

汎用コンピュータシステム「TX7/AzusA」

日本電気株式会社 コンピュータ事業部 鈴木重信 高木均 横山淳

1. まえがき

TX7/AzusA は米 Intel 社の最新 64bit マイクロプロセッサ Itanium™ を 16 個まで搭載可能な大型スケーラブルサーバです。高性能 Itanium™ プロセッサとこれを相互に接続し 16way のスケーラビリティを実現する独自チップセットの搭載により、51.2 GFLOPS という 16way スカラサーバとして世界最高(2001 年 5 月の製品発表時点で)性能を達成しています。

TX7/AzusA は NEC 独自のチップセットとサービスプロセッサ、BIOS、OS の関係により高性能、高信頼性、高可用性を同時に実現する、システムアーキテクチャを有します。このアーキテクチャにより、科学技術計算からビジネス向け用途向けに至るまで、幅広い用途に対応できるプラットフォームとなっています。

本稿では、このような特長を備えた TX7/AzusA システムのハードウェアについてご紹介いたします。

2. システム構成

図 1 に汎用コンピュータシステム TX7/AzusA のシステム構成を示します。

汎用コンピュータシステム TX7/AzusA はノードあたり 16CPU から構成され、全体で 7 ノードから成るシステムです。各ノードは 1000BASE-SX スイッチを経由し接続されており、ファイルサーバ、スーパーコンピュータシステムと連携した利用が可能となっています。

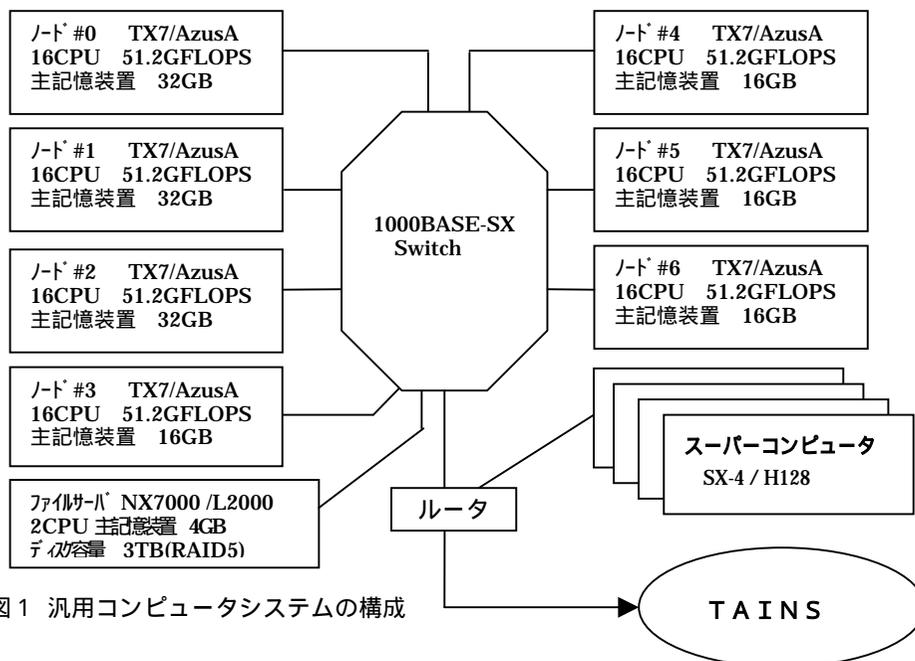


図 1 汎用コンピュータシステムの構成

3 . ハードウェア構成

3 . 1 特長

表1にTX7/AzusAの諸元を示します。特長を要約すると以下のようになります。

- (1) 最大16個の64ビットプロセッサItanium™を搭載可能なSMPサーバです。
- (2) メモリは最大64GB搭載可能で、システム総合転送性能は16.8GB/sに達します。
- (3) IOスロットとして最大128スロットのPCIスロットをサポートします。
- (4) 短いメモリレイテンシと高いメモリ/IO転送性能を実現し、高性能を追求しています。
- (5) プロセッサ/メモリ/IOの構成に柔軟に対応可能なセルアーキテクチャを採用しています。
- (6) セルを構成単位とするパーティショニング機能をサポートし複数OS搭載による柔軟なシステム構成を実現可能です。
- (7) データパス系にはECCを付与し、アドレス、制御信号系にはパリティなどを備えて障害検出能力を向上させ、高信頼度システムの構築に対応しています。
- (8) サービスプロセッサを備え、障害情報の収集、障害発生箇所の指摘、障害箇所の自動切り離し/自動再立ち上げを実現しています。
- (9) 冗長構成の電源/ファンを採用する他、各所でホットプラグをサポートしており、可用性の向上を図っています。

表1 TX7/AzusA 1ノードの諸元

項目	諸元
プロセッサ	Intel Itanium™(800MHz-4MB L3 cache)
プロセッサ数	16 個
メモリ容量	32GB / 16GB
IOスロット数	最大 64
本体サイズ	600(W) × 1040(D) × 1800(H)mm
重量	420kg

3 . 2 基本構造

図2に本体装置のハードウェアシステム構成を示します。

本体装置は、プロセッサ/メモリ/IOコントローラを持つセルカード4枚を中核に、セルカード間を接続するアドレスネットワーク/データクロスバ、PCIアダプタボックス(PCIBOX)、およびサービスプロセッサ(SP)からなります。各セルカードは4プロセッサ、8GBのメモリの搭載と、4本のIO接続用専用インタフェースGiga Stream Link(GSL)介してPCIBOXの接続が可能です。

プロセッサとメモリ、GSLインタフェースをバランスよくセルカードに搭載したセルアーキテクチャを採用しているため、システムに要求される能力の向上に応じて、段階的にコストを抑えた増設で応えることが可能です。同じく、IOは8PCIスロットを備えたPCIBOX単位での増設が可能となっています。

このように、厳密には4枚のセルカードに分散されたcc-NUMA(Cache Coherent Non-Uniform Memory Access)型の構造をしていますが、後述するようにソフトウェアから見たメモリアクセス時の振る舞いは16プロセッサのUMA(Uniform Memory

Access)型のSMP (Symmetrical Multi Processor)システムに近いものとなっています。このため比較的容易に大規模構成システムの性能を引き出すことが可能となります。

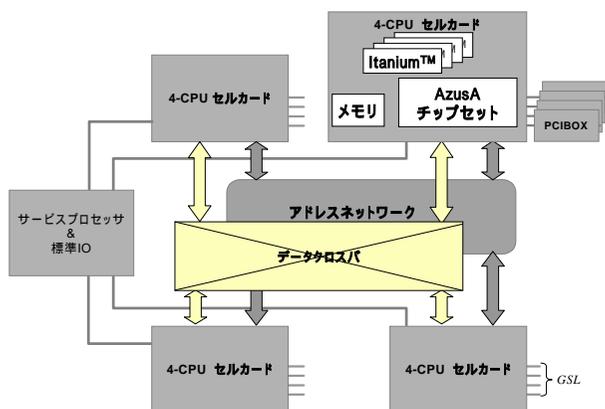


図2 TX7/Azusa System Block Diagram

3.3 Itanium™プロセッサの特長

Itanium™プロセッサは、Intel社が開発した新しい64ビットアーキテクチャであるItanium™アーキテクチャを採用した最初のプロセッサです。表2にItanium™プロセッサの仕様一覧を示します。

表2 Itanium™の仕様

項目	仕様
汎用レジスタ数	整数 128、浮動小数点 128
パイプライン段数	10 段
実行ユニット	ALU × 4 分岐ユニット × 3 単精度浮動小数点ユニット × 2 倍精度浮動小数点ユニット × 2
アドレス空間	仮想メモリ空間 16PB(16 × 10 ¹⁵ B)(54bit) 実メモリ空間 16TB(16 × 10 ¹² B)(44bit)

Itanium™アーキテクチャは命令の並列処理を効率化することで高性能を達成しており、その特徴としては「明示的に並列性を記述した命令のサポート」、「命令レベルの並列性をさらに強化する機能のサポート」の2つをあげることができます。

明示的な並列性を記述するために、Itanium™アーキテクチャでは3つの命令と3つの命令の組み合わせを表現するテンプレートからなる「命令バンドル」という概念が採用されており、プロセッサは命令をバンドル単位で実行していきます。Itanium™プロセッサは1サイクル当たり2バンドルを同時実行可能で、バンドル当たり1つの倍精度浮動小数点演算命令を含むことが可能です。加えて、積和演算命令をサポートしているため1サイクルあたり4つの倍精度浮動小数点演算が可能で、800MHzプロセッサのピーク性能は3.2GFLOPS(倍精度)となります。(単精度浮動小数点演算の場合はこの2

倍のサイクル当たり8演算実行可能で、ピーク性能は6.4GFLOPS(単精度)となります。) プロセッサには複数の命令バンドルを同時に効率良く実行するための多数のレジスタや実行ユニットなどの大量のリソースを装備しています。

命令レベルの並列性をさらに強化するために、Itanium™プロセッサは、メモリレイテンシの影響を最小限に抑えるための機能である「スペキュレーション」、命令の分岐を排除するための機能である「プレディケーション」、さらには分岐予測による分岐コストの削減といった先進的な機能を豊富にサポートしています。

これらの命令の並列処理を効率化する工夫により、Itanium™は高い処理性能を実現しています。

3.4 Itanium™プロセッサのキャッシュ構成

一般的なスカラ型マイクロプロセッサと同様に Itanium™プロセッサもキャッシュメモリの有効活用が性能向上のかぎとなります。そこで表3 および図3に Itanium™プロセッサのキャッシュメモリの構造をご紹介します。

表3 Itanium™プロセッサのキャッシュ詳細仕様

	Level1 (命令)	Level1 (データ)	Level2	Level3
容量	16KB	16KB	96KB	4MB(2MB 版も有)
実装場所	On die	On die	On die	パッケージ内
ラインサイズ	32Byte	32Byte	64Byte	64Byte
アソシビティ	4way	4way	6way	2way
書き込み方式	Read only	Write through	Write back	Write back
レイテンシ (整数)	2 サイクル	2 サイクル	6 サイクル	21 サイクル
レイテンシ (浮動小数点)	-	2 サイクル	9 サイクル	24 サイクル

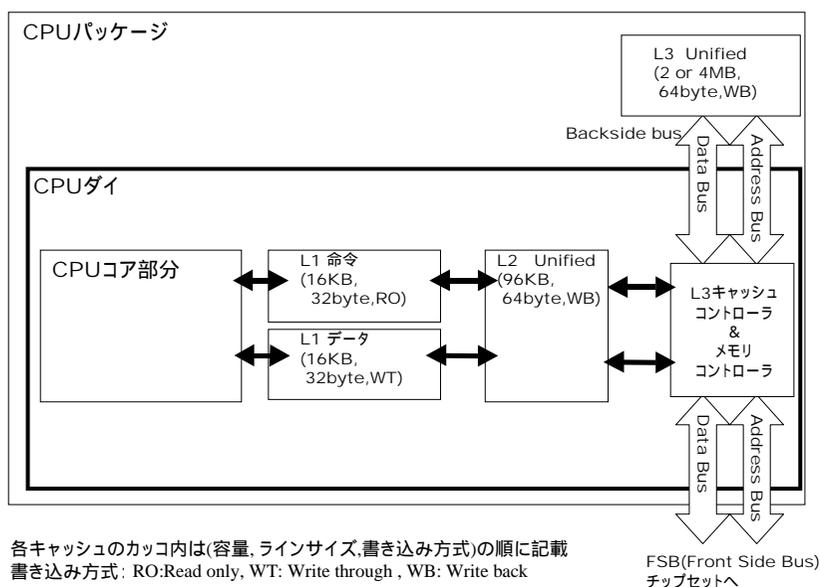


図3 Itanium™プロセッサのキャッシュ構成

3.5 短いメモリレイテンシの実現

TX7/AzusAはメモリレイテンシが短い点も大きな特長の一つです。あるプロセッサからのメモリレイテンシは、自セルカード内のメモリからのデータ読み出しの場合で200ns以下、他セルカードのキャッシュメモリまたはメモリからの読み出しの場合でも300ns以下を実現しています。この、およそ1:1.5という両者の比率は、既存の一般的なcc-NUMA型サーバ(1:3より悪い場合が多い)に対し大幅に改善されており、そのメモリアクセスにおける振る舞いはSMP型サーバに近いものとなっています。このため、cc-NUMAを意識したプログラミングテクニックを使用せずとも、比較的容易に性能をひき出すことが可能となっています。もちろん、cc-NUMA向けのプログラミングテクニックも利用することにより一層の性能発揮も可能です。

3.6 高い転送性能の達成

図4にTX7/AzusAの各所のバンド幅を示します。

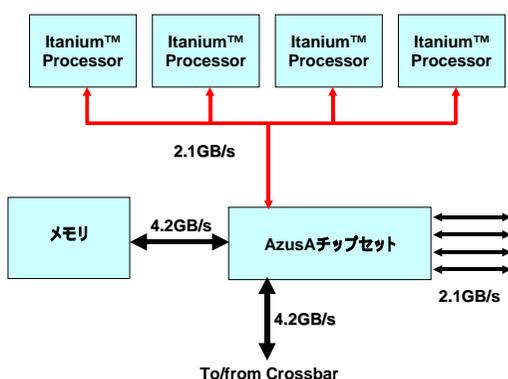


図.4 セルカード内のバンド幅

プロセッサバスとIOからのバンド幅はともに2.1GB/sでバランスしており、さらにメモリアクセスバンド幅はその合計を処理できる4.2GB/sを満たしています。またセルカード間も2.1GB/sのパスを双方向に備え、バンド幅が処理のボトルネックとならないように構成されています。

なお、セルカード内のメモリは最大4wayインタリーブが可能なメモリブロック2つからなり、セルカード内のインタリーブ数は8wayに達し、セルあたり4.2GB/sのメモリバンド幅を十分に活かすよう設計されています。

3.7 IO構成

TX7/AzusAのIOシステムは、基本IOとシステム構成に応じてPCIカードを増設するための拡張IO機能を提供する複数のPCIBOXから構成されます。PCIBOXには64ビット、66MHzタイプで64スロットをサポートすることが可能です。

4. オペレーティングシステム

TX7/AzusAでは、オペレーティングシステムにLinuxを採用しています。

TX7/AzusA の Linux は、オープンなコミュニティである IA-64 Linux プロジェクトが移植したカーネルをベースにしており、大容量メモリ(最大 64GB)、大容量ファイルストレージのサポート、SMP 向け機能強化など、大規模サーバとして必要な機能強化が行われています。

また、Linux ディストリビューションとしては、RedHat Linux for Itanium Processor を採用しています。同パッケージには IA-64 向けに多数のオープンソースのパッケージが含まれております。

5 . むすび

以上、汎用コンピュータシステム TX7/AzusA について述べました。今後も、科学技術用途における大規模高速演算のご要求や、IT 社会基盤の構成要素としてのサーバの重要度は増す一方です。弊社は継続して IA-64 サーバを開発することにより、さらに高い処理性能、拡張性、信頼性を実現する製品を提供してまいります。

写真 1 . TX7/AzusA 本体外観



写真 2 . セルカード外観

