

## [大規模科学計算システム]

### Express5800/A1080a-D のハードウェア

那須 康之<sup>1</sup> 鈴木 健一<sup>1</sup> 谷岡 隆浩<sup>2</sup>

#### 1. はじめに

Express5800/A1080a-Dは、米インテル®社の最新64bit 8コア インテル® Xeon® プロセッサをノードあたり4台(32CPUコア)搭載可能なccNUMA構造のスカラ並列型演算サーバです。1つの筐体内に2つのノードをコンパクトに収容した製品です。Express5800/A1080a-Dはメインフレームやスーパーコンピュータの技術とノウハウを取り入れた、NEC独自開発のBIOSやファームウェア、サービスプロセッサにより高性能、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ています。本稿では、Express5800/A1080a-Dシステムのハードウェアについてご紹介します。

#### 2. ハードウェア構成

##### 2.1. 特長

表1にExpress5800/A1080a-D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) ノードあたり最大4台の64ビット 8コア インテル® Xeon®プロセッサを搭載可能なccNUMA(Cache Coherent Non-Uniform Memory Access)アーキテクチャサーバです。
- (2) メモリはノードあたり最大1TB搭載可能です。
- (3) IOスロットとしてノードあたり最大7スロット(うち1スロットは内蔵HDDコントローラのスロットと兼用)のPCI-Express 2.0 スロットをサポートします。
- (4) 低いメモレイテンシと高いメモリ/IO転送性能により、高いシステム性能を追求しています。
- (5) プロセッサ間を高速で広帯域のインターフェイスで直結したグルーレスアーキテクチャを採用しています。
- (6) 高速インターフェイスにはトランザクションリトライ機能によって間欠故障を救済する機能を備え、高信頼システムの構築に対応しています。
- (7) 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。
- (8) 冗長構成の電源/ファンを採用する他、電源/ファン/HDDの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。

---

<sup>1</sup> 日本電気株式会社 IT ハードウェア事業本部 システム製品技術グループ

<sup>2</sup> 日本電気株式会社 IT ハードウェア事業本部 サーバ事業部

**表1 Express5800/A1080a-D 諸元**

項目	諸元
プロセッサ	8 コア インテル® Xeon® プロセッサ 最大 4 台 (32CPU コア) × 2 ノード
最大メモリ容量	1TB (16GB DIMM 使用時) × 2 ノード
標準 IO スロット数	PCI-Express 2.0 × 7 スロット × 2 ノード
筐体サイズ	448(W)×842(D)×307(H) mm (突起物、レール含む)
重量	79.6Kg (レール含む)

写真1にExpress5800/A1080a-D (2ノード) の外観を示します。



**写真1. Express5800/A1080a-D (2ノード) 外観**

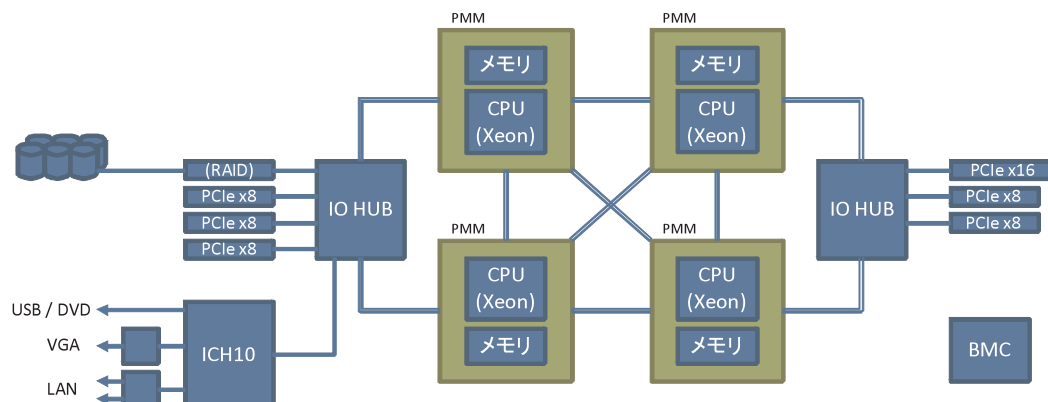
## 2.2. 基本構造

図1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、プロセッサとメモリを搭載するプロセッサ・メモリモジュール (PMM) をノードあたり最大4式と、メインボード上に配置されるIO HUBおよびIOコントローラハブ (ICH10) を主要部品として構成されます。各々のプロセッサおよびIO HUB間はインテル® Quick Path インターコネクトによりポイント・トゥー・ポイント接続されており、最短経路でCPU間のデータ送受信が可能な論理構造になっています。

プロセッサ・メモリモジュールには、モジュールあたり16枚のDDR3 DIMMを搭載でき、8GB容量のDIMMを用いた場合ノードあたり512GBまで、16GB容量のDIMMを用いた場合はノードあたり1TBまで拡張可能です。

I/OスロットはノードあたりPCI Express Gen2 x8スロット6本、x16スロット1本がメインボード上に配置されます。InfiniBand 4x QDRや10G Ethernetのような高速インターフェイスボードを利用する際にも十分余裕のあるスループットを確保しており、HPCC(High Performance Computing Clusters)やデータセンター内のクラスタシステム間の接続など、高速な通信イン

ターフェイスが求められる分野での利用にも適しています。なお、装置内に内蔵ハードディスクを搭載する場合には、x8スロットのうち1本にSAS RAIDカードを搭載して内蔵ハードディスクへの接続に用います。



**図1 Express5800/A1080a-D System Block Diagram(1ノードあたりの構成)**

### 3. Xeon® プロセッサ 7500番台

#### 3.1. Xeon® プロセッサ 7500番台の特長

Express5800/A1080a-Dは、米インテル®社製のインテル® Xeon® プロセッサ 7500番台を搭載し、サーバあたり最大4台(32CPUコア)搭載可能です。表2にXeon® プロセッサ 7500番台のなかで 最上位プロセッサであるX7560の仕様を示します。

**表2 インテル® Xeon® プロセッサ X7560 の仕様**

項目	仕様 (X7560)
コア種別	Nehalem コア (45nm High-k プロセス)
コア数	8 コア
マルチスレッド対応	対応 (コアあたり 2 スレッド)
汎用レジスタ数	整数 16、浮動小数点 8
発行ユニット	4 命令(micro-ops)/サイクルを発行
実行ユニット	6 命令(micro-ops)/サイクルを実行 <ul style="list-style-type: none"> <li>・ 3 メモリオペレーション</li> <li>・ 1 ロード</li> <li>・ 1 ストアアドレス</li> <li>・ 1 ストアデータ</li> <li>・ 3つのALU(演算装置)機能</li> </ul>
アドレス空間	仮想メモリ空間 48bit 実メモリ空間 44bit

図2にXeon® プロセッサ 7500番台の内部構造を示します。

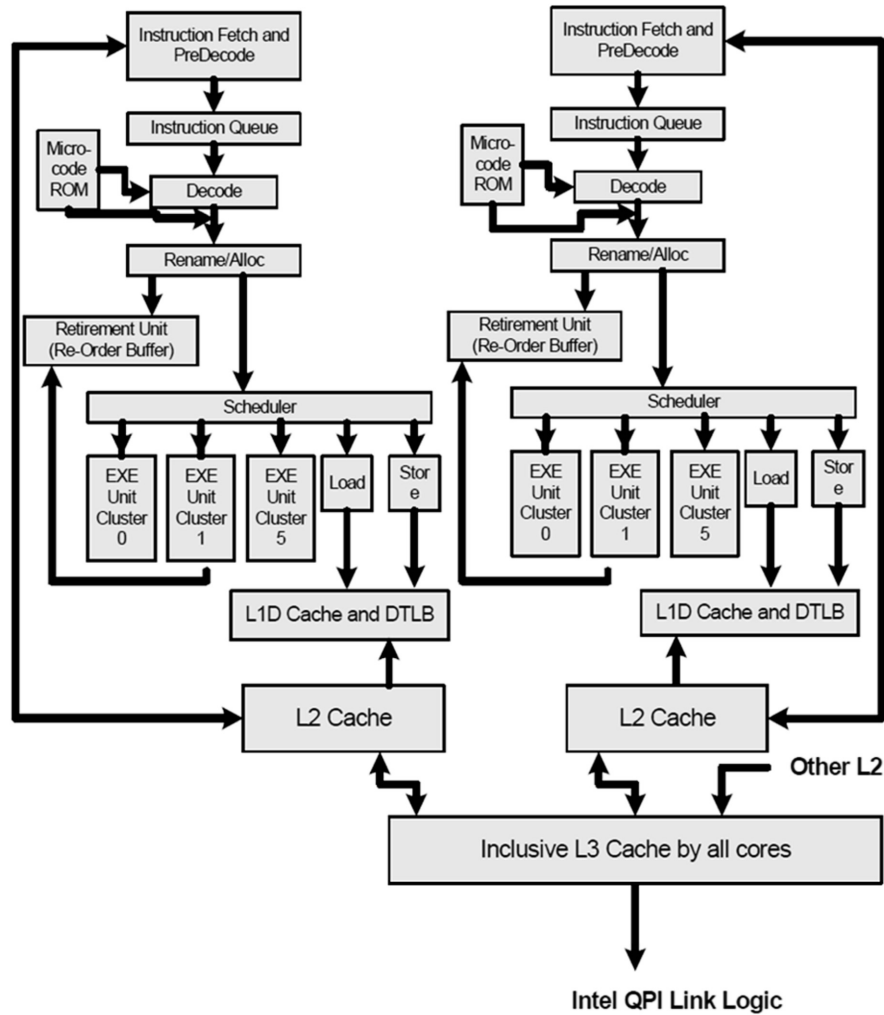
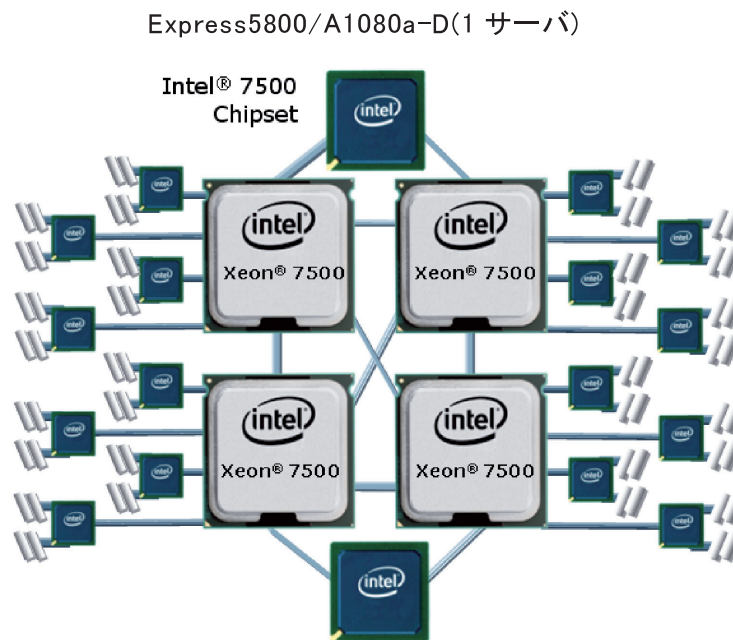


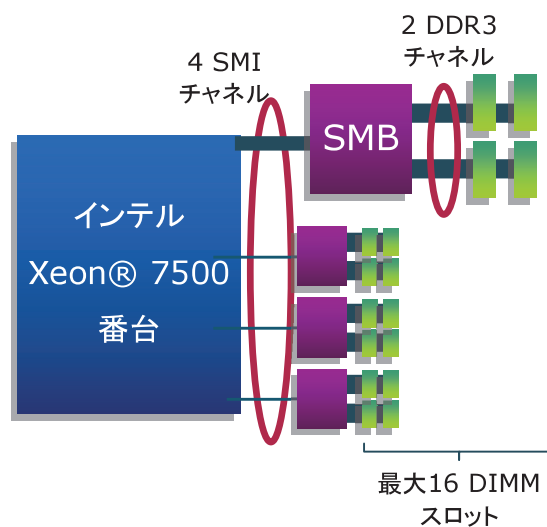
図2 インテル® Xeon® プロセッサ 7500番台の内部構造

MP系(4ソケット以上)のXeon® プロセッサとして初めて、メモリコントローラの内蔵や、これまでのFSB(Front Side Bus)を置き換えるインテル® Quick Path インターコネクタに対応し、広帯域幅かつ低レイテンシでのポイント・トゥー・ポイント接続を実現しました。インテル® Quick Path インターコネクタは、最高6.4GT/sの転送レートで、1リンクあたりの帯域は25.6GB/sに達します。図3に、Express5800/A1080a-Dのトポロジを示します。



**図3 Express5800/A1080a-Dのトポロジ**

図4に示すように、プロセッサ内のメモリコントローラは、4つのSMI(Scalable Memory Interface)チャンネルを持ちます。SMIには2つのDDR3チャンネルを持つSMB(Scalable Memory Buffer)が接続され、SMBあたり最大4枚、プロセッサあたり最大16枚のDDR3 Registered-DIMMを接続可能で、これまでに無い広大なメモリ空間を実現します。



**図4 インテル® Xeon® プロセッサ 7500番台でのメモリ接続**

### 3.2. Xeon® プロセッサ 7500番台のキャッシュメモリ仕様

一般的なスカラ型マイクロプロセッサと同様にインテル® Xeon® プロセッサ 7500番台もキャッシュメモリの有効活用が性能向上の鍵となります。表3にXeon® プロセッサ 7500番台のキャッシュメモリ仕様を示します。

**表3 インテル® Xeon® プロセッサ (X7560)のキャッシュ詳細仕様**

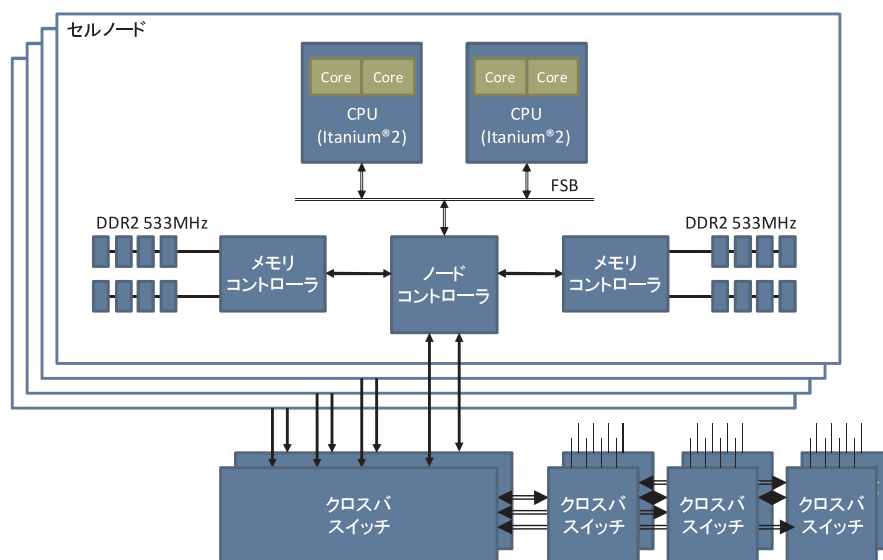
キャッシュ階層	形式	容量
L1 (命令)	コア毎	32KB
L1 (データ)	コア毎	32KB
L2 (命令・データ)	コア毎	256KB
L3 (命令・データ)	コア共有	24MB

## 4. ハードウェア性能

インテル®Xeon® プロセッサ 7500番台は、1クロックサイクルあたり最大4回の演算を実行可能なNehalemマイクロアーキテクチャと、単一CPUソケット内に最大8プロセッサコアを内蔵したマルチコアCPU構造を採用しています。これによるCPUソケットあたりの最大計算能力は72.3GFLOPSに上ります。Express5800/A1080a-Dは、ノードあたり最大4CPUまで搭載することができますので、ノードあたり最大289.3GFLOPS、従来の機種(TX7/i9610)との比較ではCPUコアあたりで1.4倍、CPUソケットあたりで5.6倍もの計算能力を有します。

メモリのスループットとレイテンシについても、Express5800/A1080a-D では従来のTX7/i9610 に比べても高い性能を得ることができます。

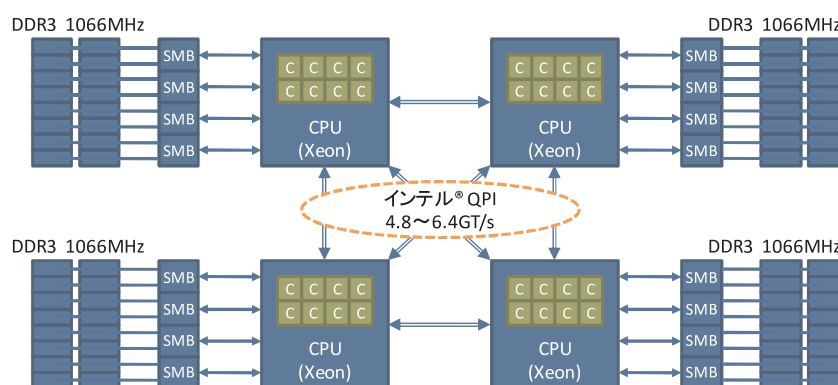
TX7/i9610 では、システムのスループットを最大化するため、図5に示すように2個のCPU毎にFSBを分割した論理構造を基本単位(セルノード)とし、最大32CPUまでの密結合マルチプロセッサシステムを構築できる構成を採用しています。



**図 5 TX7/i9610 のプロセッサ-メモリ間接続構成**

一方 Express5800/A1080a-D では、インテル® Xeon® プロセッサ 7500 番台がプロセッサコアとメモリコントローラを単一の CPU ソケット内に実装するため、図 6 に示すようにメモリは SMB(Scalable Memory Buffer)を介して CPU に直接接続される論理構造をとります。このため、プロセッサコアから同一ソケット内のローカルメモリへのアクセス時には DDR3 DIMM のスループットを十分に活用する事ができるとともに、ccNUMA に最適化されたアプリケーション環境下では CPU ソケット数に応じてシステムのスループットを伸ばすことも可能になります。

TX7/i9610と Express5800/A1080a-D の最大構成でのメモリのスループット(理論値)の比較では、プロセッサコアあたりで Express5800/A1080a-D が約 1.4 倍上回る結果が得られます。



**図 6 Express5800/A1080a-D のプロセッサ-メモリ間接続構成**

このような両者の論理構造の違いにより、Express5800/A1080a-Dでは良好なメモリレイテンシを得ることが可能になりました。プロセッサコアから最も近いローカルメモリまでのレイテンシで比較した場合、Express5800/A1080a-DのレイテンシはTX7/i9610の約40%(理論値)です(図7)。さらに、Express5800/A1080a-Dではインテル® Quick Path インターコネクトを介してリモートメモリへも1ホップでアクセスできますので、リモートメモリへのアクセスも高速です。

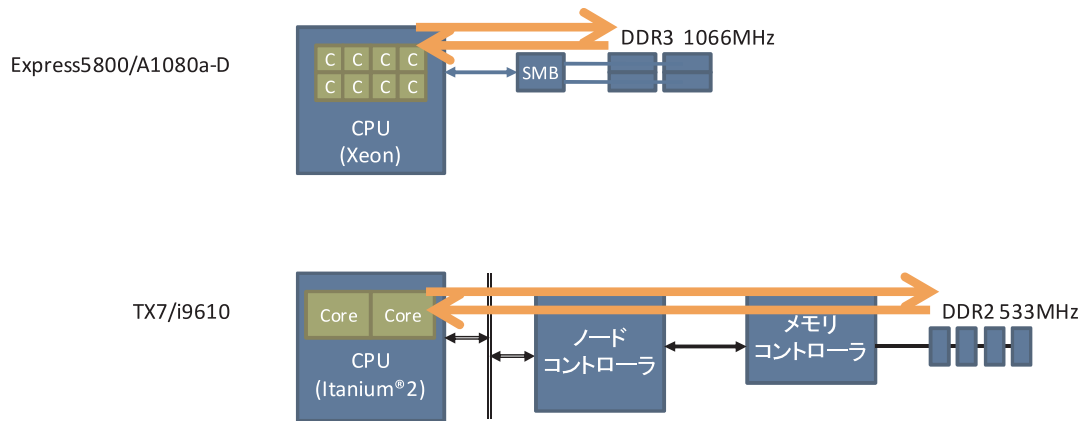


図7 ローカルメモリへのアクセス経路比較

加えて、Express5800/A1080a-DではccNUMAに対応しています。ccNUMAはメモリが物理的に分散して配置されプロセッサコアからメモリまでのレイテンシが均一でない装置構成で、キャッシュコヒーレンシを維持しながら各々のメモリを全てのプロセッサから共通の物理メモリとして利用できるようにするメモリアーキテクチャです。Express5800/A1080a-DではACPI(Advanced Configuration and Power Interface)にccNUMA対応の機能を実装しており、Linux OSとの組み合わせでccNUMAに対応したアプリケーションを利用する事ができます。

優れたメモリバンド幅とレイテンシを実現する論理構造と、ccNUMAに最適化可能な機能により、Express5800/A1080a-Dのアーキテクチャは従来の機種以上に高い性能を実現することができます。

## 5. むすび

以上、スカラ並列型演算サーバ Express5800/A1080a-Dについて述べました。科学技術用途における大規模高速演算のみならず、IT社会基盤の構成要素としても、今後ますますサーバの重要度は増していきます。NECは、最先端の技術を駆使して、さらに高い処理性能、拡張性、信頼性を実現するサーバ製品を継続して提供してまいります。