

SC18 参加報告

スーパーコンピューティング部門は、2018年11月11日から16日までの間、米国ダラスで開催されたSC18（The International Conference for High Performance Computing, Networking, Storage, and Analysis）に参加し、研究展示を行いました。本稿では、SC18に参加した東京大学情報基盤センター教職員が現地で見聞きした中から気になったことを記します。



会場の KAY BAILEY HUTCHISON CONVENTION CENTER DALLAS 前からダラスの街並みを臨む

SC18 について

SC18は、高性能計算（HPC）分野に於いて最大級の国際会議であるとともに、様々な情報技術関連企業や研究所・大学等の技術展示会としても知られています。今回の開催地は、テキサス州デンバーで、会場は KAY BAILEY HUTCHISON CONVENTION CENTER DALLAS で行われました。以前は Supercomputing-XY（XY:開催年）という名称が用いられていましたが、1997年に SC-XY という現在の名称に変更されています。今回は Supercomputing-88 から数えて30回目の開催となりました。期間中は0度を下回ることもあり、寒い日が続きました。



KAY BAILEY HUTCHISON CONVENTION CENTER DALLAS のロビー

各種ランキングについて

毎年の SC ではスーパーコンピュータの性能に関する様々なランキングが更新されます。最も有名なスパコンランキングである Top500、スパコンの電力性能を競う Green500、実際のアプリケーションに近いと言われる HPCG や、スパコンの IO 性能を競う IO-500 などがあります。本稿では Top500 から見える全体的な傾向について述べます。

表 1 Top500 の上位 10 のスーパーコンピュータ (<http://www.top500.org/> より抜粋、一部編集)

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Rmax / Rpeak	Power (kW)
1	Summit	2,397,824	143,500.0	200,794.9	0.71	9,783
2	Sierra	1,572,480	94,640.0	125,712.0	0.75	7,438
3	Sunway TaihuLight	10,649,600	93,014.6	125,435.9	0.74	15,371
4	Tianhe-2 (MilkyWay-2)	4,981,760	61,444.5	100,678.7	0.61	18,482
5	Piz Daint	387,872	21,230.0	27,154.3	0.78	2,384
6	Trinity	979,072	20,158.7	41,461.2	0.48	7,578
7	ABCI	391,680	19,880.0	32,576.6	0.61	1,649
8	SuperMUC-NG	305,856	19,476.6	26,873.9	0.72	N/A
9	Titan	560,640	17,590.0	27,112.5	0.65	8,209
10	Sequoia	1,572,864	17,173.2	20,132.7	0.85	7,890

Top500 (<http://www.top500.org/>) は世界のスーパーコンピュータの性能を LINPACK という係数行列が密の連立一次元方程式を解くベンチマークの処理速度によって競うものです。1993 年の開始以来、6 月にヨーロッパで行われる会議である ISC と、本会議 SC に

て年2回の更新を続けています。

今回のランキングにおいては、6月のランキングで初登場した、米国 DOE/SC/Oak Ridge National Laboratory の保有する Summit が引き続きトップとなりました。Top10 にランクインしたシステムの顔ぶれに大きな変化はありませんでしたが、いくつかのシステムでは LINPACK の性能が向上しています。Summit は LINPACK スコアを 122.3 PFLOPS から 143.5 PFLOPS へと伸ばしています。6月のランキング3位から2位へと順位を向上させた、DOE/NNS/Lawrence Livermore National Laboratory の Sierra も、71.6 PFLOPS から 94.6 PFLOPS へと大きくスコアを向上させています。Summit および Sierra は NVIDIA 社の最新 GPU である Tesla V100 を搭載しており、LINPACK スコアの向上は LINPACK ベンチマークプログラムやシステム全体のチューニングが進んだ結果であると考えられます。また、Rmax/Rpeak で計算される実行効率が最も低かった Trinity も性能向上により順位を上げています。ランキングの新顔として、ドイツ Leibniz Rechenzentrum の SuperMUC-NG が8位に登場しました。SuperMUC-NG は Intel Xeon Skylake 世代のプロセッサを多数搭載したシステムです。日本勢としては、6月のランキング同様、産総研の ABCI システムが7位にランクインしています。

本センターが運用するスパコンでは、Oakforest-PACS が Top500 の14位となりました。前回まで Top500 にランクインしていた Reedbush-L, Reedbush-H システムは、今回のランキングでは500位以内にランクインできませんでした。Top10 システムの変動は大きくありませんが、Top500 全体としては多数のシステムが新たにランクインしていることがうかがえます。

ゴードン・ベル賞

ゴードン・ベル賞とは、科学的に意味のある数値計算の演算性能や計算速度を競い、特に優れた結果を得た研究グループに与えられる賞です。毎年の SC において最終候補者が発表し、論文と講演をもとに SC の最終日に受賞者が決定・表彰されます。

今年には東京大学の市村強准教授らによる” A Fast Scalable Implicit Solver for Nonlinear Time-Evolution Earthquake City Problem on Low-Ordered Unstructured Finite Elements with Artificial Intelligence and Transprecision Computing”、オークリッジ国立研究所（アメリカ）の Robert M. Patton らによる” 167-PFlops Deep Learning for Electron Microscopy: From Learning Physics to Atomic Manipulation”、ローレンス・バークレー国立研究所（アメリカ）の Thorsten Kurth らによる” Exascale Deep Learning for Climate Analytics”、ユーリヒ総合研究機構（ドイツ）の Evan Berkowitz らによる” Simulating the Weak Death of the Neutron in a Femtoscale Universe with Near-Exascale Computing”、清華大学（中国）の Heng Lin らによる” ShenTu: Processing Multi-Trillion Edge Graphs on Millions of Cores in Seconds”、オークリッジ国立研究所（アメリカ）の Wayne Joubert らによる” Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction” の計6件の研究が最終候補にノミネートされました。6件のうち5件の研究がアメリカの Summit や Sierra という、NVIDIA 社の最新の GPU である Tesla V100 を搭載したシステムを用いており、Volta 世代の GPU の性能の高さが印象的でした。また、Tesla V100 を用いた研究では、単精度・倍精度浮動小数点演算だけでなく、半精度浮動小数点演算を用いた高速化が多く見られました。低精度浮動小数点数の利用には、浮動小数点演算の高速化、使用メモリ量の削減といったメリットがあり、近年注目されている性能最適化手法の一つです。

今年の受賞は2件で、Sustained Performance Prize でオークリッジ国立研究所（アメリカ）の Wayne Joubert らによる” Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction” が、Scalability and Time to Solution でローレンス・バークレー国立研究所（アメリカ）の Thorsten Kurth らによる” Exascale Deep Learning for Climate Analytics” が受賞しました。どちらの研究においても、Tesla V100 に搭載されたテンソル・コアという半精度浮動小数点演算による行列の積和算を高速化するための特別な演算器を利用していました。Tesla V100 の倍精度浮動小数点演算性能が 7 TFlop/s であるのに対し、半精度浮動小数点演算性能は 120 TFlop/s であり、Summit 全系での半精度浮動小数点演算性能は 3 EFlop/s にも及びます。特に、Joubert らによる研究は世界ではじめて1秒あたりの演算回数がエクサ回を突破したアプリケーションであることが講演中に強調されており、テンソル・コアを有効に活用できたことがゴードン・ベル賞受賞の大きな要因となったという印象でした。

東京大学情報基盤センターによる展示

東京大学情報基盤センターは昨年同様、「ITC/JCAHPC, The University of Tokyo」の名義によるブース展示を行いました。筑波大学計算科学研究センターと共同で設立した最先端共同HPC基盤施設をJCAHPCと呼び、2016年12月より共同でOakforest-PACSスーパーコンピュータの運用を行なっています。今回のブース展示においては、筑波大学計算科学研究センターによる「CCS/JCAHPC, University of Tsukuba」と協力し、両方のブースに訪れてもらったお客様にピンバッジをプレゼント行うことで、一体感のある展示となるよう工夫を行いました。また、例年同様に情報基盤センターの提供する計算資源や研究事例を紹介するポスターの展示を行い、パンフレット・チラシ・グッズの配布を行いました。また、恒例となっているブースでのプレゼンテーションでは、Oakforest-PACSなどを使った最新の研究など、以下の9件の講演が行われました。

***Accelerating nonlinear dynamic implicit finite-element earthquake simulation on Oakforest-PACS**

Kohei Fujita (Earthquake Research Institute, The University of Tokyo)

***Hardware efficient Chebyshev Filter Diagonalization at Scale**

Gerhard Wellein (Friedrich-Alexander Universität Erlangen-Nürnberg)

***Communication avoiding matrix solvers for extreme scale nuclear CFD simulations on Oakforest-PACS**

Yasuhiro Idomura (Japan Atomic Energy Agency)

***Overview of the Recent Achievements by the Cori System**

Richard Gerber (NERSC/Lawrence Berkeley National Laboratory)

***Automatic Tuning for Parallel FFTs on Intel Xeon Phi Clusters**

Daisuke Takahashi (Center for Computational Sciences, University of Tsukuba)

***Production Hardware Overprovisioning: Real-world Performance Optimization using an Extensible Power-aware Resource Management Framework**

Ryuichi Sakamoto (Information Technology Center, The University of Tokyo)

***Performance evaluation of gravitational octree code on Volta GPU**

Yohei Miki (Information Technology Center, The University of Tokyo)

***A Methodology for Batching Matrix Kernels in HPC Applications**

Osni Marques (Lawrence Berkeley National Laboratory)

***Recent Activities of ITC/U.Tokyo in JHPCN towards Exascale/Post-Moore Era**

Kengo Nakajima (Information Technology Center, The University of Tokyo)



ブース展示の様子 (ゲストによるブースプレゼンテーション)



電力性能を考慮したメイントラック論文について

A Divide and Conquer Algorithm for DAG Scheduling under Power Constraints.

近年、電力制約下でアプリケーションの性能を最適化する研究が着目されています。A Divide and Conquer Algorithm for DAG Scheduling under Power Constraints はシカゴ大学の Gökalp Demirci らによって発表された、電力制約下で DAG を有するタスクの電力性能の最適化に関する研究でした。タスク間の依存関係を守りつつ電力性能を最適化するための効率よいタスクスケジューリング手法である Divide and Conquer Algorithm を提案し、貪欲法を用いた従来の DAG タスクスケジューリングと比較し、75% の性能向上を達成していました。

従来の研究では電力制約下で既知の消費電力を持つタスクを効率よくスケジューリングする手法が提案されていますが、本研究ではさらに各タスクに対し DVFS を適用し、更なる電力効率の改善を目指したことが新しい点です。各タスクは DVFS によって消費電力と実行時間のトレードオフを持ちます。そのため、高い周波数でタスクを実行した場合、タスクの実行時間は短くなりますが、同時に実行できるタスク数は電力制約のため少なくなる可能性があります。この結果、すべてのタスクが終了するまでの時間が延びる恐れがあります。よって、DAG タスクスケジューリングにおいては各タスクの実行順序とタスクを実行する周波数を考慮しつつ、さらに電力制約を守りつつタスク実行を行う必要があります。

提案手法では電力制約下での DAG タスクスケジューリングのコストを $O(\log n)$ にする方法が提案されました。初めにタスク間の依存関係を考慮したタスクスケジューリングを行い、タスク間に依存のない複数のタスクの組み合わせを探していきます。次に依存関係の無いタスクを並列に実行することを考え、電力の最適化、すなわち最適な動作周波数の探

索を行います。この2つを繰り返すことで電力性能の最適化を低い計算コストで行います。

評価では既存の DAG を有するタスクやシンセティックに生成したタスクに対し貪欲法と提案手法である Divide and Conquer Algorithm の4つの比較を示していました。この結果、特に電力制約を厳しくした場合において、Divide and Conquer Algorithm が貪欲法と比較し、75%の電力性能の改善を達成しています。一方で電力制約を緩くした場合、Divide and Conquer Algorithm の優位性は低下し、貪欲法の方が良い結果となる傾向が示されていました。これより、厳しい電力制約下では Divide and Conquer Algorithm が優位であると述べられていました。

実際の HPC センターでは複数のユーザーが投入した多数のジョブが同時に実行されており、今後はこのような多数のジョブ間での電力融通が課題になると考えられます。また、CPU のプロセスバラつきに対する最適化も今後の課題と考えられます。

次回、SC19 は 2019 年 11 月 17 日から 22 日の日程でコロラド州デンバーにて開催される予定です。

(スーパーコンピューティング研究部門 星野 哲也)