スーパーコンピュータ「SX-4シリーズ」の開発思想と概要

ＮＥＣ　スーパーコンピュータ販売推進本部
渡　辺　　貞
(E-mail：watanabe@sxsmd.ho.nec.co.jp)

■開発思想

SX-4シリーズは、従来のSX-2、SX-3シリーズに比べて、総合的な観点での利点を、より深く追求する開発思想になっています。これは、次のようなユーザ動向、技術動向に基づいています。

ユーザの動向

依然として超高速化への要求には限りがなく、しかも、低価格化に対する要求が一層強くなってきています。これには消費電力、設置面積などの運用コストの低減も含まれています。

ローエンド領域では、価格性能比の良いワークステーションやサーバへのシフトが起きています。また、より高い性能領域では、ワークステーションクラスタや、汎用マイクロプロセッサを多数用いた超並列機(MPP)を使って、少ない投資で何とか高性能が得られないかという並列化の試みが行われています。

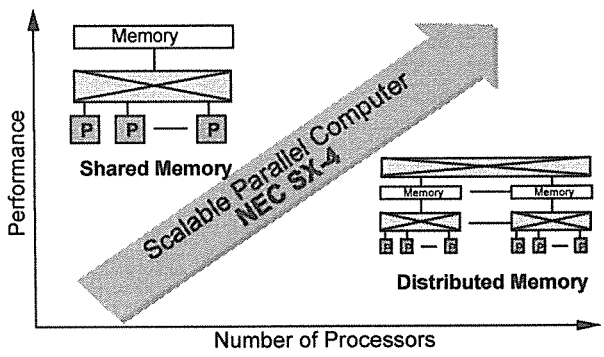
また、高性能といっても、単一ジョブの高性能に加えて、複数ジョブのスループット性能も重要視されていることには変わりありません。

さらに、最近のネットワークコンピューティングにおける相互利用の観点から、標準化への要求が強くなってきています。

アーキテクチャの動向

ベクトルプロセッサを用いた、従来の共有メモリ型マルチプロセッサシステムの他に、ベクトルプロセッサの分散メモリ型マルチプロセッサシステムや、ネットワーク上の複数のワークステーションを用いたワークステーションクラスタ、超並列機が試みられています。これらは大別すると、スカラ型かベクトル型かのプロセッサ方式と、共有メモリ型か分散メモリ型かのメモリ方式に分かれます。

プロセッサ方式については、高いメモリ性能を持ったベクトルプロセッサの方がより広い領域で安定した高性能が得られ、メモリ方式については、運用性、並列化技術の現状や使い易さを考えると、共有メモリ方式の方がはるかに優れていると言えます。しかし、共有メモリ方式はその物理的限界から、ある範囲を超えた超高性能領域では分散メモリ方式にならざるを得ないところがあります。



いずれにしても、価格と機能／性能の兼ね合いになります。現状では、限られた用途のものを除き、ベクトルプロセッサを用いた従来の共有メモリ型マルチプロセッサシステムが依然として主流であり、分散メモリ方式での並列化への動きは極めてゆっくりとしたものになっています。

テクノロジの動向

過去、スーパーコンピュータには、その素子の高速性からバイポーラによるECL(Emitter Coupled Logic)回路が当然のように使われてきました。この回路は、SX-2、SX-3シリーズにも使われましたが、本質的に集積度が小さく、消費電力が大きいことから、今後この回路を用いても、装置としての性能向上度をいままでのようには望めないと予想されています。

これに代わるテクノロジとして、最近、ガリウムひ素の回路やCMOS(Complementary Metal-Oxide Semiconductor)回路が、その性能レベルに応じたスーパーコンピュータに採用されてきています。

特にCMOS回路は、汎用マイクロプロセッサの多くで使われ、競争により高速化技術は飛躍的に進み、これを用いた装置の性能は、ECL回路を用いたものにかなり接近してきています。ECL回路に比べると10分の1の消費電力、10倍の集積度が期待でき、低価格化が可能なため、スーパーコンピュータの高性能領域でも有望視されています。

開発方針

これらの動向をふまえ、SX-4シリーズを以下の方針のもとに開発しました。

(1)価格性能比、設置性の大幅な向上

・CMOSテクノロジ採用による、低価格化、小型化、低消費電力化、空冷化

(2)幅広い性能レンジの製品をスケーラブルに提供

・短ベクトルにも強い高性能ベクトルプロセッサ

・使いやすい共有メモリ方式のマルチプロセッサシステム

・共有メモリを単位とした分散メモリ方式のマルチプロセッサシステムと分散メモリ間高性能ネットワーク

(3)標準化の推進

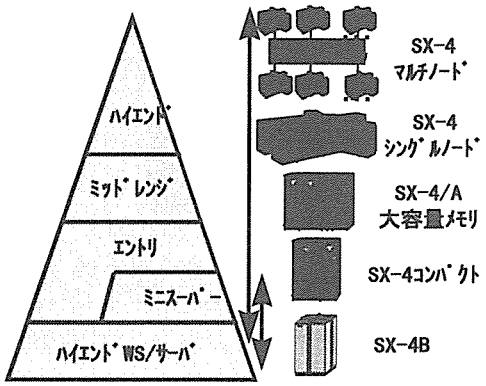
・SX-3シリーズで十分に実績のあるUNIXベースのOS、SUPER-UXの強化、拡張

・従来のIBM及びCRAYデータ形式に加えて、ﾜｰｸｽﾃｰｼｮﾝとの親和性を高める　IEEEデータ形式のハードウェアでのｻﾎﾟｰﾄ

・入出力インタフェース、ネットワークに、SCSI、HIPPI、Ethernet、FDDI、ATMなどのサポート

・従来のFORTRAN77、C、C++などに加え、Fortran90のｻﾎﾟｰﾄ

・分散メモリ型の並列処理で世界標準になりつつある、MPI(Message Passing Interface)、HPF(High Performance Fortran)などの並列処理機能のサポート



(4)使いやすさの追求

・シングルシステムイメージ(SSI)の提供
・SX-2,SX-3 シリーズで十分に実績のある自動ベクトル化、自動並列化、プログラミング開発環境の強化及び拡張
・クロスコンパイラの提供
・SX-3 シリーズとのロードモジュール互換

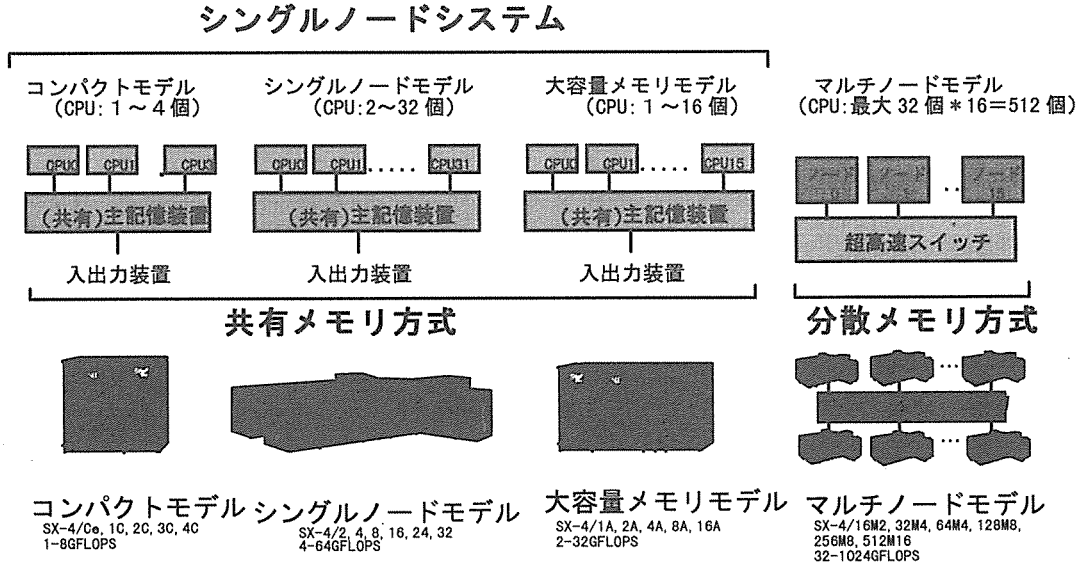
■ハードウェアの概要

スケーラビリティに優れたマルチプロセッサシステム

SX-4 シリーズは共有メモリと分散メモリアーキテクチャを融合した新しいタイプのスーパーコンピュータで、メモリ共有型のシングルノードシステムと複数のシングルノードシステムをクラスタ接続し、共有メモリと分散メモリとを組み合わせたマルチノードシステムとから構成されます。

シングルノードシステムは、最大 4 台までのプロセッサをコンパクトな筐体に実装して設置性を高めた 1～8GFLOPS の性能を有するコンパクトモデル（最大主記憶容量は 4 Gバイト）、2～32 台のプロセッサで 4～64GFLOPS の性能を有するシングルノードモデル（最大主記憶容量は 16 Gバイト）、及び共有メモリ型では世界最大の 32 Gバイトの主記憶容量を実現し、1～16 プロセッサで 2～32GFLOPS の性能を有する大容量メモリモデルからなります。

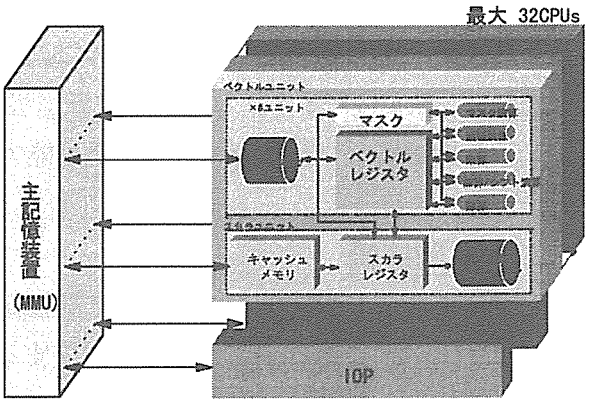
マルチノードシステムは、超高速なクロスバースイッチからなるノード間接続装置、または HIPPI スイッチを介して最大 16 ノードをクラスタ接続することにより、最大 512 台のプロセッサを接続でき、1TFLOPS のベクトル性能を実現しています。



高性能プロセッサ

各プロセッサは8セット（モデルCeは4セット）のベクトルユニットと1個のスカラユニットから構成されます。ベクトルユニットは各々独立に並列動作が可能な論理演算、乗算、加算／シフト演算、除算の合計4本の基本演算パイプラインと18Kバイトのベクトルレジスタから構成され、超高速、高集積なLSI1チップにより実現しています。

したがって、各プロセッサは144Kバイトのベクトルレジスタを備え、32本のベクトル演算パイプラインを同時に動作させることが可能です（モデルCeは各半分）。



中央処理装置(CPU)の構成

スカラユニットはRISCアーキテクチャを採用し、LSI1チップで実現したスカラ演算およびアドレス演算用の128個のスカラレジスタとスカラ演算パイプラインをもつスカラプロセッサと、128Kバイトのキャッシュメモリから構成されます。さらに、スーパースカラ方式を採用することにより1クロック当たり同時に2命令を処理し、高速なスカラ処理を実現しています。

また、SX-4シリーズでは浮動小数点データ形式として、IEEE形式、IBM形式、CRAY形式の3種類をハードウェアでサポートしており、各データ形式はプログラムごとに切り替えて使用することが可能なため、さまざまなシステムとの高い親和性も確保しています。

主記憶装置

高速ベクトル処理と並列処理を効率よく実行するには、プロセッサに大量のデータを連続的かつ高速に供給する必要があります。そこで、主記憶装置の記憶素子として超高速な4Mビットシンクロナス SRAM（大容量メモリモデルは高速な16Mビットシンクロナス DRAM）を採用することにより、シングルノードシステムの共有メモリはプロセッサ当たり16Gバイト／秒（モデルCeは8Gバイト／秒）、ノード当たり最大512Gバイト／秒のデータ供給能力を備えた高性能メモリシステムを実現しています。マルチノードシステムでは最大8Tバイト／秒のスループットとなります。主記憶容量についてはシングルノードモデルで最大16Gバイト（大容量メモリモデルは32Gバイト）、シングルノードモデルをノードとしたマルチノードシステムで最大256Gバイト（大容量メモリモデルの場合には512Gバイト）の容量を実現しています。

拡張記憶装置

入出力動作時間を大幅に短縮するために16MビットのDRAMを採用した高速、大容量の拡張記憶装置を磁気ディスク装置と主記憶装置との間の階層記憶装置として準備しています。

拡張記憶装置は、超高速のファイル装置やプログラムのスワッピング装置として、またディスクキャッシュとして利用できるほか、FROTRAN プログラムの配列データ空間として利用することによ

り、主記憶容量を超える大規模なプログラムを実行することができます。

シングルノードシステムでは最大32 Gバイトの容量で、最大16 Gバイト／秒のデータ転送性能を、マルチノードシステムでは最大384 Gバイトの容量と192 Gバイト／秒のデータ転送性能という大容量かつ超高速データ転送性能を実現しました。

高速入出力処理

高速な入出力処理を実現するために、シングルノードシステムでは最大64本のHIPPIチャネルを接続し、6.4 Gバイト／秒のデータ転送性能を、マルチノードシステムでは768本、76.8 Gバイト／秒の大規模かつ超高速の入出力処理能力を実現しています。

ノード間接続

マルチノードシステムの各ノード間接続として2種類の方法を準備しています。ひとつは標準インタフェースであるHIPPI経由で接続する方法で、100 Mバイト／秒のデータ転送速度でHIPPIスイッチを経由して各ノード間を接続します。もうひとつはクロスバースイッチ構造のノード間接続装置(IXS:Internode Crossbar Switch)を介して接続する方法で、各ノード間を最大8 Gバイト／秒の高性能なインタフェースで接続します。

最先端のテクノロジの採用による設置性向上

400万トランジスタを収容可能な0.35 μ mの最先端プロセスによる超高速、高速積CMOS LSIを全面的に採用するとともに、主記憶装置には超高速シンクロナスSRAM素子(大容量メモリモデルは高速なシンクロナスDRAM素子)を採用しています。これらのテクノロジと高密度実装技術の採用により、全面的な空冷方式を実現し、従来のスーパーコンピュータに比べて大幅な低消費電力化、省スペース化を図り、設置性を格段に向上しています。

豊富なRAS機能

最新のテクノロジを採用することにより、ハードウェアの信頼性を飛躍的に向上させるとともに、RAS機能として主記憶装置と拡張記憶装置に誤り検出訂正符号(ECC)の採用をはじめ、回路の二重化などによる誤り検出回路の組み込み、故障発生時には故障個所を自動的に指摘するビルトイン診断機能(BID)、さらには故障個所を自動的に切り離して継続運転を行う自動再構成処理機能を備えています。また、障害情報の自動収集、サービスセンターへの自動通報、センターからの遠隔保守により迅速な障害対応と予防保守を実現し、システムの総合的な信頼性・稼働性・保守性を高めています。

■オペレーティングシステム

SX-4シリーズには、UNIX SystemVをスーパーコンピュータ向けに強化した最新鋭のオペレーティングシステム、SUPER-UXを搭載します。以下にSUPER-UXの特長を説明します。

オープンシステム指向

SUPER-UXは、UNIXの使い勝手の良さを受け継ぐとともに、様々な標準機能を積極的に取り入れています。例えば、API、コマンドインタフェースとしてPOSIXインタフェースをサポートし、各種流通ソフトウェア、ユーザプログラムの移植を容易にしています。また、分散コンピューティング環境(DCE)として、分散ファイルシステムDFSや、セキュリティ機能Kerberos認証システムをサポートしています。

さらに、UNIX標準プロトコルTCP/IP、およびネットワーク機能の一元管理を行うDNS、ネットワーク管理のためのSNMPなどの標準プロトコルをサポートすることにより、柔軟なネットワーク構築を可能とします。また、ネットワーク間の接続のためのBGP-4のサポートも検討しています。なお、今後の標準化動向についてもタイムリーに取り入れていく予定です。

高速性の追求

SUPER-UXでは、UNIXの特長を生かしつつ大幅な機能強化を行い、スーパーコンピュータにふさわしい高速性を実現しています。

(1)強力な並列処理機能

1ノード当たり最大32台のマルチプロセッサのサポートにより、カーネルからコンパイラ、ライブラリに至るまで効果的な並行動作が可能となり、システム全体の高性能化を実現しています。また、マルチタスク方式による並列処理によって各プログラムのターンアラウンドタイムも短縮することができます。さらに、並列プログラミング環境の基盤機能として、POSIX1003.4a相当のスレッド機能や、システムコール、3C/3S/3Mライブラリをリエントラント化したスレッドセーフなライブラリを提供しており、並列処理を強力に支援します。

(2)高速入出力・ネットワーク接続

SUPER-UXでは、バッファレスデータ転送、ファイル領域の連続割り当てなどにより入出力性能の大幅な高速化を実現したSFS(Supercomputing File System)や大容量入出力装置に適したSFS／H(Hybrid SFS)をサポートしています。また、SFSに対して拡張記憶装置を大容量キャッシュとして使用できるキャッシュ機能や、ディスクストライピング機能を実現した高速入出力サブシステムIAS(Intelligent I/O Accelerator Subsystem)を提供し、徹底した高速性の追求を行っています。

さらに、超高速HIPPIインタフェースによりHIPPIスイッチ、高速ディスクアレイ装置、マスデータプロセシングシステム(MDPS)、ソニー製ID-1テープライブラリ装置、高速画像処理装置(HIPS)、Ethernet,FDDI,ATMなどの高速入出力装置・ネットワーク接続が可能です。

充実した運用管理

SUPER-UX は、小型の高性能演算サーバから超大型のスーパーコンピュータまで幅広く適用できる、融通性の高いオペレーティングシステムです。インストレーションを簡易化し、導入からサービスインまでを短期間で行えるようにするとともに、各種システム統計情報や性能解析ツールを提供し、システムチューニングを強力に支援しています。

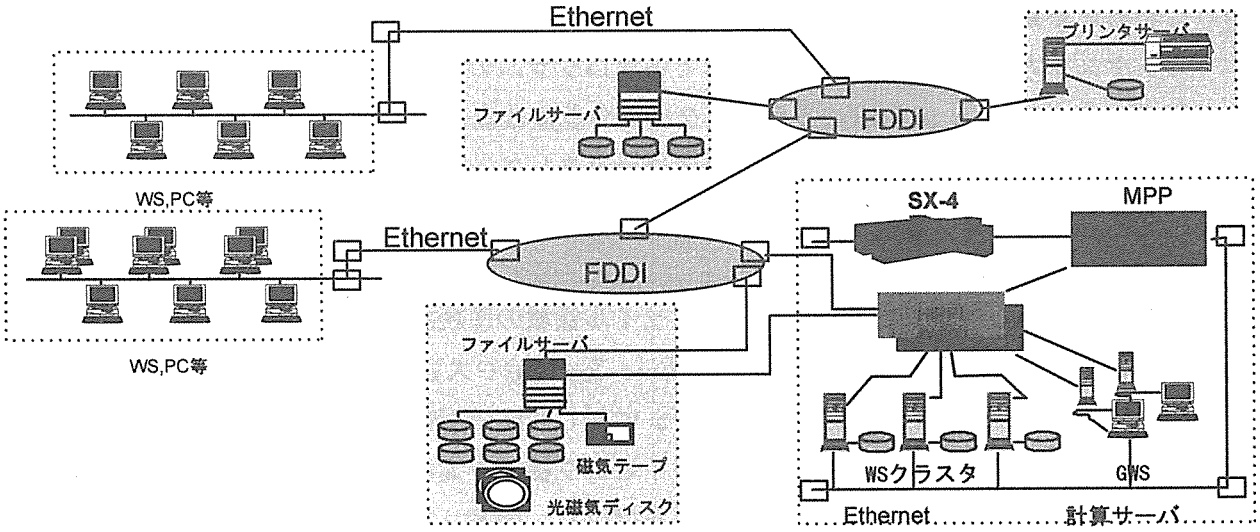
また、任意の時点で実行中のプログラムを中断／再開させるチェックポイント／リスタート機能、ファイルアーカイビングシステム SX-BackStore、個別電源制御、環境異常監視、自動オペレーションなどの自動運転機能、ネットワーク上の複数のホストマシンを 1 台の管理用ホストマシンで統合的に管理する分散管理機能 NetAdmin など多彩な機能を備えています。

多様なネットワーク環境への対応

SUPER-UX は、多様なネットワーク機能を取り入れており、計算センターのようなマルチベンダネットワーク環境においても計算サーバとして適用可能な十分な機能を有しています。

例えば、UNIX 標準プロトコル TCP/IP、BSD の全通信機能に加え、DNS、SNMP など標準プロトコルもサポートしており、ユーザニーズに応じた最適なネットワークを構築することができます。さらに BGP-4 など最新のプロトコルのサポートも検討しています。また、API としては、socket システムコール、TLI ライブラリをはじめ、X-Window（クライアント）システム、グラフィック・ユーザ・インタフェース Motif を提供しており、アプリケーションプログラムを極めて容易に作成できます。

さらに、HIPPI、Ethernet、FDDI、ATM といった高速の LAN をサポートしており、高速、大容量通信を実現しています。



大規模・高信頼システムへの対応

SUPER-UX は、シングルノードモデルにおいて、最大 16 G バイトの主記憶装置、および最大 32 G バイトの拡張記憶装置、大容量メモリモデルでは最大 32 G バイトの主記憶装置をサポートしており、超高速スーパーコンピュータにふさわしい大規模プログラムの実行にも適しています。また、マルチボリュームを仮想ボリュームの一形態としてサポートしているので、論理的にはテラバイトをはるかに超える大容量ファイルを容易に構築できます。各種障害処理に関しては、周辺装置の障害に

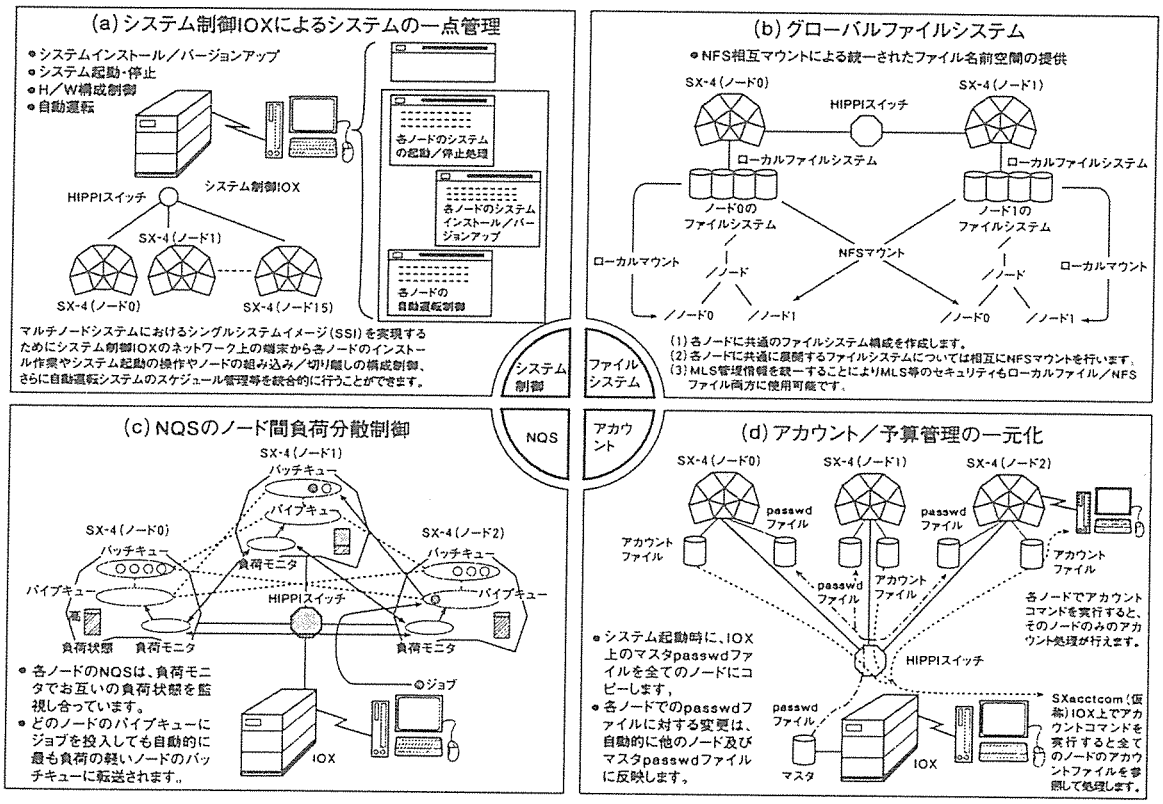
対するハードウェアとソフトウェアの２段階の障害回復制御、マルチプロセッサを構成する各プロセッサの障害に対する自動切り離し／組み込み機能、メモリ／入出力制御装置の構成制御機能などを備えており、高い信頼性を確保しています。

マルチノードシステムへの対応

マルチノードシステムにおいてもシングル・システム・イメージ(SSI)を実現しており、使いやすく快適な利用環境を提供します。システムの立ち上げ・監視、運用管理などは、単一のシステム制御 IOX で総合的に行えます。SUPER-UX 上の運用管理についても、グローバルファイルシステム、NQS のノード間負荷分散、全ノードに対する総合的なアカウント／予算管理などの機能を提供しています。このように、ユーザ側からはマルチノードシステム全体を１つのシステムとして利用することができます。

また、システム全体の運用管理を外部 IOX で実現するため、ノードごとのバージョンアップ、一部ノードの隔離運転などきめ細かな運用も可能となります。

さらに、コンパイラの自動並列化機能、マクロ／マイクロタスクによる並列化機能に加え、データ・パラレルを取り入れた HPF 言語、メッセージ・パッシングを標準化した MPI ライブラリをサポートしており、並列処理すべてのプログラミングモデルに柔軟かつ最適に対応させることができます。



マルチノードシステム(クラスタ制御)の機能概要

■言語および支援ツール

SX-4シリーズのハードウェアは、SX-3シリーズと比べ、さまざまな点で大きく拡張されています。スーパースカラの採用によるスカラ性能の強化、最大32台までの共有メモリ型マルチプロセッサ構成（ノード）のサポート、また最大16ノードまでの分散メモリ型クラスタ構成（マルチノード）のサポートなどが、その代表的なものです。これらの特長を十分に生かすために、言語および支援ツールも大幅に機能強化しています。ここでは、主な強化内容について説明します。

言語処理系としては、従来からのFORTRAN77/SX、C/SX、C++に加えて、FORTRAN90/SXが利用可能です。また、複数のノードにまたがる分散メモリ型の並列処理のために、新たにHPF/SXとMPI/SXをサポートしています。

FORTRAN90/SX

言語処理系の中核をなすのがFORTRAN90/SXです。その特長は次のとおりです。

①言語仕様

最新の規格であるFortran90規格を完全にサポートしています。Fortran90言語の特長をまとめると次のようになります。

・配列に対する直接の演算や代入
・動的割り付け／解放とポインタ
・構造体を利用するための構造型プログラム単位で共通のデータの宣言や手続きを記述可能にするとともに、仕様の一部を外から隠ぺいすることを可能とするモジュール
・内部手続きや手続きの再帰呼び出し

従来のFORTRAN77から大きく拡張されていますが、上位互換性があるため、既存のFORTRAN77プログラムの翻訳も可能です。

②最適化機能

「ループの自動アンローリング」や「手続きのインライン展開」などに加えて、手続き間にまたがってデータの参照関係を追跡する「手続き間解析」機能が強化され、その結果を並列化以外の種々の最適化にも利用できるようになっています。またSX-4シリーズのスーパースカラを十分に活かすために、「命令の並べ替え」機能を大幅に強化しています。

③自動ベクトル化機能

「多重ループの一重化や入れ替え」、「隣接するループの融合」、「外側ループのベクトル化」、「条件ベクトル化」など、FORTRAN77/SXで定評のある自動ベクトル化機能を受け継いでいます。なお、これらは、DOループだけでなく、Fortran90言語で追加された配列式や配列代入文に対しても適用されます。

④自動並列化機能

「手続き呼び出しを含むループの並列化」や「条件並列化」など、これもFORTRAN77/SXで定評のある自動並列化機能を受け継いでいます。なお、従来と同様に、自動並列化はマイクロタスクの上に構築されていますが、最大32台という多数のプロセッサを用いた場合でも十分な効率が得られるように方式を改善し、マイクロタスクの入れ子を可能とするとともに、オーバヘッドを最小にしています。

⑤ XMU 配列機能

拡張記憶装置(XMU)上に配列を割り当てる機能が追加されています。

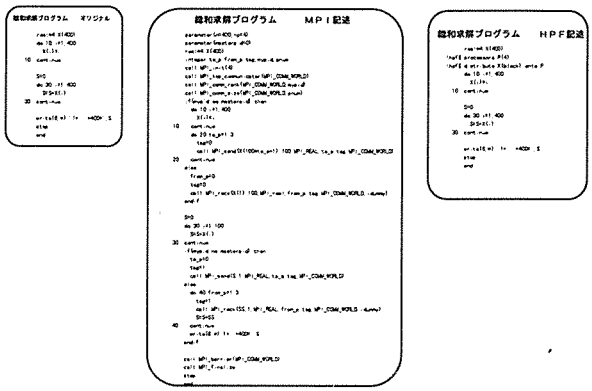
XMU に割り当てたい配列を共通ブロックまたはモジュールに置き、その名前を翻訳時のオプションで指定するだけで利用できます。

⑥クロスコンパイラのサポート

ワークステーション上で動作するクロスコンパイラも提供されています。したがって、手元のワークステーション上で翻訳・リンク処理まで行うことができます。

HPF/SX

HPF は、分散メモリ型の並列処理を指向した Fortran90 の拡張仕様の名称で、世界中から並列計算機のメーカやユーザが集まって標準化作業を行っているものです。HPF では、分散メモリに対するデータの配置をユーザに指定させ、それに基づいてプログラムの並列化を行うという「データパラレル」の考え方が採用されています。配置の指定は、指示行の形で行います。SX-4 シリーズでは、この HPF をサポートしています。



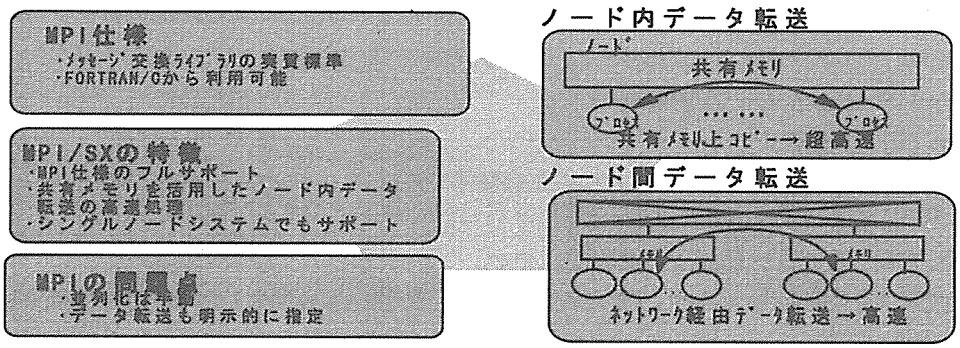
MPIとHPFのｺｰﾃﾞｨﾝｸﾞ例

MPI/SX

分散メモリ型の並列処理のためのもう１つのアプローチは、ユーザに分散メモリを完全に意識させ、タスク構造の定義から同期処理、分散メモリ間のデータの転送まで、すべてを明示的に行わせようという「メッセージ交換」の考え方です。

こちらも、HPF と同様に、世界中から並列計算機のメーカやユーザが集まって、プログラムから使用されるライブラリのインタフェースの標準化が行われました。これが MPI です。SX-4 シリーズでは、この MPI に基づいたライブラリを提供します。

MPI/SX では、分散メモリ型のマルチノードシステムだけでなく、シングルノードシステムにおいても共有メモリ内でデータ転送のオーバーヘッドが非常に小さいことを利用し、最適な処理を行うように考慮し、スケーラビリティの高い並列化を実現しています。



ＭＰＩ/ＳＸの特徴

充実したプログラミング開発環境

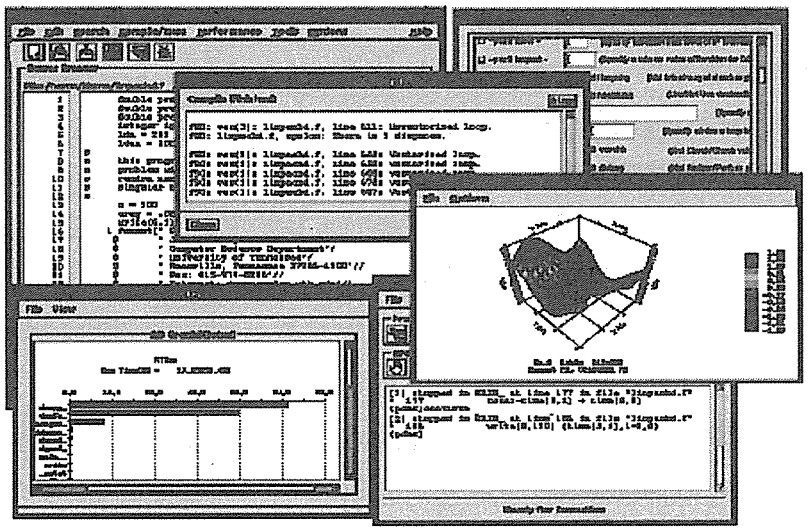
SX-4シリーズでは、ネットワーク上でスーパーコンピュータとワークステーションが接続された環境で、より快適かつより効率的にプログラミングが行えるように、次のような特長を備えた統合開発環境ツールＰＳＵＩＴＥを提供しています。

・ソース編集、翻訳、実行、デバッグ、及びチューニングといった一連の作業を統一的で使い易いＧＵＩで提供しています。

・スーパーコンピュータの負荷を軽減するために、プログラムの実行、デバッグ以外はクロスコンパイラと一体となってワークステーションで処理することができます。

・最適化、ベクトル化及び並列化されたプログラムに対してもソースレベルでデバッグ・チューニングが可能です。

・ＦＯＲＴＲＡＮやＣ言語のみならず、分散メモリ型並列処理にも対応しています。



統合開発環境ツールPSUITEの画面例

■おわりに

このようにSX-4シリーズは、超高速化、低価格化などのユーザニーズに応えて、1GFLOPS～1TFLOPSまでの1000倍の性能レンジをスケーラブルにカバーし、ユーザにとって使いやすいスーパーコンピュータを目指して開発を行ってきました。

今後も多様なニーズを十分に把握し、使いやすいスーパーコンピュータを提供していく予定です。

（備考）UNIX：X/Openｶﾝﾊﾟﾆｰﾘﾐﾃｯﾄﾞが独占的にライセンスしている米国ならびに他の国における登録商標です。

IBM ：International Business Machines Corporationの登録商標です。

CRAY：Cray Research,Inc.の登録商標です。

X-Window：MITの商標です。

Motif：Open Software Foundation,Inc.の登録商標です。

Ethernet：米国Xerox社の登録商標です。

NFS ：米国Sun Microsystems社の登録商標です。

NQS ：NASA Ames Research Centerのために Stering Softwareが開発した Network Queueing Systemです。