

# 各種計算科学アプリケーションにおける NEC SX-Aurora TSUBASA システムの性能評価 (1)

西川 武志<sup>†1</sup>

**概要** : FOCUS スーパーコンピュータシステム (以下、FOCUS スパコン) に 2018 年 6 月に追加した NEC SX-Aurora TSUBASA システムの各種計算科学アプリケーション (線形演算、流体計算、分子動力学法) での性能評価を行ったので報告する。Intel Xeon CPU を搭載した FOCUS スパコン A,B,C,D,E,F,G,H システムと比較してベクトル化されたソースコードでは当然ながら著しく性能が優位であるが、スカラコードであっても Xeon E5-2698v4-2.2GHz (20 コア) をノードに 2 基搭載した倍精度浮動小数点理論演算性能 1.15TFLOPS の F システムに比して平均数倍の速さであった。これはノードあたりの理論メモリ転送バンド幅が NEC SX-Aurora TSUBASA システムが F システムの 8 倍であるためと思われる。線形演算、流体計算、分子動力学法、それぞれの性能傾向についても報告する。

**キーワード** : インセンティブ設計, 計算センター運用, 運用統計, 並列度向上

## 1. はじめに

計算科学振興財団 (FOCUS) が運用する産業界向けエントリースーパーコンピュータシステム「FOCUS スパコンシステム」 [1-5] は汎用 CPU に Intel Xeon CPU を採用し産業界に計算資源を提供している。アクセラレータとしては E システムに Xeon Phi 5110P を 1 ノードに 4 基, 全 48 ノード 192 基を, F システムに NVIDIA Tesla P100 を 1 基, 全 2 ノードを具備しているが 2018 年 10 月現在, Xeon Phi の利用はほとんど無く, NVIDIA Tesla P100 もわずかに利用されているに過ぎない (表 1)。

表 1 Xeon Phi 5110P 搭載 E システム (48 ノード) と Tesla P100 搭載 F システム (2 ノード) の提供および利用ノード時間とその割合 (2016 年度から 2018 年度)

年度	E (Xeon Phi 5110P)		F (Tesla P100)	
	ノード時間	割合	ノード時間	割合
2016	495	0.1%	未導入	
2017	84	0.02%	469	3%
2018	0	0.0%	128	1%

産業界の利用では利用者が直接アクセラレータに移植するというは過去 3 年度でわずか数例であり、表 1 に示した 2016 年度に Xeon Phi が利用されたのは Xeon Phi 対応電子状態計算の商用アプリケーションによるものであったが、ライセンス料が汎用 CPU 向けに加えて Xeon Phi 向けに追加でかかるためコストパフォーマンスが汎用 CPU での並列利用に劣っていたため現在では利用が全く無い。

Tesla P100 の利用も FOCUS の標準運用が 24 時間および 72 時間のバッチキューによる利用のため数日から数ヶ月もかかるような深層学習目的には適していない。Gaussian 16 Rev. B から Tesla P100 に Hartree-Fock 計算と DFT 計算において対応した [6] が 1 ノード内のみで利用可能であり

汎用 CPU1 ノードより 2 割程度コストパフォーマンスが優位なため汎用 CPU の複数ノード並列による割引や性能向上に劣っており利用が少ない。

FOCUS スパコンシステムでの利用は計算資源量のおおよそ 2/3 が商用アプリケーションやオープンソースソフトウェアによる利用が占めているが残り 1/3 は自作アプリケーションである。

Xeon Phi や Tesla P100 がそれぞれアクセラレータ向けにソースコードを改変しないと性能がなかなか出ない状況のため自作アプリケーション利用者の利用が進まない状況であるが、FOCUS スパコン V システム (NEC SX-Aurora TSUBASA Type 10B をノードあたり 1 基搭載, 全 2 ノード) は 2018 年 7 月から公開運用したが 2018 年 10 月末までに 168 ノード時間の利用があり運用時間に対する利用割合は 3%と Xeon Phi をはるかに凌駕し、2017 年度の Tesla P100 に匹敵している。

FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果をこれまで報告 [1-5] しているが利用者は現在利用しているシステムから容易には新しいシステムに移行しないことも明らかになっている。

このような状況において利用移行・向上を図る目的で NEC SX-Aurora TSUBASA システムの各種計算科学アプリケーション (線形演算、流体計算、分子動力学法) での性能評価を行った結果を報告する。

## 2. 性能評価対象 FOCUS スパコンシステム

FOCUS スパコンシステムの概要についてはこれまでの報告 [1-5] に概要を述べているが、今回、性能評価の対象とした A, D, F, H, V の各システムの基本仕様を述べる。

### 2.1 FOCUS スパコン A, D, F, H, V システム概要

FOCUS スパコン A, D, F, H, V システムの概要は次の通りである。

#### (1) A システム (224 ノード)

高並列化環境 (40Gbps QDR-Infiniband 接続)

<sup>†1</sup>(公財)計算科学振興財団  
 Foundation for Computational Science

CPU : Xeon L5640 (Westmere-EP) 2.26 GHz 6 コア×2  
108GFLOPS, RAM : 48GB, HDD : 500GB

## (2) D システム (80 ノード)

高並列化環境 (56Gbps FDR-Infiniband 接続)

CPU : Xeon E5-2670 v2 (Ivy-Bridge) 2.5 GHz 10 コア×2  
400GFLOPS, RAM : 64GB, HDD : 6000GB

## (3) F システム (60+2 ノード)

高並列化環境 (56Gbps FDR-Infiniband 接続)

CPU : Xeon E5-2698 v4 (Broadwell) 2.2 GHz 20 コア×2  
1152GFLOPS, RAM : 128GB, HDD : 6000GB

2 ノードには PCI 版 NVIDIA Tesla P100 をそれぞれ 1 基搭載

## (4) H システム (136 ノード)

高密度高並列化環境 (34 ノード/3U シャーシ, シャーシ間  
40Gbps Ethernet ×16 シャーシ内ノード間 10Gbps Ethernet  
×2 接続)

CPU : Xeon D-1541 (Broadwell) 2.1 GHz 8 コア×1  
205GFLOPS, RAM : 64GB, SSD : 512GB

## (5) V システム (2 ノード)

NEC SX-Aurora Tsubasa ベクトルエンジン環境 (56Gbps  
FDR-Infiniband 接続)

CPU : Xeon Gold 6148 (Skylake) 2.4 GHz 20 コア×1  
1024GFLOPS, RAM : 96GB, HDD : 240GB

PCI-Express 接続で NEC SX-Aurora TSUBASA Type 10B (周  
波数 1.4GHz 8 コア 2.15TFLOPS, メモリ帯域 1.22TB/s,  
HBM2 メモリ 48GB) をノードあたり 1 基搭載

V システムのホスト Xeon Gold 6148 を評価した場合は VH,  
ベクトルエンジンを評価した場合は VE と以後示す。

## 2.2 NEC SX-Aurora TSUBASA

NEC SX-Aurora TSUBASA は Xeon Phi 5000 シリーズ  
(Knights Corner) や NVIDIA の GPGPU のような  
PCI-Express を介して汎用 CPU の演算のアクセラレータと  
して動作するハードウェアと類似のハードウェア構成を取  
りながら Xeon Phi 7000 シリーズ (Knights Landing) のよ  
うにアクセラレータを通常の Linux 環境が動作しているか  
のように利用できる。

開発元の NEC ではベクトルプロセッサ+x86/Linux アー  
キテクチャと名付け[7]アプリケーション演算処理を行う  
ベクトルエンジン (VE) 部と、主に OS 処理を行うベクト  
ルホスト (VH) 部により構成し、VH 上で動作する VEOS  
というソフトウェアによって VE を VH からシームレスに  
利用できる。VH のファイルシステム上にある実行オブジ  
ェクトは、適切な環境設定によって、Intel アーキテクチャ  
のものは VH 上の Xeon CPU において、VE のオブジェクト  
のものは VE において利用者が意識すること無く実行される。

VE 上のプロセスは VEOS の働きによって利用者だけでな  
く VH 上のシステムソフトウェアからも VH プロセスと同  
等に見えるため Linux プロセス間通信等もコードやスク

リプトを特別な変更すること無く利用できる。このことは  
塩月等による SX-Aurora TSUBASA におけるプロセス間通  
信の性能評価の報告に詳しい[8]。

## 3. 性能評価と考察

性能評価においては Xeon CPU はインテル Parallel Studio  
XE 2018.0.3.222 に含まれる Fortran, Intel MPI, Intel Math  
Kernel Library(MKL)により作成した実行モジュールを用い、  
NEC SX-Aurora TSUBASA の VE は NEC SX-Aurora  
TSUBASA Fortran コンパイラ, NEC MPI, NEC Numeric  
Library Collection の Version 1.0 を用いた。

### 3.1 線形計算 (DGEMM)

ピーク性能の上限として DGEMM を使用した正方行列の  
乗算(N=3 万, プログラムが消費するメモリサイズは 20GB  
と A, VE の搭載メモリの半分の大きさに留めた) を各シス  
テムに対して行なった。各システムに対する性能測定結果  
と A システムに対する相対性能を表 3.1 に示す。

表 3.1 DGEMM (正方行列 N=3 万) の性能  
(スレッド数=ノード内コア数)

システム	スレッド数	GFLOPS (理論 ピーク)	相対 性能 A=1
A: L5640 (Westmere)	12	115 (108)	1 (1)
D: E5-2670 v2 (Ivy-Bridge)	20	427 (400)	3.7 (3.7)
F: E5-2698 v4 (Broadwell)	40	1,341 (1,152)	12 (11)
H: D-1541 (Broadwell)	8	264 (205)	2.3 (1.9)
VH: Gold 6148 (Skylake)	20	1,368 (1,024)	12 (9.5)
VE: Type10B	8	2,131 (2,150)	19 (20)

Fortran の system\_clock サブルーチンを利用した経過時間ベ  
ースで GFLOPS 値を算出したため、Xeon システムでは理論  
ピーク性能を 5%程度超過してしまっている。A システムに  
対する相対性能も F, H, VH では理論性能比以上の比となっ  
ている。

### 3.2 姫野ベンチマーク (DGEMM)

姫野ベンチマークは非圧縮性流体の解放である Poisson 方  
程式を 3 次元直行座標系による差分法により離散化し、  
Jacobi 反復法により近似解を求めるプログラムである[9]。

20年以上に渡って公開され、渡部等の報告では倍精度版が作られ評価されているが、今回は2001年11月26日作成のVersion 3.0, OpenMP 対応の単精度版を用いて評価を行なった。各システムに対する性能測定結果と A システムに対する相対性能を表 3.2 に示す。

表 3.2 姫野ベンチマーク (L: 512x256x256) の性能  
 (スレッド数=ノード内コア数)

システム	スレッド数	GFLOPS	相対性能 A=1
A: L5640 (Westmere)	12	8	1
D: E5-2670 v2 (Ivy-Bridge)	20	17	2.1
F: E5-2698 v4 (Broadwell)	40	23	2.9
H: D-1541 (Broadwell)	8	11	1.4
VH: Gold 6148 (Skylake)	20	43	5.3
VE: Type10B	8	285	36

A システムに対する相対性能は H システムをのぞいておよそノードあたりのメモリ帯域に比例している。

### 3.3 嶋ベンチマーク

嶋ベンチマークは、川崎重工業・嶋英志氏 (現 JAXA) が作成された圧縮性流体解析に特徴的な近似リーマン流速の部分を用いて Roe 法の質量流束を流用した SHUS (Simple High-resolution Upwind Scheme) [11] のコア部分を抜き出したベンチマークコードである。圧縮性数値流体力学コードの特徴を持たせることを狙いソースコードは Fortran の倍精度逐次版で、主要ループに平方根など四則演算以外の計算が入り、メモリアクセスに比べ演算数が多いのが特徴で姫野ベンチマークより実際に使われるコードに近づけたものである。本コードは並列化されていないため、各コンパイラの自動並列化機能を用いて並列化を行なった。Intel Fortran のオプションは“-O3-parallel”を、NEC Fortran のオプションは“-O4 -mparallel”を使用した。各システムに対する性能測定結果と A システムに対する相対性能を表 3.3 に示す。

姫野ベンチマークよりも A システムに対する他のシステムの相対性能の開きが小さくなった。演算密度が高い分、メモリバンド幅の影響が小さくなったためと考えられる。

表 3.3 嶋ベンチマーク (LDA=4001, IMAX=JMA=4000) の性能 (最高性能時のスレッド数, スレッド数<ノード内コア数)

システム	スレッド数	GFLOPS	相対性能 A=1
A: L5640 (Westmere)	12	26	1
D: E5-2670 v2 (Ivy-Bridge)	18	39	1.5
F: E5-2698 v4 (Broadwell)	10	46	1.8
H: D-1541 (Broadwell)	6	28	1.1
VH: Gold 6148 (Skylake)	16	81	3.1
VE: Type10B	4	220	8.5

### 3.4 md\_omp ベンチマーク

OpenMP.org でサンプルとして公開されていた md\_omp.f90 (Author: Bill Magro, Kuck and Associates, Inc. (KAI), 1998) を用いて Xeon では Intel Fortran 2018, SX-Aurora TSUBSA では nfort を用いて OpenMP による並列化の速度を測定した。当該プログラムは(カルテシアン座標次元, 粒子数)で配列が宣言されており、今回は 3 次元, 2048 粒子と非常に小規模な系での性能測定を行なった。

表 3.4 md\_omp の性能 (スレッド数=ノード内コア数)

システム	スレッド数	GFLOPS	相対性能 A=1
A: L5640 (Westmere)	12	2.2	1
D: E5-2670 v2 (Ivy-Bridge)	20	3.4	1.5
F: E5-2698 v4 (Broadwell)	40	4.4	1.8
H: D-1541 (Broadwell)	8	3.3	1.1
VH: Gold 6148 (Skylake)	20	4.4	3.1
VE: Type10B	8	2.2	1.0
VE: Type10B(vec.)	8	132	60

配列が(3, 2048)と宣言されており、ベクトル長が 50 以下と非常に短いため VE の性能が出ていない。2 次元配列を x,y,z の 3 つの独立した配列を用いたもの書き換えたものでは

132GFLOPS と 60 倍も高速化された。

### 3.5 MDCORE ベンチマーク

md\_omp はあまりにも小規模であり実際に利用される分子動力学法のコードとはかけ離れているため、筆者が分子化学研究所、産業技術総合研究所、東京工業大学でのスーパーコンピュータシステム調達でのベンチマーク用に作成した 3 次元周期境界条件系の古典分子動力学法プログラム (粒子数  $N=64$  千, 相互作用数  $N(N-1)/2+26*N*N=1.1e11$ ) を用いて性能評価を行なった。

表 3.5 MDCORE (3 次元周期境界条件分子動力学法, 粒子数  $N=64$  千, 相互作用数  $N(N-1)/2+26*N*N=1.1e11$ ) の性能 (スレッド数=ノード内コア数)

システム	スレッド数	GFLOPS	相対性能 A=1
A: L5640 (Westmere)	12	4.8	1
D: E5-2670 v2 (Ivy-Bridge)	20	22.3	4.7
F: E5-2698 v4 (Broadwell)	40	53.4	11
H: D-1541 (Broadwell)	8	4.1	0.86
VH: Gold 6148 (Skylake)	20	21	4.5
VE: Type10B	8	111	23

相互作用が十分多くかつベクトル長も 250 を超えるため VE の性能が A システムの 23 倍と著しく良好なものとなっている。

## 4. まとめ

DGEMM, 姫野ベンチマーク, 嶋ベンチマーク, MDCORE のいずれでも SX-Aurora TSUBASA の VE は良好な性能を Xeon CPU と比較して示した。このような性能を Xeon Phi や GPGPU のように, md\_omp 以外で, 特別にソースコードを書き換えること無く得られたことは VE のこれからの普及の可能性を示した。

## 5. 今後の課題

今後は, より多くの応用プログラムや MPI を用いた並列化効率の測定を様々な問題サイズに対して行う必要がある。

**謝辞** 評価に利用したのベンチマークプログラムを公開してくださっているベンチマークプログラムを提供していただいた, 理化学研究所姫野龍太郎氏, JAXA 嶋英志氏, 評価のためベクトルエンジン Type 10B を 2 基搭載した SX-Aurora A300-2 を試用させていただきいただいた NEC グローバル P F 本部関係者各位ならびに FOCUS スーパーコンピュータシステムの運用や利用者の開拓に尽力されている計算科学振興財団の同僚と利用してくださっている利用者各位に, 謹んで感謝の意を表する。

## 参考文献

- [1] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) ,2015-HPC-149(2),1-4 (2015-06-19).
- [2] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果 II, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) ,2016-HPC-157(10),1-5 (2016-12-14).
- [3] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果 III, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) , 2017-HPC-161(3),1-5 (2017-09-12).
- [4] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果 IV, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) , 2018-HPC-163(17),1-5 (2018-02-21).
- [5] FOCUS スーパーコンピュータシステムにおける並列課金インセンティブの効果 V, 西川 武志, 研究報告ハイパフォーマンスコンピューティング (HPC) ,2018-HPC-166(1),1-4 (2018-09-20).
- [6] Gaussian 16 Rev. B.01 Release Notes, <http://gaussian.com/relnotes/>
- [7] NEC SX-Aurora TSUBASA ウェブページ <https://jpn.nec.com/hpc/sxauroratsubasa/features/index.html>
- [8] SX-Aurora TSUBASA におけるプロセス間通信の性能評価, 塩月 信智, 江川 隆輔, 滝沢 寛之, 研究報告ハイパフォーマンスコンピューティング (HPC) , 2018-HPC-165 (21),1-6 (2018-07-23).
- [9] 姫野ベンチマーク, <http://acc.riken.jp/supercom/documents/himenobmt/>
- [10] 渡部善隆, 南里豪志, 藤野清次: Himeno BMT によるハイパフォーマンスコンピュータの性能評価, 情報処理学会研究報告ハイパフォーマンスコンピューティング (HPC) ,2003(83(2003-HPC-095)),137-142 (2003-08-04)
- [11] Shima, E. & Tadamas, J.: Role of CFD in aeronautical engineering (No.14) -AUSM type upwind schemes, Proc. 14th NAL Symp. Aircraft Comput. Aerodynamics (1997).