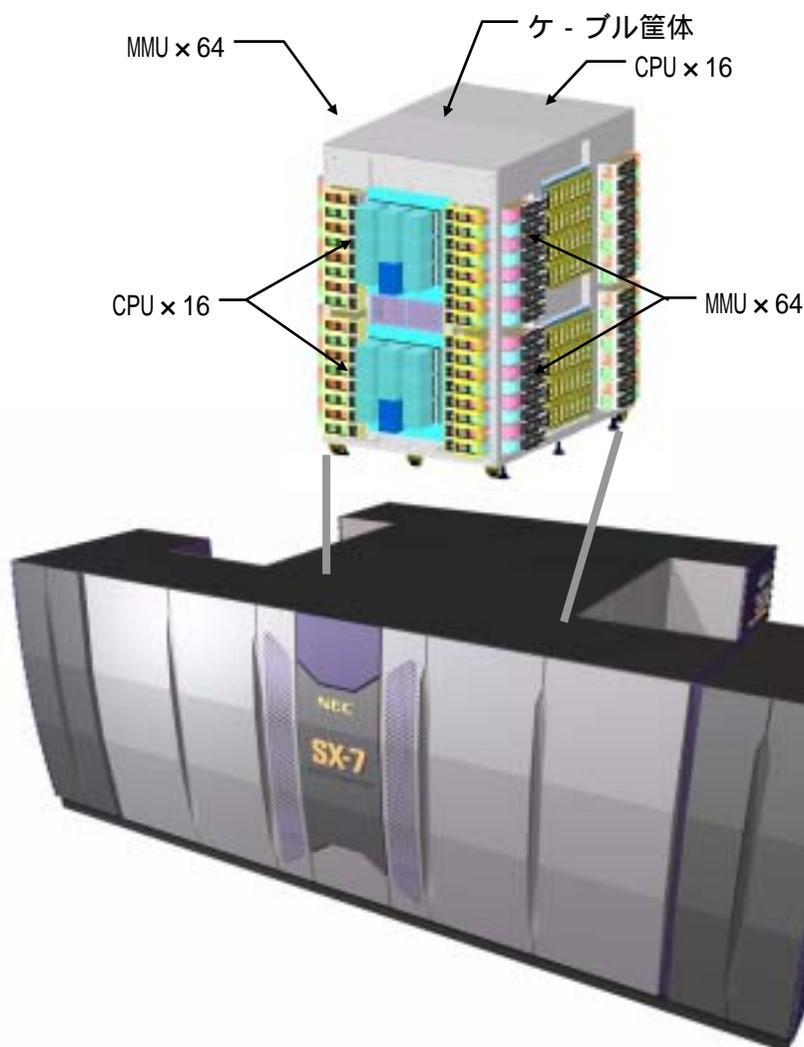


図 8 SX-7 の外観と筐体構造

さらに、システム全体を高速動作させるため、高密度 LSI 実装技術、高性能ケーブル接続技術、高効率冷却技術、高性能電源モジュールを採用しています。

CPU/MMU パッケージといった超高速動作を要求されるパッケージにおいては、超高速 LSI の性能を最大限に引き出すために LSI と配線基板間の接続方式としてベアチップ実装方式を採用しています。LSI を直接高密度プリント配線基板に実装することにより、従来 LSI ケース部で発生していたノイズ、信号遅延などを削除し、高速伝送を実現しています。

CPU パッケージには、1 チップベクトルプロセッサを搭載しています。また、MMU パッケージには、1 個の制御用 LSI と、6 個の RAM を搭載したメモリキャリア 16 枚を搭載しています。メモリキャリアには 6 個の RAM を搭載しており、MMU パッケージ当たり 96 個の RAM を搭載しています。写真 2 に MMU パッケージの外観、写真 3 にメモリキャリアの外観を示します。

高速伝送を要求される CPU/MMU パッケージ間の接続には、コネクタ付き高速細径同軸ケーブルを採用しています。このケーブルは伝搬遅延時間が 3.8ns/m と伝送特性に優れるとともに、狭ピッチでの同軸ケーブル多極化により、ノード内の全 CPU/MMU パッケージ間の高速伝送と多信号接続を実現しています。

また、LSI の高速・高集積化に伴って、LSI 発熱量が従来機種と比べて約 3 倍に増大しています。さらに高密度実装により、装置全体の発熱密度も高くなっています。このため、熱流体シミュレーションを利用し、最適冷却設計を実施しました。これにより、全部品について規定温度以下での動作を可能とし、高い信頼度を実現しています。

電源開発においては、従来機種と比較して低電圧化が進み、給電系の電圧ドロップ、インダクタンス低減ならびに負荷変動応答速度向上が課題となりました。これらの課題に対してアクティブクランプ+同期整流回路を採用し、高効率および従来機種電源モジュール比約 2 倍の応答速度を実現した DC-DC コンバータを採用するとともに CPU/MMU パッケージにプラグインコネクタを介して 1 対 1 接続する分散給電方式を採用しました。

4. 終わりに

以上、SX-7 のハードウェアの概要について説明しました。SX-7 は、スーパーコンピュータの要件である CPU の高い演算性能とそれに見合う CPU - 主記憶間の強力なデータ供給能力のバランスを重視し、ユーザに使い易い大規模共有メモリシステムとして開発しました。今後も、ユーザのご要望を取り入れながら、各種の研究分野の発展に寄与するスーパーコンピュータを開発していきたいと思っております。



写真2 MMUパッケージの外観

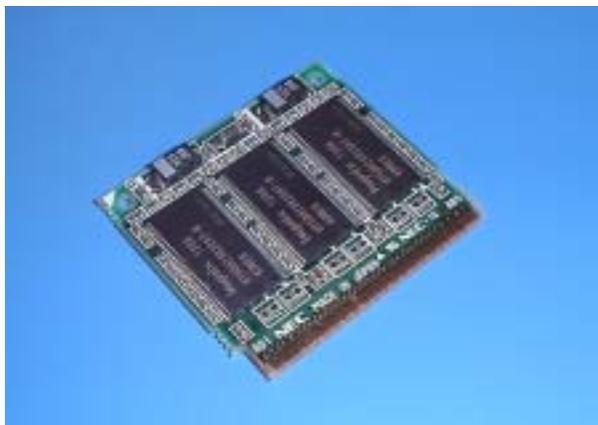


写真3 メモリキャリアの外観