



新型スーパーコンピュータ SX-Aurora TSUBASA



日本電気株式会社
ITプラットフォーム事業部
第三ソリューション基盤統括部
主任

やまだ ようへい
山田 洋平



日本電気株式会社
ITプラットフォーム事業部
第三ソリューション基盤統括部
主任

いまい てるゆき
今井 照之

1. はじめに

近年、スマートフォンの普及やIoT (Internet of Things) の進展により、大量かつ多様なデバイスがインターネットに接続され、これらのデバイスを介した膨大なデータ収集、蓄積がされるようになった。これら蓄積データから新しい知見を得る、またはリアルタイムに制御や予測を行うといったビッグデータ解析や機械学習が高い注目を集めている。

NECは2017年10月に新型スーパーコンピュータSX-Aurora TSUBASAを発表した。SX-Aurora TSUBASAの特徴は大量データの高速処理である。このため、シミュレーション等を主とするスーパーコンピューティング用途のみでなく、ビッグデータ解析や機械学習の高速化用途が期待されている。本稿ではSX-Aurora TSUBASAの構成及び特徴について紹介する。

2. SX-Aurora TSUBASAのコンセプト

SX-Aurora TSUBASAは、NECが30年以上にわたり開発・販売を行ってきたベクトル型スーパーコンピュータSXシリーズの最新モデルである。

従来SXは非常に高いメモリアクセス性能及びコア性能により、気象予報や科学技術計算分野で高い実行性能を実現してきた。この性能特性は大量データの高速処理が必要となるビッグデータ処理に必要とされるものである。一方で、従来SXはユーザ向けのライブラリの種類に限られる点や、許容されるシステム構成が最小でもラック単位の規模である点など、ハイエンドHPC (High Performance Computing) 領域に特化した製品であり、ビッグデータ解析等の新領域適用に向けた課題があった。

本背景により、従来SXの高性能をより広範な分野に適用するために、SX-Aurora TSUBASAは以下の3つのコンセプトに基づき開発された。

① 高性能

従来SXの特徴である高メモリアクセス性能を継承するため、最先端メモリテクノロジーを採用しプロセッサ当たり、及び価格当たりで世界最大のメモリ帯域を実現。これにより、HPC、ビッグデータ解析、及び機械学習等の高メモリ帯域を要求するプログラムを高速化。

② 使いやすさ

新領域のユーザが違和感なく利用可能な環境を提供するために、広く一般に普及しているLinuxをベースとしたオペレーティングシステム (OS) を採用。Linuxと高い互換性のあるシステムコール・ライブラリの提供による高いプログラム移植性を実現。また、従来SXからベクトルアーキテクチャを継承し、ベクトル化コンパイラにより特殊なプログラミングモデルを使用することなく高実行性能を実現。

③ フレキシビリティ

システム構成の柔軟性を高めるために、ハードウェアの形態をPCI Expressカード型とし、x86/Linuxサーバに搭載。これにより、ワークステーションのような小規模モデルから、大規模スーパーコンピュータまでを幅広くカバー。

3. SX-Aurora TSUBASAのハードウェア

SX-Aurora TSUBASAのハードウェアは、PCI Express (PCIe) カードであるVector Engine (VE) 及びそれを搭載するx86サーバで構成される。本章では主にVEについてその構成及び特徴を紹介する。

3.1 Vector Engineカード

図1はVEカードの外観である。カード中央のCPUモジュールは、CPU LSI及びメモリを1モジュール実装としたものである。VEカード上には本CPUモジュールのほか、



電源回路や制御用IC等が置かれ、VEカード周辺部にPCIe Gen3 x16対応のエッジコネクタ及び外部給電端子を備える。VEカードの消費電力は300W以下で、カード形状とともにPCIe規格に準拠する。

VEカードは、搭載されるx86/Linuxサーバの環境に応じて以下の3種がある。カード上のファンによる空冷を行うアクティブクーリングタイプ、x86/Linuxサーバ内のファンにより空冷を行うパッシブクーリングタイプ、及びスーパーコンピュータ向けに高効率冷却を提供する水冷タイプである。

世界一のメモリアクセス性能を有するプロセッサをPCIeカード化し、冷却方式の高いフレキシビリティを実現したことにより、デスク上で利用可能なタワー型から大規模スーパーコンピュータまでの構成を可能としている。



■図1. Vector Engineカード (上: カード外観、下: 内部ボード)

3.2 Vector Engine CPUモジュール

VE CPUモジュールは、単一CPUとして世界一のメモリ帯域となる1.22TB/sを実現している。図2にVE CPUモジュールの外観を示す。メモリにHigh Bandwidth Memory 2 (HBM2) を採用し、CPU LSIと6個のHBM2を世界で初めて1パッケージ化することにより、世界一のメモリ帯域を実現している。

HBM2は広帯域が要求されるハイエンド領域をターゲットとして開発された新しいメモリ規格である。HBM2は4～8枚のメモリチップをThrough Silicon Via (TSV) と呼ばれるチップを貫通する電極で接続した3次元積層構造となっている。これに1024bit幅の外部インタフェースを備えることで、1個のHBM2あたり最大256GB/sの帯域を実現している。この値は現在一般的なx86サーバに使用されている

DDR4メモリ (PC4-21333、21.3GB/s) の10倍以上にあたる。

CPU/HBM2間は2.5次元 (2.5D) 実装を採用し、1個のCPUに6個のHBM2を超広帯域で接続している。2.5D実装とは、通常LSI間の接続に使われる有機材料ではなく、シリコン製のインタポーザによりLSI間を接続する実装形態である。2.5D実装により、LSI間の超高速接続及びLSI/インタポーザ間の熱膨張率等の物理特性が揃うことにより、接続部の機械的な信頼性が増す。さらには、CPUとメモリを近傍に配置することにより、高メモリ帯域を実現しながらメモリアンタフェースの消費電力を下げることができる。



■図2. Vector Engine CPUモジュール

3.3 Vector EngineのCPUアーキテクチャ

VE CPUモジュールに搭載されているCPU LSIは、従来のSXベクトルアーキテクチャを継承しつつ、SX-Aurora TSUBASA向けに新設計されたCPUである。アーキテクチャの大幅な見直しにより、VE CPUは、前機種であるSX-ACEのCPUに対して約10倍の演算性能向上及び約5倍のメモリ帯域拡大を実現した。

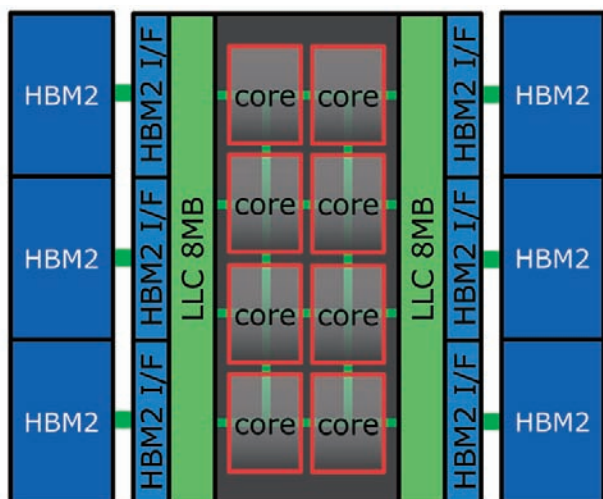
図3に示すように、CPU LSI内には8個のコアが含まれている。各コアはベクトル型アーキテクチャを採用し、単一命令により最大256個のデータを処理するベクトル命令をサポートしている。さらに、データを効率的に並列処理可能なハードウェア機構を備えることにより、コア当たりの演算性能307GFlops (倍精度浮動小数点時) 及びコア当たりの最大メモリ帯域400GB/sの世界トップクラスの超高実効性能コアを実現している。

なお、本性能は、プログラムを後述するベクトルコンパイラでコンパイルすることにより、ユーザがCPUアーキテクチャの違いを強く意識することなく利用することができる。

コアとHBM2メモリアンタフェースの間には合計16MBのキャッシュ (Last Level Cache、LLC) が備わり、コア/キャッシュ間は二次元メッシュ形状のメモリネットワーク



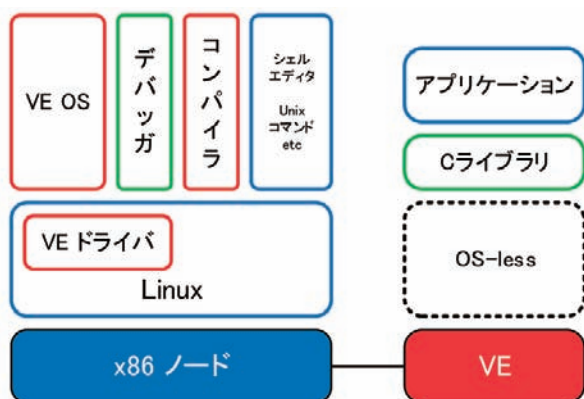
で接続される。HBM2による広いメモリ帯域を活かせるよう、これらも十分な帯域を確保した設計が行われている。



■図3. Vector Engine CPUの構成

4. SX-Aurora TSUBASAのソフトウェア

本章では、SX-Aurora TSUBASAのソフトウェアについて述べる（図4）。



■図4. SX-Aurora TSUBASA のソフトウェアスタック

SX-Aurora TSUBASAの最大の特徴は、アクセラレータ型のハードウェアでありながらアクセラレータ用の特殊なプログラミングなしに使用できる点である。本特徴は、x86ノード上のLinuxに実装されるソフトウェアスタックにより実現されている。

VE上にはOSが存在せず、アプリケーションのプロセスのみが動作する。x86ノード上ではVEにOS機能を提供するVE OS、VEが実行するプログラムをデバッグするデバッガ、VE用のコンパイラ、シェル、エディタ、その他の通常のLinuxノードで利用できるツール類が動作する。

4.1 コンパイラ

VE用コンパイラとして、C/C++及びFortranコンパイラが提供されている。VE用コンパイラの特徴を以下に列挙する。

- 自動ベクトル化：コンパイラがプログラム中でベクトル化可能な部分を自動的に検出し、ベクトル化したコードを生成する
- 自動並列化：コンパイラがプログラム中の並列性を自動的に検出し、並列化したコードを生成する
- 標準準拠：コンパイラはISO C11/C++14、Fortran 2003規格をサポートしている

前二者はいずれも並列性を利用した最適化であり、自動ベクトル化は単純ループなどの狭い範囲で特に有効であるのに対し、自動並列化はサブルーチンコールを含むような広い範囲の並列性に対しても有効であるという特性がある。

以下に自動ベクトル化の一例を示す。図5に示すCソースコードでは、forループ内で配列の各要素に同一の演算を行っている。コンパイラはこの演算がベクトル化可能であることを自動的に検知し、ベクトル命令（vld.nc、vfadd.a）を用いたコード（図6）を生成する。

これらにより、ユーザは、新規に特別な言語を使用することなく、SX-Aurora TSUBASAを使用するプログラムを開発することができる。また、既存のアプリケーションを

```
int function() {
    double a[1000], b[1000], c[1000];
    int i;
    ...
    for (i = 0; i < 1000; ++i) {
        a[i] = b[i] + c[i];
    }
    ...
}
```

■図5. Cソースコード

```
lvi      %s58
vld.nc   %v63,8,%s59
lea      %s57,-16000
and      %s56,%s57,(32)0
lea.s1   %s55,-1(,%s56)
adds.l   %s54,%s60,%s55
vld.nc   %v62,8,%s54
vfadd.d  %v61,%v63,%v62
lea      %s53,-8000
and      %s52,%s53,(32)0
```

■図6. 命令コード



修正することなく、VE用コンパイラでコンパイルするだけでSX-Aurora TSUBASAで動作させることができる。

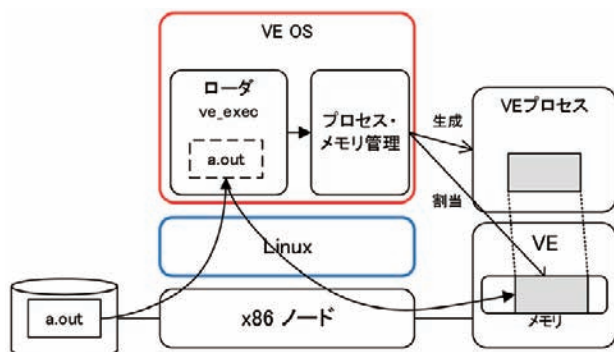
4.2 VE OS

SX-Aurora TSUBASAは、Linuxホスト上にVE OSを動作させることにより、入出力などのシステムコールを含む既存のプログラムをそのまま実行可能としている。本節では、VEプログラムの動作と、それを支えるVE OSの機能を説明する。

4.2.1 VEプログラムの開始

ユーザはVE上で実行するプログラムを指定して、シェルからローダ`ve_exec`を起動する。例えば`a.out`を起動するには、「`ve_exec ./a.out`」を実行する。

VEプログラムの動作を図7に示す。



■図7. VEプログラムの動作

ローダ`ve_exec`は実行ファイルを読み込み、プロセス・メモリ管理機能にVEプロセスの作成を指示する。プロセス・メモリ管理機能はVEプロセスを生成し、当該VEプロセス用にVEのメモリを割り当てる。プロセスの生成が完了すると、ローダはVEプロセスのメモリにプログラムを転送し、VEプロセスの実行開始をプロセス・メモリ管理機能に要求する。プロセス・メモリ管理機能はVEに実行開始指示を送ることで、VEにおいてプログラムの実行が開始される。

4.2.2 システムコール

本項では、VEにおけるシステムコールの実装について説明する。VE OSでシステムコールを処理するハンドラには、以下の2種類の実装がある。

- 単純オフロード: Linuxが提供する同一のシステムコールを呼び出し、その結果をVEプロセスに返す。入出力・

ファイルシステム・ソケットの操作などは単純オフロードハンドラで実装される。

- 各システムコール固有の実装: VE OSが自身の管理するデータを用いて処理する。メモリ管理（`mmap`、`munmap`、`brk`など）やプロセス・スレッド関連（`fork`、`clone`、など）はVE固有の資源を操作するため、固有の実装である。

VEプロセスがシステムコール命令を実行するとVEプロセスが停止され、VEからVE OSにシステムコール割込が通知される。VE OSは割込を受信すると、システムコール番号に基づき対応するハンドラを呼び出す。

大多数のシステムコールは単純オフロードで実装されるので、VE OSはLinuxとの高い互換性を低コストに実現できる。

4.2.3 VE用ライブラリ

VE上で通常のプログラムを実行可能とするため、通常のLinuxと同等のCライブラリ（`libc`）が提供される。VE用`libc`はISO C99、C11、及びPOSIXに準拠したLinux向けのオープンソースの`libc`を移植したものであり、通常のLinuxと同等の機能を提供する。これにより、標準CやPOSIXの関数を使っているアプリケーションのコードを修正することなくVE上で実行できる。

5. おわりに

新型スーパーコンピュータSX-Aurora TSUBASAは、従来のHPC領域だけではなく、ビッグデータ領域の高速化を目指し開発された。幅広い分野での利用を可能とするために、従来の大規模計算機センタのシステムだけではなく、デスクサイドのタワー型ワークステーションまで単一のアーキテクチャで提供することを目指した。

SX-Aurora TSUBASAは、NECが30年以上開発・提供してきた実績あるベクトルスーパーコンピュータの技術と、近年のサーバのデファクトスタンダードであるx86・Linuxを融合することにより、1) 高メモリバンド幅による高い処理性能、2) 自動ベクトル化・自動並列化を行うC/C++、Fortranコンパイラによる使い易さ、3) 多くのユーザが慣れ親しんでいるx86/Linuxによる標準的な環境を実現した。

今後NECはSX-Aurora TSUBASAを従来からのHPC領域のみならず、AI、ビッグデータ解析、資源探査、画像解析、セキュリティなどの新しい領域のプラットフォームとして提供することにより、安全・安心・高効率・公平で豊かな社会の実現に貢献していく。