

「京」の後の時代を支えるスパコン



# ⑥ いまどきのベクトル型スパコン —高性能ベクトルコアと高バンド幅メモリ で高い実行効率を追求—

小林広明 | 東北大学

## ベクトル型スパコンとは？

皆さんはベクトル型スパコンと聞いて何を思い浮かべるのだろうか？ Crayをはじめとする古典的スパコンだろうか？ 確かに1命令でたくさんのデータを処理するベクトルISA（命令セット）に大きな違いを見出せないかもしれないが、それを支えるマイクロアーキテクチャ（プロセッサ構成方

式）は、その時代の最新の半導体技術とアーキテクチャ技術が取り入れられ、大きく進歩してきた。2009年9月3日に開催のFIT（情報処理学会情報科学技術フォーラム）において、「21世紀はベクトルコンピューティングの時代!？」というタイトルで企画セッションを行ったが、いまやGPU, Xeon Phiそしてx86プロセッサでもベクトル命令や処理機構が当たり前となり、プログラムのベクトル化は、

プロセッサのマルチコア化に伴うプログラムの並列化とともに、プロセッサの性能を引き出す上で重要なプログラム最適化技術となっている。

東北大学が運用するスパコンSX-ACEやSXシリーズの最新のモデルであるSX-Aurora TSUBASAなどいわゆるベクトル型スパコンの特徴は、複数のデータを単一の命令で処理できるベクトル演算命令やベクトルロード・ストア命令等を有することに加え、そのマイクロアーキテクチャやシステムアーキテクチャに高性能ベクトルコアや高バンド幅メモリサブシステムを導入していることにある。

図-1はSXを含む代表的なプロセッサのコアあたりのFlop/s性能とメモリバンド幅を、図-2はプロセッサソケットあたりの同性能を示したものが、ベクトル型スパコンのコアあたりの演算性

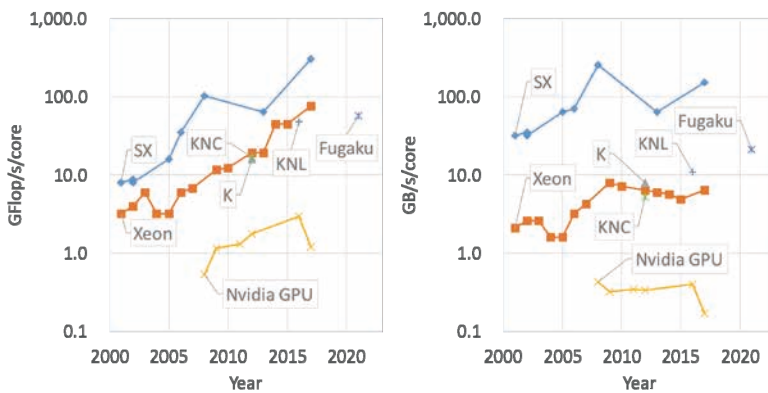


図-1 コアあたりの演算・メモリ性能

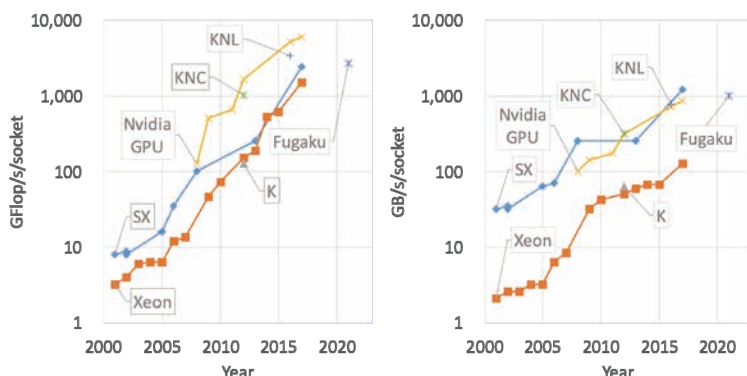


図-2 プロセッサソケットの演算・メモリ性能

能およびメモリ性能は群を抜いており、たとえば SX-Aurora TSUBASA と Xeon Skylake を比べるとコアに関しては、Gflop/s 性能で約 4 倍、メモリバンド幅性能で 24 倍、ソケットに関しては Gflop/s 性能は 1.6 倍程度であるが、メモリバンド幅性能は 10 倍近くの性能差があり、SX のベクトルプロセッサは同時代の HPC 向けプロセッサに比べ高いコア性能とメモリバンド幅性能を有していることが分かる。

これらのことから、ベクトル型スパコンは高い B/F (メモリバンド幅性能 bytes/s と演算性能 flop/s の比) を持つシステムともいえる。高いメモリ性能や B/F を有するシステムは、後述するように流体や構造計算などメモリ制約である多くの実アプリケーションの実行において、高い実効性能が得られることが知られている<sup>1)</sup>。

本稿では、東北大学サイバーサイエンスセンターで 2015 年から運用しているベクトル型スパコン SX-ACE と 2018 年に発売された最新の SX シリーズのモデル SX-Aurora TSUBASA のシステム構成について述べ、これらのシステム向けプログラム

開発とその高速化技術に関する研究から得られた知見を紹介し、ベクトル型スパコンの現状と今後への期待について述べる。

## 東北大学大規模科学計算システム SX-ACE

### システム構成

東北大学で運用する SX-ACE システムは 5 クラスタから構成される。各クラスタは 512 ノードが IXS (Internode Crossbar Switch) という専用の多段クロスバーネットワークで接続され、うち 2 クラスタはクラスタ間も IXS 接続されている。1 つのノードは 256Gflop/s の性能を有するベクトルプロセッサ 1 つと 64GB のメモリサブシステムが 256GB/s のメモリバンド幅性能で接続されている。したがって、1 クラスタのピーク性能は 131Tflop/s、総メモリバンド幅は 131TB/s で、B/F は 1.0 となる。

図-3 に SX-ACE のプロセッサアーキテクチャを示す。SX-ACE プロセッサは 4 つのベクトルコアから構成され、これらのコアはクロスバース

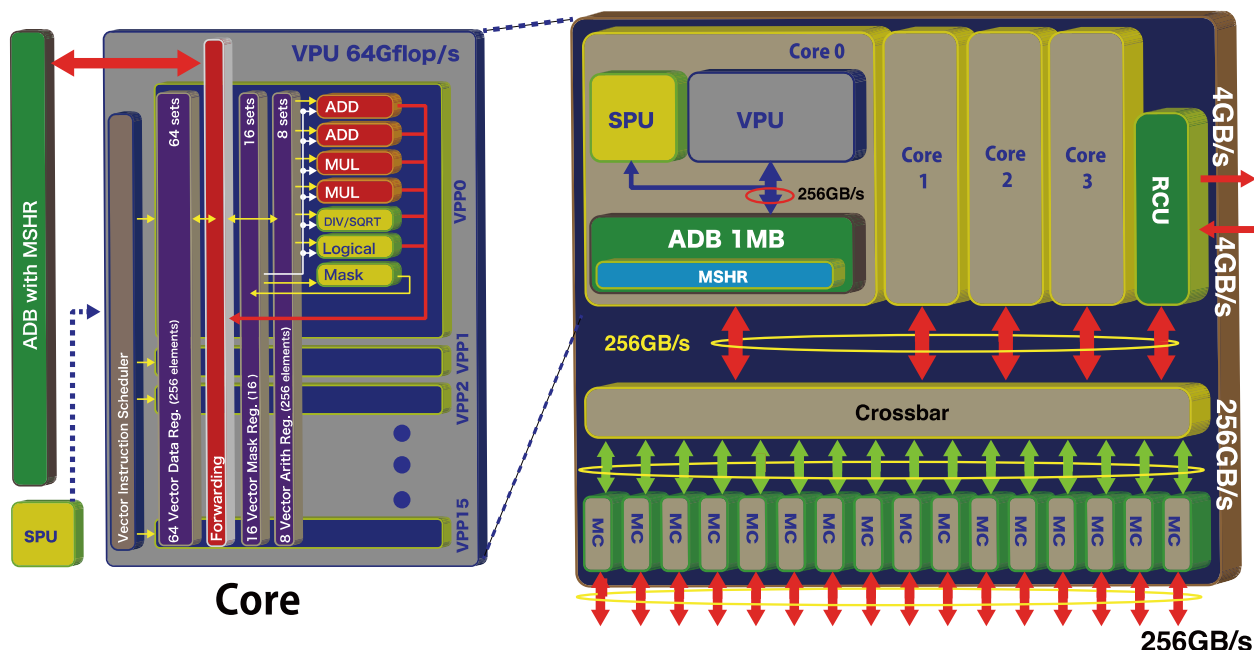


図-3 SX-ACE プロセッサアーキテクチャ



イッチを介してメモリサブシステムへ接続されている。各コアは、主にスカラプロセッシングユニット (SPU)、ベクトルプロセッシングユニット (VPU)、不要なメモリアクセスを削減する MSHR (Miss Status Handling Register) 機能付きのオンチップベクトルデータキャッシュ (ADB, Assignable Data Buffer) から構成される。VPU は 256 ベクトル長を持つベクトル命令を処理する。コア占有型の ADB は 1 MB の容量を有し、再利用性の高いベクトルデータを選択的に保持でき、データアクセス時の ADB ヒット時には広帯域・短レイテンシでのデータアクセスを実現すると同時に、オフチップメモリへのアクセス削減に貢献する。4つのコアはクロスバースイッチを介してメモリに接続され、アプリケーションの特性に応じて選択的にコアを活性化する機能により 1 コアあたり最大で 256GB/s のメモリ帯域を占有利用することが可能であり、この機能によりプログラムの特性を考慮して B/F を 1 から 4 に調整することもできる。SPU は 1Gflop/s の性能を有し、主に命令デコード、VPU に対するベクトル命令発行、およびスカラ系命令の処理を行う。加えて、ベクトル処理能力の向上のために、ベクトルロード・ストア命令の実行順序を変えて高速化するアウトオブオーダー実行やベクトルパイプ間でのデータフォワーディング機構なども兼ね備えている。

## SX-ACE の応用分野と性能

SX-ACE の高いメモリバンド幅や高性能コアによる効率の良いベクトル処理能力は、多くの科学技術分野や工学分野での研究開発に活用されている。具体的には、流体科学、電磁解析、熱解析、気象気候、災害科学、材料物性、宇宙などその分野は広範囲におよび、これらの成果は国産近距離ジェット機、発電タービン、高性能タイヤ設計など産業界にもインパクトを与えている。

また、東北大学と大阪大学の SX-ACE が中心となって実現されたリアルタイム津波浸水被害推計システムは、南海トラフ沖など鹿児島から静岡までの約 6,000km にもおよぶ太平洋沿岸域での大規模地震発生時には、ただちに緊急シミュレーションモードに両機関の SX-ACE が設定され、地震発生から 30 分以内に 30m メッシュの精度で 6 時間分の津波浸水被害情報を政府や自治体へ提供する。大規模津波被害発生時に救助や復旧などを迅速化することを目的として開発された本システムは、現在、内閣府の防災システムの一部に組み込まれ、24 時間・365 日運用されている。また津波浸水被害シミュレーションにおいては、同じ規模のシミュレーションを行う京コンピュータと比較して、SX-ACE は 1/32 以下のコア数で要求時間を達成できることが明らかになっており、SX-ACE の高い実効性能がここからも分かる<sup>2)</sup>。

次に、SX-ACE の有する高いメモリバンド幅や B/F がもたらす、高い実効性能の例を世界的な性能指標である HPCG (High Performance Conjugate Gradient) ベンチマークを例に説明する<sup>1)</sup>。

我々は SX-ACE の性能を最大限に引き出すために、超平面法の導入とベクトル長と ADB ヒット率向上の両面を考慮した問題サイズ的最適化、再利用性の高いデータの選択的 ADB 配置など多く高速化技術を HPCG ベンチマーク向けに開発した。その結果、図-4 に示すように実行効率 (ピーク性能に対する達成性能の

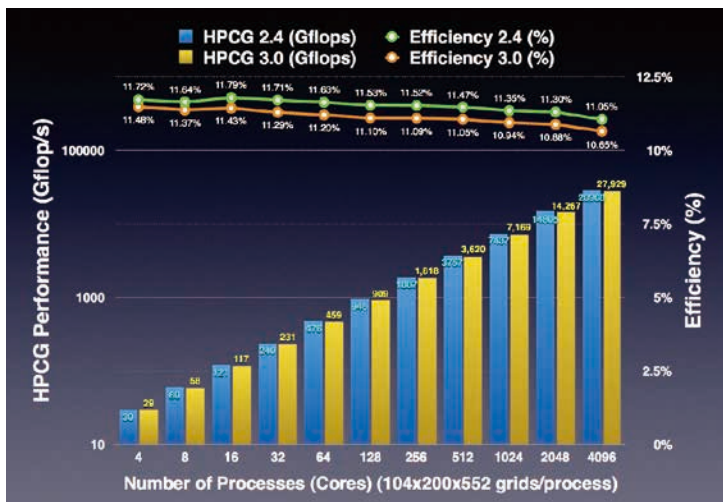


図-4 SX-ACE の HPCG 性能

比) 11% 前後の非常に安定した台数効果を得ることができた。

図-5 に HPCG ベンチマークに登録されているシステムの性能比較 (2016 年 11 月 SC16 登録時点) を示す。同図から分かるようにスカラプロセッサや GPU などアクセラレータを搭載するスパコンが 0.3 ~ 5% 程度の実効性能であるのに対し、SX-ACE は約 11% と優れた実行効率を達成している。

実行効率の低さをカバーするためには、同図の Required Peak 倍のピーク性能を少なくとも有する必要がある、システムの示すピーク性能の意味するところがメモリ性能によって大きく変わってくることを示唆している。

以上の SX-ACE でのプログラム開発の経験から、メモリ制約のベンチマークやアプリケーションで高い実行効率を達成するには、演算性能とメモリ性能のバランスが重要であると言える。

## 最新のベクトル型スパコン SX-Aurora TSUBASA

### SX-Aurora TSUBASA の構成

#### SX-Aurora TSUBASA のシステム概要

SX スパコンシリーズは、前述のように高性能

ベクトルコアと高メモリバンド幅メモリの組合せで、多くのメモリ制約のアプリケーションにおいて高い実効性能を実現してきた。その一方、SX のオペレーティングシステムは System-V をベースとした Super-UX が導入されており、近年の Linux ベースのオープンソースの実装に困難が生じて、利用拡大の制約となっていた。そのような問題を解決するために、SX シリーズの持つ高いベクトル実効性能とデファクトスタンダードな Linux のプログラミング環境の汎用性を得るために、新たに開発されたのが SX-Aurora TSUBASA である。

SX-Aurora TSUBASA はベクトルプロセッサを搭載する Vector Engine (VE) と、Linux OS 機能を実行する x86 Vector Host (VH) から構成される新しいシステムアーキテクチャを採用している。この構成は従来のアクセラレータのシステムアーキテクチャと似ているが、その設計コンセプトは大きく異なる。SX-Aurora TSUBASA では、アプリケーションは VE で実行され、Linux のシステムコールなどの OS 機能のみが VH にオフロードされる。このようなシステムアーキテクチャと実行モデルを採用することで、従来のアクセラレータに比べ 2 つの大きな利点がある。1 つは、アプリケーションが VE で実行されるため、VE と VH のデータ転送を最

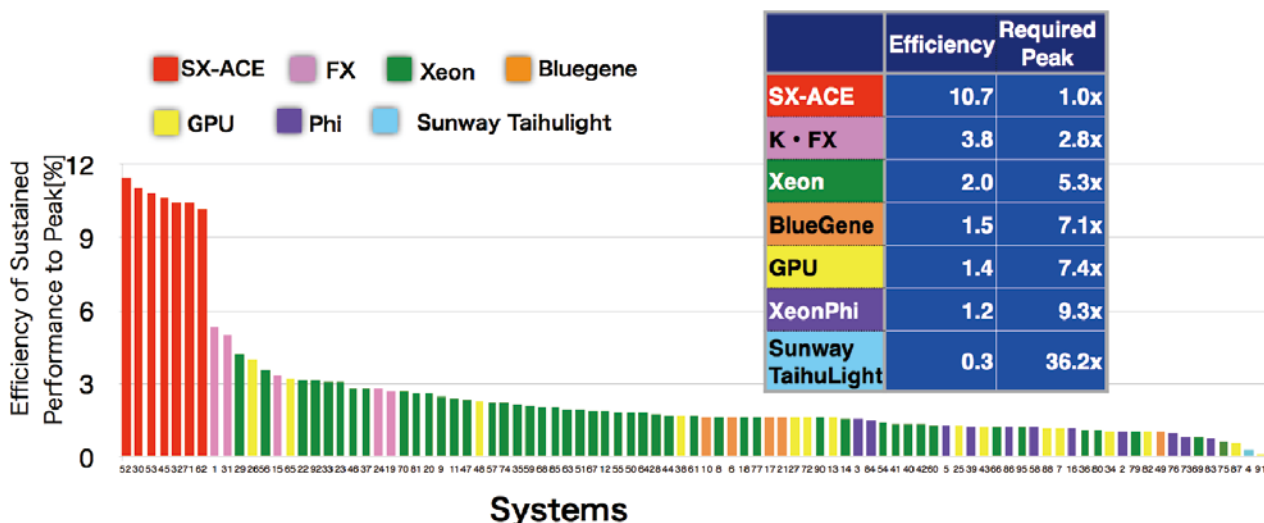


図-5 HPCG ベンチマークの実行効率



小限におさえることができることである。2つ目は、VE でプログラムを実行するため、従来の SX と同様、コンパイラが自動ベクトル化、自動並列化を行うことで特別なプログラミングをしなくてもベクトルプロセッサでプログラムを実行することができる点である。SX-Aurora TSUBASA は、従来のアクセラレータの実行モデルの問題点を解決し、これまでの SX スパコンと同様の使い勝手の良さを維持しつつ、Linux 標準環境を利用者に提供している。

### ベクトルエンジン (VE)

図-6 左図に VE の CPU パッケージを示す。長方形の形をしたベクトルプロセッサが中央に配置されて、6つの HBM2 (High Bandwidth Memory 2) メモリがベクトルプロセッサの両サイドに3つずつ配置されている。6つの HBM2 実装は世界で初めての実装になる。この最先端技術によって、ベクトルプロセッサは、1.22 TB/s もの世界最高のメモリバンド幅と容量 48GB のメモリサブシステムを実現している。また、演算性能は、単精度浮動小数点演算で 4.90 Tflop/s、倍精度浮動小数点演算で 2.45 Tflop/s とメモリ性能とバランスのとれた設計になっている。

図-6 右図に示すように、ベクトルプロセッサは、8つのベクトルコア、16 MB のラストレベルキャッシュ (LLC)、2次元メッシュメモリネットワーク、6つの HBM2 メモリインタフェースから主に構成

されている。たとえば Type 10A の VE では、各コアが 1.6 GHz の周波数で動作し、単精度浮動小数点演算 (SP) で 614.4 Gflop/s、倍精度浮動小数点演算 (DP) で 307.2 Gflop/s の演算性能を達成している。16 MB の LLC は各コアと 2次元メッシュで接続され、キャッシュバンド幅は 3.0 TB/s を実現している。各 HBM2 メモリインタフェースは 204.8 GB/s のバンド幅で HBM2 と接続されている。ベクトルプロセッサは PCI Express Gen3 のインタフェースを介して、VH および DMA (Direct Memory Access) エンジンと接続されており、DMA エンジンにより VE 間および VH-VE 間のデータ転送を行う。ベクトルプロセッサは 16 nm FinFET プロセスで製造されており、14.96 mm × 33.00 mm の面積に約 48 億トランジスタが集積されている。GPU とほぼ同じサイズにパッケージされた VE と HBM2 メモリの消費電力は、VE カード単体で 300 W 以下である。

### SX-Aurora TSUBASA の実行モデル

SX-Aurora TSUBASA では、基本的に VE でアプリケーションが実行され、OS に関するシステムコールや VH での実行が適しているスカラ演算などを VH 側にオフロードする (図-7 左図 VE execution モードと呼ぶ)。図-7 左図では、VE で実行されるベクトル処理中心のアプリ (V AP) から Linux システムコールやスカラ演算部分 (S) が

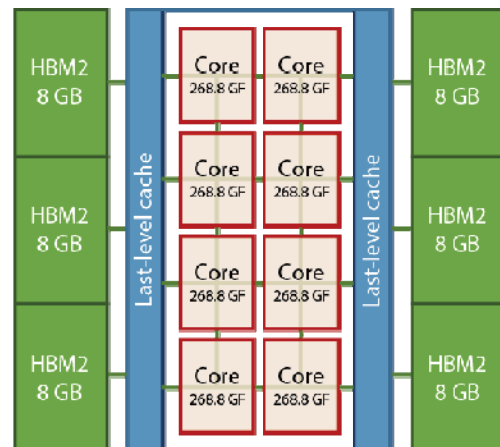


図-6 VE の構成：ベクトルチップ写真 (左) とシステムアーキテクチャ (右)

VHへオフロードされる様子を示している。

VHでは、プロセスのスケジューリングや、アプリケーションから呼ばれるシステムコールの処理など、OSの機能を担当する。さらに、システムコールなどの処理は自動的にVHへオフロードされるため、VHとVE間のやりとりを明示的に指示する必要などはない。またGNU C Library (glibc) はVE向けに移植されているため、glibcプログラミングが可能である。言い換えると、アクセラレータと異なり、SX-Aurora TSUBASA上でアプリケーションを実行するために特別なプログラミングは不要であり、C/C++やFortranなどの標準的なプログラミング言語で書かれたプログラムをそのままコンパイルすることで、SX-Aurora TSUBASA上で実行することができる。これによって、従来のベクトル型スパコンと同様の使い勝手の良さを実現している。

また、VHが適した計算を明示的にオフロードするVH offload実行モードも有する。これにより、スカラ処理などVHで処理する方が効率的である部分をユーザが指定しVHで強制的に処理させることができる。

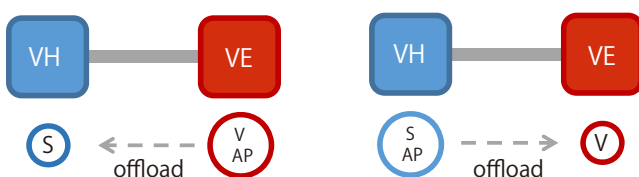


図-7 SX-Aurora TSUBASA の実行モデル

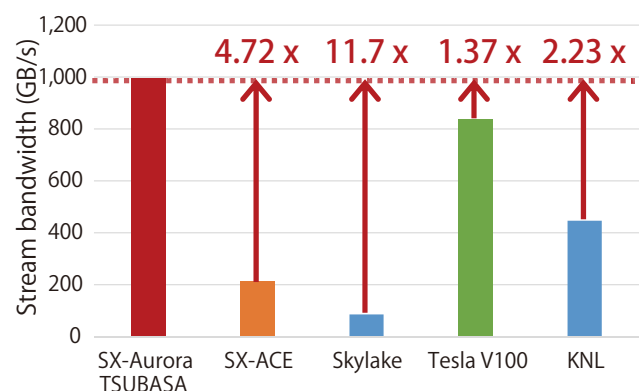


図-8 STREAM ベンチマークによる性能評価

一方、従来のGPUのようにVH側でプログラムを実行し、ベクトル処理部分をVE側にオフロードする機能も提供し、GPU向けに開発されたプログラムの移植性も考慮されている(図-7右図VE offloadモードと呼ぶ)。図-7右図では、VHで実行されるスカラプロセッサ向けに開発されたアプリ(S AP)からベクトル演算部分(V)がVEにオフロードされている様子が見られる。

### SX-Aurora TSUBASA の性能

ここでは、スパコンの性能評価指標として国際的に使われるSTREAMベンチマークとHimenoベンチマークの2つを用いたSX-Aurora TSUBASAの性能について紹介する。併せて、新たに導入された実行モデルについても評価結果を報告する。評価の詳細については文献3)を参照されたい。

#### STREAM ベンチマーク性能

図-8に、スパコンの実効メモリバンド幅を測定するための標準ベンチマークプログラムであるSTREAMカーネルを用いたSX-Aurora TSUBASAの実効メモリ性能を示す。比較のために、SX-ACE、Intel Skylake、NVIDIA Tesla V100、Intel KNLの性能も測定した。同図から分かるようにSX-Aurora TSUBASAでは、SX-ACE、Skylake、Tesla V100、KNLに比べて、それぞれ4.72倍、11.7倍、1.37倍、2.23倍と高い実効メモリバンド幅を示している。

#### Himeno ベンチマーク性能

Himenoベンチマークは非圧縮性流体解析コードの性能評価のために考案されたもので、ポアソン方程式解法をヤコビ反復法で解く場合に主要なループの処理速度を計測するものである。19ポイントステンシル計算が高コストであるために、主にコンピュータのメモリバンド幅の性能に大きく左右されるという特徴がある。

図-9に、Himenoベンチマークによる単精度演算での性能評価結果(flop/sによる実効性能値)を示す。横軸はシステム名を示す。ここでは、3つの



SX-Aurora TSUBASA のシステム (10B, 10A, 20B, それぞれ1VEで測定) の性能に加え, 比較のために SX-ACE, FX100, 富岳, Tesla V100 (以上, それぞれ1プロセッサについて測定. なお, 富岳の性能はISC 2019の発表資料による), Xeon Skylake (2プロセッサで測定) の性能も併せて示した. 同図下段には, システムのスペック Peak Perf. (DP) (Tflop/s) で示す倍精度演算性能 Tflop/s 値と, Memory BW (TB/s) で示すメモリバンド幅性能 TeraBytes/s) を示している. データは問題サイズの最も大きい XL (1,024 × 512 × 512) を用いた.

Himeno ベンチマークはメモリ負荷が高く, グラフから分かるように高いメモリバンド幅が Himeno ベンチマークにおいても高い実行効率として表れている (なお, さらなる最適化でメモリアクセス命令を削減したことにより, 10B の性能は文献3) より性能が約10%向上している. また, 10A, 20B は東北大学サイバーサイエンスセンター・高性能計算技術開発 (NEC) 共同研究部門で先行評価した結果による).

### ベクトルオフローディング性能

図-10にSX-Aurora TSUBASA を用いてマイクロベンチマークを実行した場合の性能を示す. 本マイクロベンチマークは, ヤコビ計算 Jacobi 1とその結果を出力するI/O, その後に再度ヤコビ計算を行う Jacobi 2の3つから構成されている. 実行は, VE でプログラムを実行する VE execution, I/O を明示的に VH へオフロードする VH offload, ヤコビ計算を VE にオフロードする VE offload の3通りで行った. VE execution の場合, I/O の部分がほかの2つに比べ大きくなり, 総実行時間が長くなっていることが分かる. 一方, VH offload や VE offload の場合, スカラ処理に向いているI/OがVHで実行されるため, VE execution に比べ, 総実行時間が短くなっている. また, 初期データの転送が必要な VE offload に比べ, VH offload では初期データの転送が不要なため, 総実行時間が一番短いことが分かる. このようにアプリケーションの特性により VE-VH を適切に使い分けるハイブリッドなプログラム実行も可能になる.

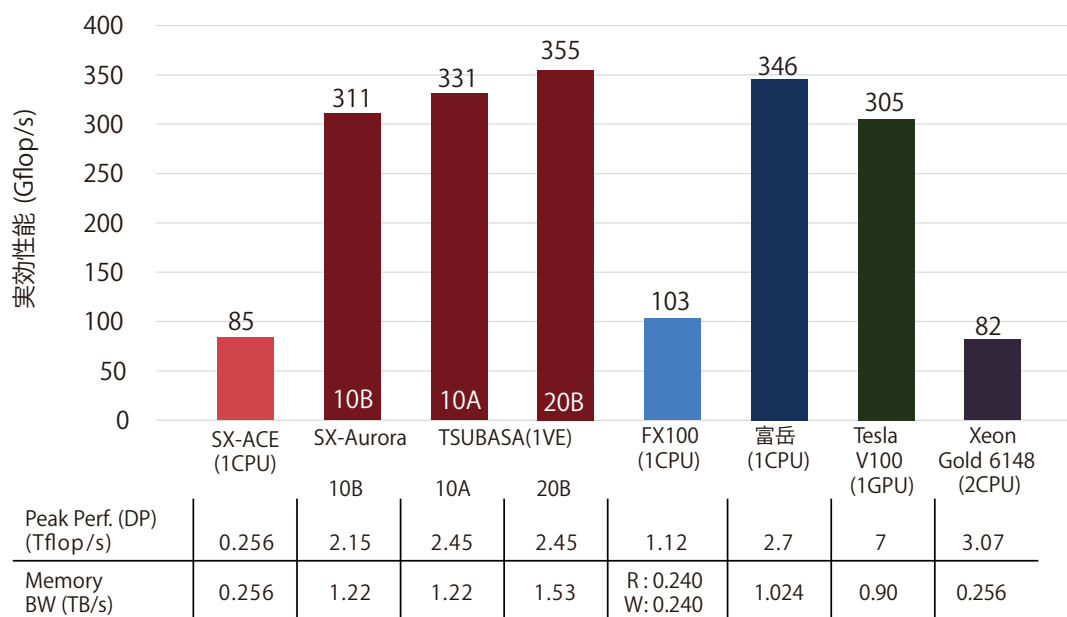


図-9 HIMENO ベンチマークによるノード性能評価

## ベクトル型スパコンユーザの広がり と今後への期待

SX-ACE を運用する代表的な機関としては、東北大学のほかに海洋研究開発機構（10 クラスタ構成、うち4クラスタはIXS相互接続で拡張クラスタ化）や大阪大学（3クラスタ構成）があるが、ここでも気候／気象シミュレーション・地震／津波シミュレーションや心臓の興奮伝播シミュレーション等で高い実行効率を達成し、我が国の最先端科学の推進や気候・気象に関する巨大なデータベースの構築に貢献している。

一方、SX-Aurora TSUBASA は、VEによる高性能ベクトルコアと高バンド幅メモリによる高いベクトル性能を維持しつつVHによるLinux環境の提供を実現する新しいベクトル型スパコンの在り方を提示している。また、これまでのSX-ACEシリーズと異なり、デスクサイドから大規模クラスタ構成まで自由に構成することもできる。SX-Aurora TSUBASA は、日本と欧米にすでに100を超えるシステムが導入され、最近のニュースとしてはドイツ気

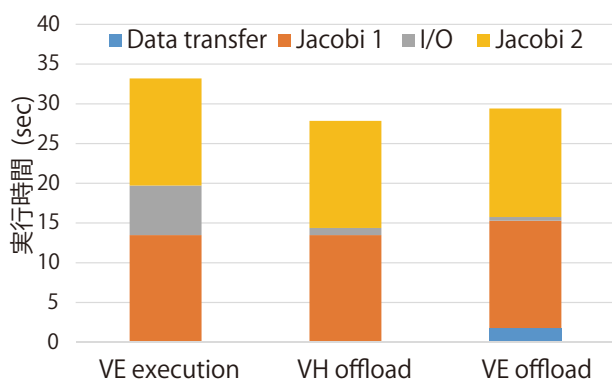


図-10 オフローディングの性能評価

象庁にSX-Aurora TSUBASA ベースの欧州最大級の気象予報システムの導入が決定している。

SX-Aurora TSUBASA 利用者コミュニティも100名規模で立ち上がり、日本と欧州のユーザが中心になって利用技術やプログラムの高速化の情報交換やベンダに対してSX-Aurora TSUBASA の利用環境の改善の提言などを行っている。

もちろん、ベクトル型スパコンは万能ではないし、応用分野に応じてさまざまなスパコンが適材適所で活用されていくものと思うが、今後、日本からの高性能ベクトル技術が世界に浸透し、新しい時代のベクトル型スパコンユーザコミュニティがますます活発になることを期待している。

### 参考文献

- 1) Egawa, R., Komatsu, K., Momose, S., Isobe, Y., Musa, A., Takizawa, H. and Kobayashi, H. : Potential of a Modern Vector Supercomputer for Practical Applications - Performance Evaluation of SX-ACE -, J. of Supercomputing, pp.1-29 (2017).
- 2) Musa, A., Watanabe, O., Matsuoka, H., Hokari, H., Inoue, T., Murashima, Y., Ohta, Y., Hino, R., Koshimura, S. and Kobayashi, H. : Real-time Tsunami Inundation Forecast System for Tsunami Prevention and Mitigation, J. of Supercomputing, pp.3093-3113 (2018).
- 3) Komatsu, K., Momose, S., Isobe, Y., Yokokawa, M., Watanabe, O., Aoyama, T., Sato, M., Musa, A. and Kobayashi, H. : Performance Evaluation of a Brand-New Vector Computer SX-Aurora TSUBASA, Proceedings of SC18 (2018).

(2019年8月1日受付)

小林広明 (正会員) koba@tohoku.ac.jp

東北大学大学院情報科学研究科教授。同サイバーサイエンスセンター・高性能計算技術開発(NEC)共同研究部門長併任。工学博士。専門は高性能コンピュータアーキテクチャとその応用。量子計算にも興味を持つ。東北大学総長特別補佐(デジタル革新担当)。