〈SX-8のハードウェア〉

SX-8のハードウェア技術(1)～プロセッサ・主記憶装置～

Central Processing Unit and Main Memory Unit of SX-8

篠原真史* 古澤一昭** 西垣泰洋*
Masafumi Shinohara Kazuaki Furusawa Yasuhiro Nishigaki
多賀谷聡* 鈴木栄司**
Satoru Tagaya Eiji Suzuki

要　旨

SX-8のCPUは，4セットのベクトルパイプラインを有することにより，16GFLOPSという高い単体性能を実現しています。また，最大128Gバイトのメモリを最大8CPUが共有することで，ノード当たり最大128GFLOPSという高い性能を実現しています。

本稿では，この高い性能を実現しているSX-8のCPU，メモリ，および高い実効性能を実現するために必須となるCPU－メモリ間のネットワークについて説明します。

A CPU of SX-8 has 4 vector pipelines and shows 16GFLOPS peak performance. The SX-8 single-node system is configured with up to 128Gbyte main memory shared by up to 8 CPUs. Therefore, the maximum performance of SX-8 single node system is 128GFLOPS.

This paper describes the CPU, main memory, and network between CPU and main memory, which play a significant role to realize high effective throughput.

1．まえがき

スーパーコンピュータSX-8は，SX-4およびSX-5，SX-6で実証された高い実効性能とノード内共有メモリの利便性を継承しつつ，さらに増大する科学技術計算ニーズに応えるため，システム全体性能の強化はもとより，単一プロセッサ性能の強化とそれに伴う主記憶データ転送能力を強化しています。

以下では，その特長を中心に中央処理装置（Central Processing Unit：CPU），主記憶装置（Main Memory Unit：MMU），およびCPU－メモリ間ネットワークについて紹介します。

2．プロセッサ

SX-8のCPUは，従来のSXアーキテクチャを継承し，さらに機能・性能の強化を図っています。

2.1　プロセッサ構成と諸元

CPUは大きく分けて，スカラユニットとベクトルユニットで構成され，プロセッサ－メモリ間ネットワークを介して，MMUに接続されています。**図1**にCPUの構成を示します。

スカラユニットは，命令の解読，ベクトルユニットへのベクトル命令の供給・起動，スカラ命令の実行を行います。ベクトルユニットは，4セットのベクトルパイプラインを備え，クロックサイクル当たり，4個の加算と乗算を同時に処理することで最大16GFLOPSのベクトル演算性能を実現しています。

2.2　ベクトルユニット

ベクトルユニットは，ベクトル演算部，ベクトル制御部から構成されます。

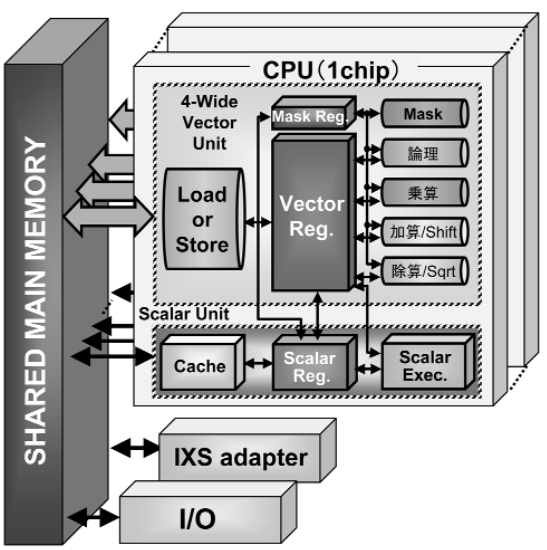

図1　CPUの構成
Fig.1　Configuration of CPU.

＊ コンピュータ事業部
Computers Division

＊＊ NECコンピュータテクノ
NEC Computertechno, Ltd.

（1）ベクトル演算部

ベクトル演算部は，おのおの独立に並列動作可能な論理演算，乗算，加算/シフト演算，除算の基本演算パイプラインに加え，マスク演算，ロード/ストアの各パイプライン，および64ビットの容量を持つ16個のマスクレジスタ，おのおの512バイトの容量を持つ72個のベクトルレジスタから構成されるベクトル演算セットを1セットとし，先端半導体テクノロジを採用した高集積LSI 1チップに4セットを収容することにより実現しています。1台のプロセッサでは，4セットのベクトルパイプラインを有することにより，合計144Kバイトのベクトルレジスタを備え，32本の基本演算パイプラインを同時に動作させることにより，強力なベクトル処理を行います。

特に，スカラユニットおよびベクトルユニットすべてを同一チップに収容したことで，命令分配，演算パイプライン間のデータ転送，さらにはスカラユニットとのデータ転送などが短時間で実現でき，演算実行のオーバーヘッドを大幅に短縮しました。

1）ベクトル演算パイプライン

おのおのの演算パイプラインは，IEEE倍精度浮動小数点データ形式およびIEEE単精度浮動小数点データ形式をサポートしています。それぞれのデータ形式の切り替えは，命令ごとに可能で，従来システムやワークステーションなどとの高い親和性を確保しています。また，IEEE浮動小数点演算仕様の丸めモードや無限大数などの特殊数の処理も，単精度/倍精度共にハードウェアで実現することで，高速な演算処理を行います。

加算/シフト演算，乗算，論理演算の各演算パイプラインは，パイプライン段数を極限まで短くすることで，短ベクトル長処理をも高い処理性能を発揮することができます。

また，SX-8で新規サポートした平方根パイプラインは，除算パイプライン回路を可能な限り共用化しつつ，倍精度浮動小数点演算で1GFLOPSという高い性能を実現しています。

ベクトルパイプ間相互のデータ転送においても，64Gバイト/秒の転送性能により，ベクトルデータのベクトルパイプ間移送，マスクビットによるベクトルデータの圧縮・伸張も高速に処理することができます。

2）ベクトルレジスタ

それぞれが最大256語まで保持可能なベクトルレジスタを72個用意し，144Kバイトの容量を有しています。

主記憶装置からベクトルレジスタへのデータ供給能力は，1セットのベクトルパイプラインで1マシンサイクル当たり1語（8バイト/語）の転送能力を持ち，4セットのベクトルパイプラインが同時に動作できるように構成しています。

各ベクトル演算パイプラインおよびロード/ストアパイプラインへのデータの供給/格納を同時に行えるようにし，機能リソースを有効に使用できるように構成してあります。

3）ベクトルマスクレジスタ

ベクトルマスクレジスタを16個備えており，それぞれを最大1ビット×256語で構成しています。

マスクビットの生成は，論理演算パイプラインを使用して並列に行うようにし，また，他のマスク付き演算とベクトルマスクレジスタ間の演算，あるいは，ベクトルマスクレジスタ間の演算と他のマスク付き演算とのチェイニング機構をより強化することで性能の向上を図っています。

（2）ベクトル制御部

ベクトル制御部は，スカラユニットからのベクトル命令の解読，並列パイプラインをインターリーブ方式で動作させるための制御，並列パイプラインに対応するベクトルデータおよびマスクビットのアライン制御，ベクトルデータ長の制御を行います。

ベクトル制御部をベクトル演算パイプラインと同一チップ上に構成できたことにより，命令起動およびリソース情報収集などにかかる時間が短縮され，ベクトル長に依存しない高い実行性能を提供します。

2.3 スカラユニット

SX-8のスカラユニットは，命令同時デコード数4，命令同時発行数4のスーパースカラアーキテクチャを採用，アウトオブオーダ実行，命令投機実行をサポートしています。

スカラユニットは，SXシリーズ互換の64ビットRISCアーキテクチャであり，64ビット×128ワードの汎用レジスタ，命令・データそれぞれ64KバイトのL1キャッシュを内蔵しており，SX-8ではピーク性能は2GFLOPSに達しています。

スカラユニットの特長を以下に示します。

（1）高精度分岐予測機構

分岐予測情報として複数回の分岐履歴情報を保持することにより分岐予測精度の向上を図っており，分岐予測失敗による命令パイプラインストールの頻度を小さくしています。

（2）4命令デコード・4命令発行のパイプライン構成

SX-8スカラユニットでは，最大4命令同時デコード可能であり，高い命令供給能力を提供します。また，命令ディスパッチ，発行，完了処理の各ステージにおいても4命令/サイクルのスループットを確保し，各パイプラインステージで命令パイプラインストールが発生しないバランスの良い構成になっています。

（3）投機実行制御

アウトオブオーダ実行機構によって，プログラム順によらず実行可能となった命令から実行することで高性能を実現しています。しかし，さらなる性能向上のためには複数の分岐命令を跨いだ，より大きな領域に対する実行可能命令の検索が必須です。

SX-8スカラユニットは，最大6つの分岐命令を跨いだ命令領域に対して実行可能命令の検索が可能です。また，実際に分岐命令が行われるまでの間，その分岐命令の先の命令を仮に実行し，分岐予測が失敗した場合には正しい状態から実行を再開する，投機的命令実行制御を採用しました。そのために48エントリのリオーダーバッファを持ち，発行制御を行っています。

3．MMUとネットワーク

MMUはCPUとリモートアクセス制御装置（Remote access Control Unit：RCU）に接続され，高速かつ均一にアクセス可能な共有メモリ方式を採用しています。

MMUではお客様の幅広いニーズに応えるために512Mビット FCRAM（Fast Cycle RAM）/DDR2-SDRAM（Double Data Rate-SDRAM）の2種類のRAMを採用し，高速アクセス・大容量メモリを提供しています。両者の大きな違いはRAMのスペックであるランダムアクセス時間（tRC）が異なる点です。FCRAMは高速であり，バンク競合による待ち時間が短いため性能に優れていますが，DDR2-SDRAMは安価という利点があります。

（1）MMUの構成

MMUは，最大構成で32枚のMMUカードから構成されます。おのおののMMUカードはCPUやRCUからのメモリアクセス要求や，プロセッサ間通信要求に対して同時に並列して動作でき，極めて高いデータ転送能力を有しています。

1MMUカード当たりのメモリ容量は，FCRAMの場合2Gバイト，DDR2-SDRAMの場合4Gバイトが実装され，システムではそれぞれ64Gバイト，128Gバイトと大容量のメモリ構成となります（**表**）。

メモリスループット性能は，1MMUカード当たり16Gバイト/秒，システムでは最大512Gバイト/秒のデータ供給能力を実現しています。また，MMUではメモリを4,096wayのインターリーブ数に分割し制御することで，高スループット性能を実現しています。

（2）MMUの特長

SX-8ではSX-4以降採用している3次元構造のMMUカード実装を引き続き採用しています。MMUカードは，プリント基板に，BGA（Ball Grid Array）封止のFCRAMまたはDDR2-SDRAMを搭載したメモリキャリアと，それを制御する高速CMOS LSIとを3次元に搭載します。高密度実装はもちろんですが，本実装方式の狙いは，RAMとCMOS LSIとの物理的な距離を近くして，RAMの高速動作を可能にすること，およびCPUとの高速信号転送を実現することにあります。

表　MMUの諸元
Table　Specifications of MMU.

項　目	諸　元
記憶容量	64Gバイト（FCRAM） 128Gバイト（DDR2-SDRAM）
インタリーブ数	4,096 way
記憶素子	512Mビット FCRAM 512Mビット DDR2-SDRAM
データ供給能力	512Gバイト/秒

ECC（Error Correcting Code）符号の採用による信頼性の向上，さらにタイミング，パリティ，2重化回路などのチェックにより高い故障検出率を実現しています。また，疑似障害発生によるチェック回路の診断機能，エラー内容から即座にエラー個所を指摘できるビルトイン診断機能などを持っており，RAS（Reliability Availability Serviceability）機能の充実を図っています。

SX-8から，MMU内にはメモリバンクキャッシュと呼ぶキャッシュ機能をCPUごとに備えています。各CPUからのロードリクエストに対しキャッシュヒットした場合はメモリレイテンシの短縮が可能となりスループット性能を向上させています。

また，MMUには通信レジスタ（Communication Register：CR）を備えており，自動並列処理やOpenMP指示行による共有メモリ内の並列処理時のCPU間の同期制御を高速に行っています。

4．プロセッサ - メモリ間ネットワーク

CPUと大容量MMUを接続するネットワークは，プロセッサネットワークとメモリネットワークで構成され，シングルノードシステムで最大512Gバイト/秒を転送する能力を備えています。

図2にプロセッサ－メモリ間ネットワークの構成を示します。

4.1　プロセッサネットワーク

プロセッサネットワークは，並列に高速演算するベクトル演算部と，複数のCPUに共有されるMMUとの間で，演算能力を最大限に引き出し，高い実効性能を得るために大容量，高速のデータ転送を行います。

（1）リクエスト転送方式

SX-8では8バイトデータ幅をメモリアクセスのリクエスト単位としています。リクエスト転送方式は個々のリクエストに目的メモリポートあるいは目的ベクトルポートに到達するための転送先情報，あるいは転送先での動作規定情

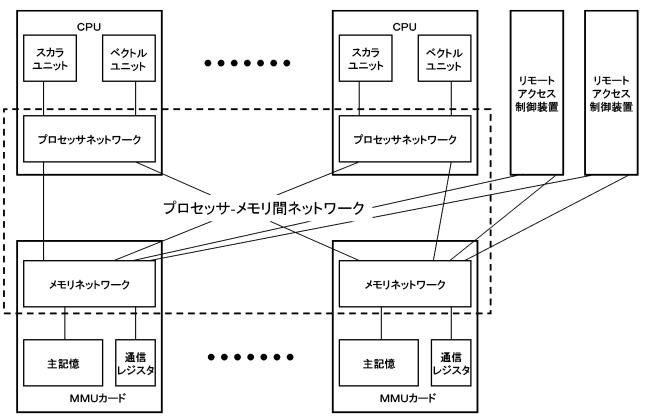

図2　プロセッサ - メモリ間ネットワークの構成
Fig.2　Configuration of processor-memory network.

報などのルーティング情報を付与し，リクエスト自らがリクエストの転送経路の選択および競合調停を行う，セルフルーティング方式を採用しています。

これにより複数のCPUと，複数のMMUカードに分散して構成されるプロセッサ－メモリ間ネットワークにおいて，それぞれのリクエストが独立してメモリアクセスする分散動作が可能になり，データ転送性能の低下なく大規模な共有メモリ構成を実現しています。

プロセッサネットワークはこのセルフルーティングに必要なルーティング情報，およびメモリアドレスを，8バイトデータ幅単位に連続的かつ並列に生成し，CPU当たり64Gバイト/秒のリクエスト転送を可能としています。

(2) ネットワーク制御

CPUはMMUと接続されるCPUポートを32ポート備えており，メモリリクエストを，個々のルーティング情報に従い所望のMMUに接続されているCPUポートに出力します。CPUが発行するメモリアクセス命令は先行する命令との間でメモリアクセスの順序を保証する必要があります。一方並列に発行される複数のリクエストが同一CPUポートに向かうとポート競合が発生しリクエスト間の競合調停が必要となります。プロセッサネットワークではこの順序保証と競合調停を効率よく行い，実効スループット性能の向上を実現しています。

CPUから送出されたメモリリクエストは，メモリアクセス後，メモリリクエストを送出したCPUポートにアウトオブオーダで戻ってきます。プロセッサネットワークでは，この複数のポートから戻ったメモリデータを効率よく所望のベクトルポートに転送し，ベクトル演算部に高速にデータを供給することで高い実効性能を実現しています。

4.2 メモリネットワーク

メモリネットワークは，最大8台のCPUと最大2台のRCUに接続され，メモリアクセスのデータ転送制御，およびプロセッサ間通信制御を行います。

(1) データ転送制御

メモリネットワークは，CPU/RCUに接続するクロスバスイッチを備え，高速のデータ転送制御を行います。

共有メモリ密結合型のマルチプロセッサシステムにおいて，高い並列処理効果を得るためには，複数CPUからの高負荷のメモリアクセス競合に対しても，効率よく競合調停を行い，極端な性能劣化が生じないようにデータ転送制御を行う必要があります。特にMMUカードの入力ポートは異なるCPUに接続されるため，入力ポート間でのアクセスパターンに相関性がなく，アクセス競合が発生しやすくなります。

メモリネットワークの各スイッチが独立にルーティング制御と競合調停を行うことにより，高い実効スループット性能と，優れたスケーラビリティを得ることができます。

(2) プロセッサ間通信制御

CPU間の通信制御は，メモリネットワークのクロスバスイッチ機能を利用して実現しています。プロセッサ間通信では，指定したCPUへの通信に加え，すべてのCPUへのブロードキャスト通信，および指定した複数のCPUへのマルチキャスト通信を行うことができます。

5. むすび

以上，SX-8のプロセッサ，MMU，プロセッサ－MMU間ネットワークを中心に紹介しました。SX-8は，高い実効性能で定評のあるSXアーキテクチャを継承しながら，新規命令のサポート，メモリレイテンシの短縮などを実施することにより，さらに実効効率の高い，コストパフォーマンスに優れたスーパーコンピュータ製品として開発しました。

今後もユーザニーズを取り入れ，最先端技術を駆使し，より優れたスーパーコンピュータを開発していく所存です。

＊ FCRAMは，富士通（株）の商標，並びに登録商標です。

筆者紹介

Masafumi Shinohara
篠原　真史　1988年，NEC入社。現在，第一コンピュータ事業本部コンピュータ事業部第四技術部技術エキスパート。情報処理学会会員。

Kazuaki Furusawa
古澤　一昭　1987年，NEC甲府（現NECコンピュータテクノ）入社。現在，コンピュータ第二技術部技術マネージャー。

Yasuhiro Nishigaki
西垣　泰洋　1991年，NEC入社。現在，第一コンピュータ事業本部コンピュータ事業部第四技術部技術主任。情報処理学会会員。

Satoru Tagaya
多賀谷　聡　1994年，NEC入社。現在，第一コンピュータ事業本部コンピュータ事業部第四技術部技術主任。

Eiji Suzuki
鈴木　栄司　1992年，NEC甲府（現NECコンピュータテクノ）入社。現在，コンピュータ第二技術部主任。