［大規模科学計算システム］

SX-Aurora TSUBASA のハードウェア

星 宗王　江副 健司　上山根 慎
日本電気株式会社

1. はじめに

東北大学サイバーサイエンスセンターで本年 10 月より新しいスーパーコンピュータ AOBA の運用が開始されます。AOBA はベクトル型システムであるサブシステム AOBA-A と、x86 システムであるサブシステム AOBA-B から構成されています。本稿では、AOBA-A を構成する SX-Aurora TSUBASA のハードウェアを紹介します。

近年、シミュレーション解析のニーズをはじめとする科学技術計算需要はますます増大しており、一方で大規模化するシステムに対して省電力化、省スペース化の要求もますます高まっています。この要請に応えるために、NEC は従来のベクトル型コンピュータを PCI Express カードのフォームファクタにパッケージし、ユーザが x86 ベースのサーバからシームレスにベクトル型コンピュータを活用できるシステムアーキテクチャを採用した、新しいスーパーコンピュータ SX-Aurora TSUBASA シリーズを 2018 年 2 月に出荷開始しました。

このたび東北大学サイバーサイエンスセンターに導入いただいたシステムは、前機種よりもメモリ性能をさらに強化し、2020 年から提供を開始した第二世代製品になります。

本稿では、AOBA-A を構成する SX-Aurora TSUBASA システムのアーキテクチャ、システム概要、ハードウェア構成、テクノロジについて紹介します。

2. Aurora アーキテクチャ

SX-Aurora TSUBASA は、NEC のベクトル型スーパーコンピュータ SX シリーズの流れを汲む製品として 2018 年から出荷開始した新しいタイプのスーパーコンピュータです。Vector Engine と呼ばれる PCI Express カードにベクトルプロセッサと主記憶を搭載し、これを Vector Host と呼ばれる標準的な x86 サーバに接続することによってシステムが構成されます。

Vector Engine に搭載されるベクトルプロセッサは、従来の SX シリーズのベクトルプロセッサ構成を踏襲しています。通常、サーバに接続される GPU や FPGA のような PCI Express カード型のアクセラレータは、ホスト側で実行されるアプリケーションの一部分を実行し、全体の処理時間の短縮を図ります。一方 Vector Engine は、コンパイルされたアプリケーション実行ファイルを丸ごとカード上で実行する、Aurora アーキテクチャを採用しています。図 1 に Aurora アーキテクチャにおける実行モデルを示します。

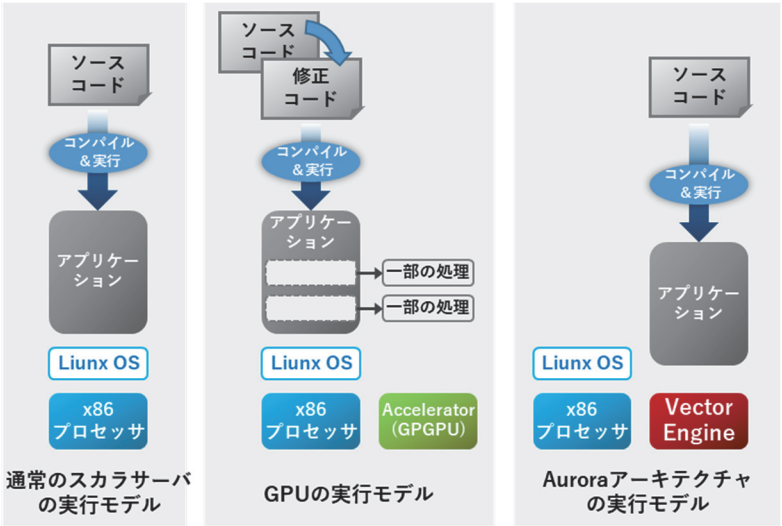

図 1. Aurora アーキテクチャの実行モデル

Aurora アーキテクチャでは、原則としてアクセラレータ向けのプログラム修正を行うことなく、SX-Aurora TSUBASA 向けコンパイラを用いてコンパイルしたプログラムをそのまま実行することができます。(さらに高い実行性能を得るためには、Vector Engine 向けのコードチューニングを実施することが望ましいです。)

この実行モデルにはいくつかのメリットがあります。第一に、ユーザにとって使い慣れた一般的な Linux OS 環境から、SX シリーズの高性能なベクトルプロセッサを利用できます。SX シリーズでは従来、専用の OS 環境を提供していましたが、SX-Aurora TSUBASA は標準的な x86 サーバと Linux OS 環境から演算処理だけにベクトルプロセッサを利用することができます。第二に、プログラム全体を Vector Engine カード上で実行することにより、PCI Express バス上の頻繁なデータ移送を避け、性能ボトルネックを解消することができます。

3. システム概要

AOBA-A は、Vector Engine カードを 8 枚搭載した Vector Host(SX-Aurora TSUBASA B401-8)計 72 台で構成され、各 Vector Host は InfiniBand Network によって接続されます。

システムの主要な諸元を表 1 に示します。

表 1. AOBA-A の諸元

Vector Engine	モデル名	Type 20B
	CPU コア数	8
	理論ベクトル演算性能	2.45 TFlops(DP)
	メモリ容量	48 GB
	メモリ帯域	1.53 TB/s
	インタフェース	PCI Express Gen3 x16
Vector Host	CPU モデル名	AMD EPYC 7402P
	CPU コア数	24
	メモリ容量	256 GB
	理論演算性能	1.075TFlops(DP)
	OS	CentOS
	搭載 Vector Engine 数	8
	ネットワーク I/F	InfiniBand HDR x2
システム	Vector Host 数	72
	Vector Engine 数	576
	CPU コア数	6,336
	総理論演算性能	1,488.6 TFlops
	総メモリ容量	45TB
	総メモリ帯域	895.68TB/s

4. ハードウェア構成

4.1 Vector Engine Type20B

AOBA-A は最新の Vector Engine Type 20B を合計 576 枚搭載しています。Vector Engine は前述したように PCI Express カードのフォームファクタにベクトルプロセッサと主記憶を搭載した SX-Aurora TSUBASA の心臓部です。図 2 に Vector Engine カードの外形を、表 2 にカードの実装仕様を、図 3 にブロック図を示します。

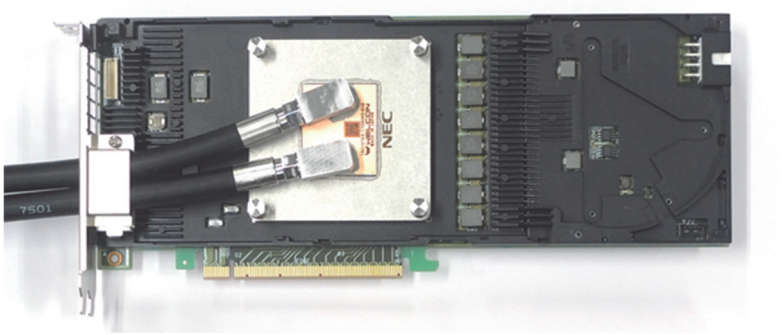

図 2. Vector Engine カード（水冷タイプ）

表 2. Vector Engine カード実装仕様（水冷ホースを除く）

Board Length	10.5 inch
Board Height	4.376 inch（PCI Express CEM3.0 maximum）
Z-Height of Component side	最大 1.370 inch（PCI Express CEM3.0×2slot 分）
Z-Height of Solder side	0.105 inch
Power Connector type	8-pin EPS 12V
Power Connector Location	EAST

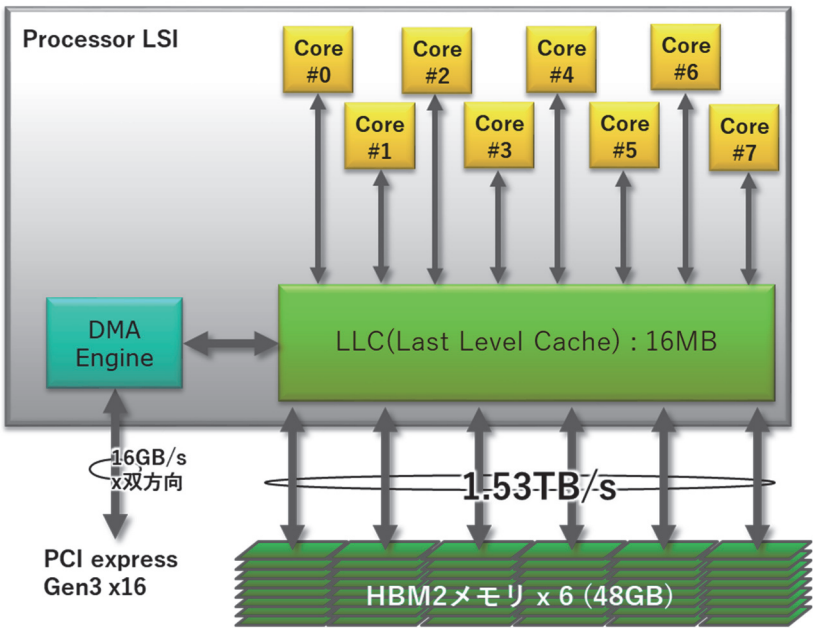

図 3. Vector Engine ブロック図

Vector Engine は 8 個の CPU コアとコア間で共有される LLC(Last Level Cache)、DMA エンジン、PCI Express インタフェース、6 個の HBM2 メモリで構成されます。

各 CPU コアは、基本的には従来の SX アーキテクチャを継承しつつ、マイクロアーキテクチャレベルで多くの改良を加えた新しいベクトルプロセッサコアです。図 4 に CPU コアの構成を示します。

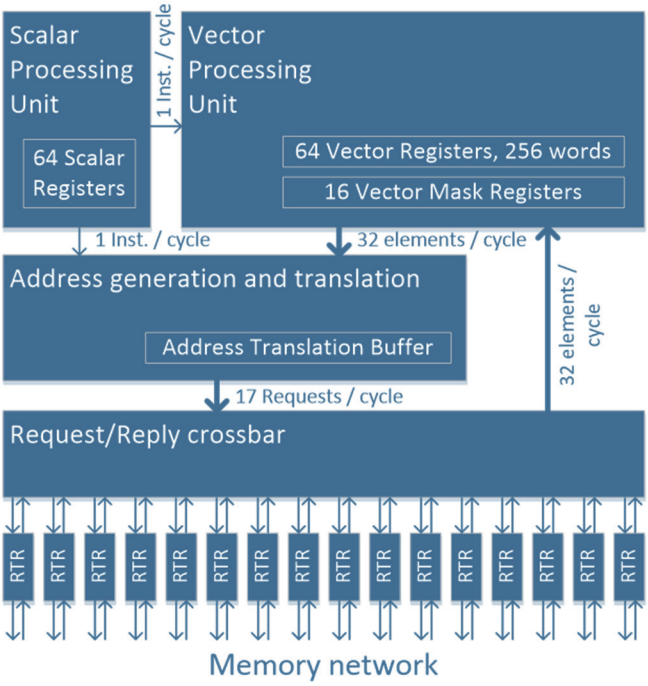

図 4. CPU コア ブロック図

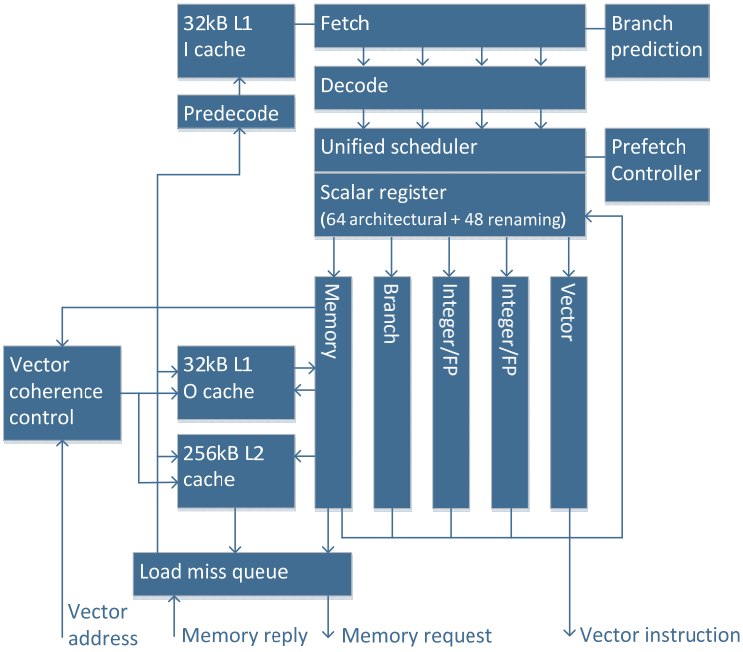

図 5. SPU ブロック図

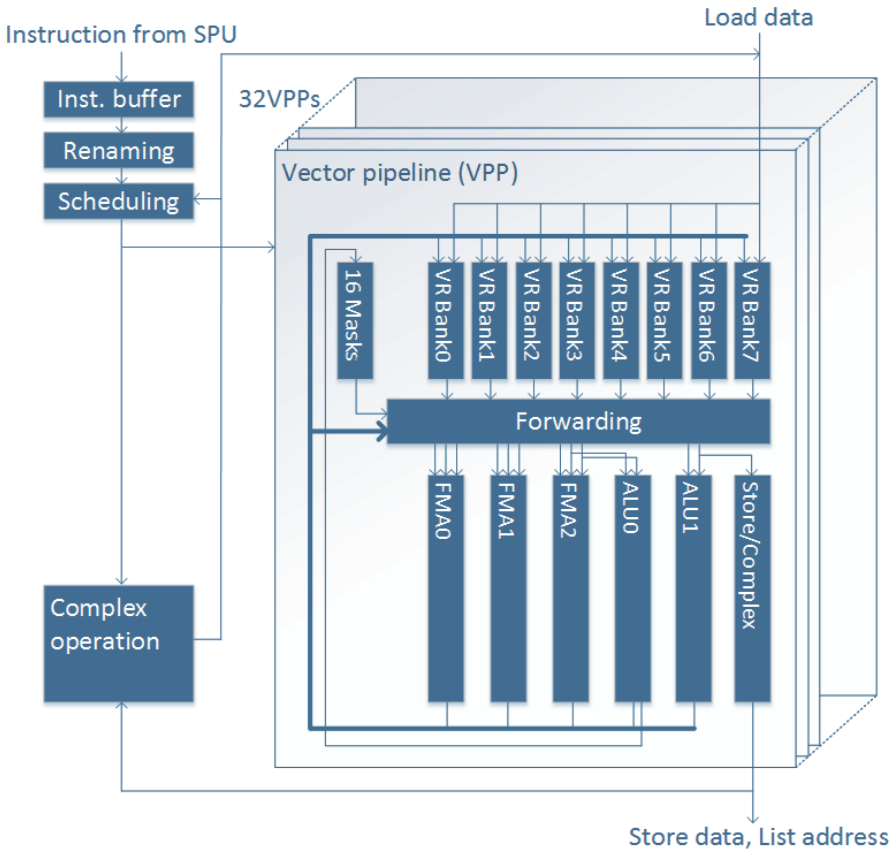

図 6. VPU ブロック図

CPU コアは、スカラ処理部(Scalar Processing Unit: SPU)、ベクトル処理部(Vector Processing Unit: VPU)の他、アドレス生成部、メモリネットワークによって構成されます。SPU は、すべての命令発行制御を行い、ベクトル命令やノード間通信命令を VPU や DMA エンジンへと発行します。1 サイクルに 4 命令の fetch/decode が可能で、VPU への専用パスを用いて 1 つのベクトル命令を発行可能です。32KB の L1 命令キャッシュ、32KB の L1 オペランドキャッシュ、256KB の L2 キャッシュを備えています。VPU は、32 本のベクトル・パイプラインから構成され、各パイプラインは最大で 3 つの 64bitFMA(浮動小数点積和演算)命令を同時に実行することが可能です。よって理論ベクトル演算性能は 32 パイプライン×3 セット×2(乗算＋加算)×1.6(GHz)=307.2GFlops となります。また、FMA 演算器は 32bit×2 セットの単精度 Packed 演算器としても動作することが可能であり、単精度浮動小数点演算性能は最大 714.4GFlops となります。各パイプラインは 1 サイクルに 64bit データのロードストアが可能であるため、コア辺りのメモリバンド幅は 8 バイト(64 ビット)×32 パイプライン×1.6GHz = 409.6GB/s(x2=ロード+ストア)となります。

コアあたりの論理ベクトルレジスタ数は 64 で、各ベクトルレジスタは 64bit の浮動小数点データを 256 要素格納できます。物理的には 256 本のベクトルレジスタを用意し、ベクトルレジスタのリネーミング機構を備えています。また、256 ビット長のマスクレジスタを 16 本備えています。

LLC は各コアで共有されるソフト制御可能なキャッシュメモリです。容量は 16MB で、ライトバック方式で制御されますが、SX-ACE の ADB(Assignable Data Buffer)に類似した機能として、ベクトル命令中のフラグによってデータ保持の優先度を指示することができます。

DMA エンジンは CPU コアと非同期に動作可能で、Vector Engine の主記憶の他、Vector Engine 内の各種レジスタ、Vector Host 内の主記憶にアクセス可能で、CPU コアだけでなく Vector Host の CPU からも利用可能です。

4.2 SX-Aurora TSUBASA B401-8

AOBA-A の構成要素(ノード)である SX-Aurora TSUBASA B401-8 の外観とブロック図をそれぞれ図 7 と図 8 に示します。B401-8 は Vector Host と 8 枚の Vector Engine で構成されます。Vector Host のプロセッサは AMD EPYC 7402P で、Vector Host あたり 1socket、24 コアを備えています。プロセッサの基本動作クロックは 2.8GHz で、シングルコアであれば最大で 3.35GHz までのブースト動作が可能です。プロセッサは 4 つの PCI Express switch と 2 つの InfiniBand HDR カードと直接接続されています。4 つの PCI Express switch はそれぞれ 2 つの Vector Engine カードと接続されています。プロセッサと InfiniBand HDR カードとの接続は PCI Express Gen4 x16 で、他の接続は PCI Express Gen3 x16 です。SX-Aurora TSUBASA B401-8 は Vector Host あたり 256GB の主記憶を備え、図 8 には図示されていませんが 960GB の SSD を備えています。

図 7. SX-Aurora TSUBASA B401-8 外観

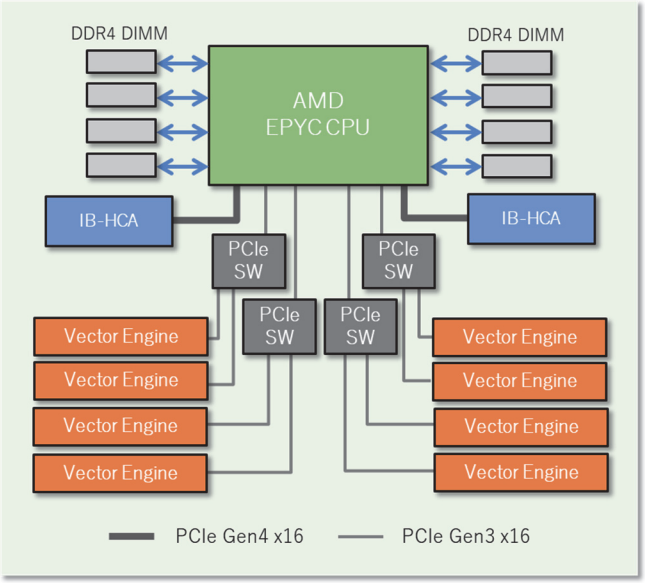

図 8. SX-Aurora TSUBASA B401-8 のブロック図

4.3 ノード間結合網

AOBA-A の 72 台の Vector Host は InfiniBand HDR による 2 段 Fat-Tree ネットワークによって接続されます。図 9 に InfiniBand ネットワーク構成図を示します。InfiniBand HDR スイッチとしては NVIDIA Mellanox QM8700 シリーズを利用します。Vector Host あたり 2 枚の HDR カードを介してエッジスイッチに接続されます。ノード間接続ネットワークは、AOBA-A だけでなく、AOBA-B、フロントエンドサーバなどの各種サーバやストレージを含めたシステム全体を、フルバイセクションバンド幅、ノンブロッキング構成により接続し、高速なデータ通信を可能とします。

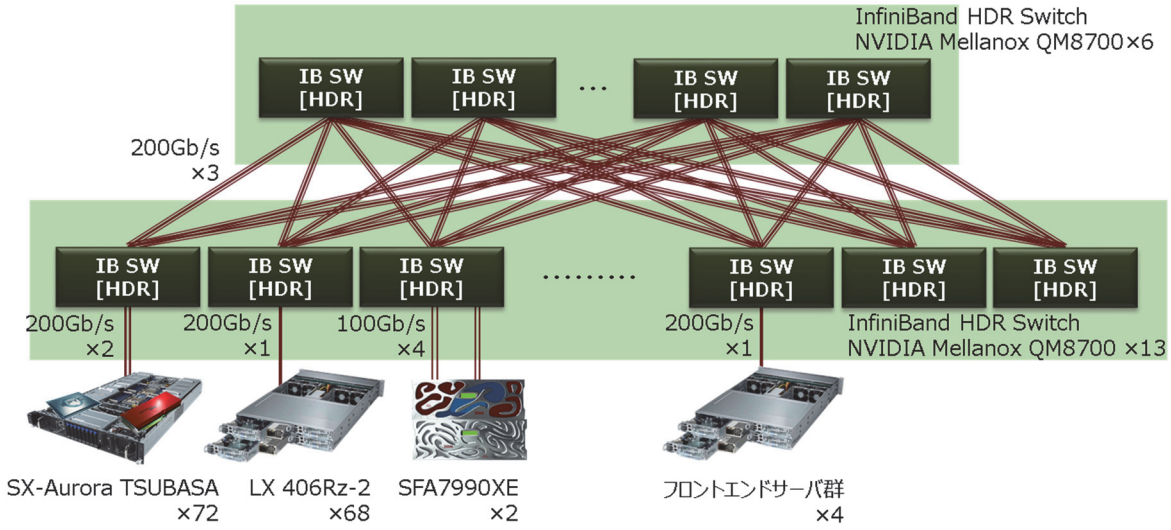

図 9. InfiniBand ネットワーク構成図

5. SX-Aurora TSUBASA のテクノロジ

5.1 LSI 技術

SX シリーズでは、従来から CMOS テクノロジによる高集積化を進め、SX-ACE では初めて 4 プロセッサを 1LSI に集積するなど、低消費電力と高性能の両立を目指した開発を行ってきました。AOBA-A の SX-Aurora TSUBASA システムに搭載している Vector Engine Type 20B では、最新の HBM2E メモリを利用することで LSI あたりのメモリバンド幅を 1.53TB/s に向上させています。LSI の外観を図 10 に、仕様を表 3 にまとめます。

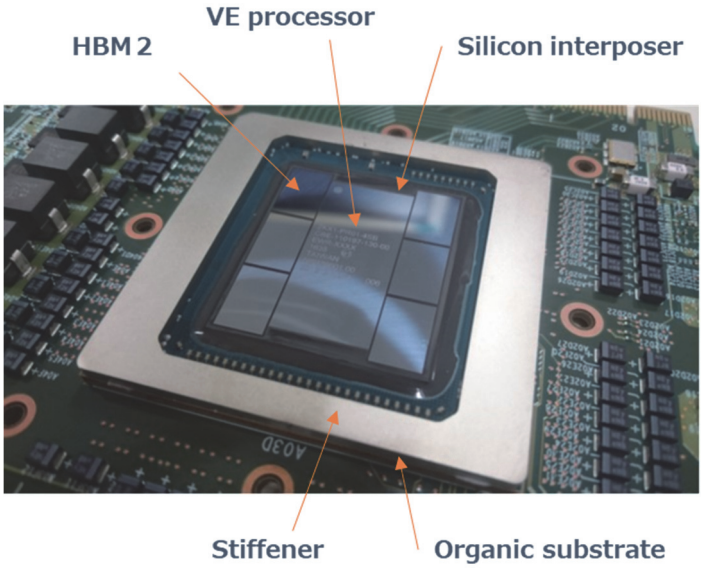

図 10. Vector Engine プロセッサ LSI 外観

表 3. CPU チップ諸元

テクノロジノード	16nm FinFET プロセス
Total gate count	274MGates
Total SRAM	283Mbits
コア電源電圧	0.89V
信号ピン数	10,676 IO total
ASIC(VE processor)サイズ	15mm×33mm
インタポーザサイズ	32.5mm×38mm
パッケージサイズ	60mm×60mm

5.2 実装/冷却技術

SX-Aurora TSUBASA B401-8 は、2U のユニットに 8 枚の Vector Engine カードを搭載しています。SX-Aurora TSUBASA B401-8 の水冷機構は in/out それぞれ 4 本の水冷ホースを備え、8 枚の Vector Engine カードと AMD EPYC 7402P 1 ソケットの計 9 つのプロセッサ LSI を冷却します。図 11 に B401-8 の内部構造図を、図 12 に B401-8 のラック搭載イメージを示します。

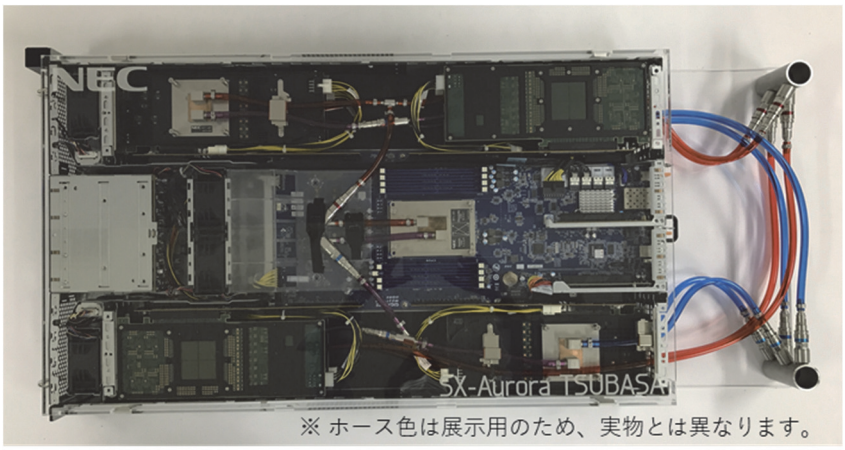

図 11. SX-Aurora TSUBASA B401-8 内部構造図

図 12. SX-Aurora TSUBASA B401-8 のラック搭載イメージ図

6. おわりに

このたび、AOBA-A に採用された SX-Aurora TSUBASA システムは、ラックあたりの演算性能を従来の SX-ACE システムと比較して 20 倍以上に高めるとともに、新しい Aurora アーキテクチャによって Linux OS 環境からベクトル型スーパーコンピュータをシームレスに利用できるようにしたことで、高性能と使いやすさの両立を実現することができました。この新しい SX-Aurora TSUBASA システム、そしてスーパーコンピュータ AOBA が、多くの方々の研究を前進させる一助になることと確信しています。NEC は今後も、様々な研究分野の発展を支えるため、高性能で使いやすいスーパーコンピュータの開発を継続していきます。