平成 24 年 2 月 1 日

筑波研究学園都市記者会　御中

筑　波　大　学

新型スーパーコンピュータ「HA-PACS」が稼働開始

ポイント

筑波大学計算科学研究センターは、最先端の超並列演算加速器クラスタ型スーパーコンピュータ、密結合並列演算加速機構実証システム「HA-PACS」（Highly Accelerated Parallel Advanced system for Computational Sciences）を平成 24 年 2 月 1 日より稼働開始しました。HA-PACS は、宇宙・素粒子・生命などの研究をけん引する目的で、平成 23 年度より導入を進めてきました。本システムにより、今後のエクサ*1 スケールへの展開を視野に入れたアプリケーション開発と、計算科学による成果獲得を目指します。

HA-PACS は、各計算ノードに高性能の演算加速装置を搭載し、コンパクトながら極めて高い演算性能を実現します。2 基の CPU と 4 基の GPU*2 を搭載した計算ノード単体のピーク演算性能は 2.99 テラフロップス（毎秒 2 兆 9900 億回）。これは GPU を搭載した超並列クラスタ型スーパーコンピュータのノード単体として世界最高性能となります。本システムは計算ノードを 268 台結合して構成され、総ピーク演算性能は 802 テラフロップス（毎秒 802 兆回）に達します。

計算科学研究センターでは、先端計算科学推進室を中心に分野間連携および学外連携のもと、素粒子・宇宙・原子核・物質・生命・地球環境の各分野におけるブレイクスルーを目指します。また、次世代計算システム開発室において GPU 間の直接通信を可能とする「密結合並列演算加速機構」を開発・実装し、GPU 間並列処理の一層の高速化を目指します。

1. 概要

国立大学法人筑波大学【学長 山田信博】計算科学研究センター【センター長 佐藤三久】は、宇宙・素粒子・生命などの研究をけん引する最先端の超並列演算加速器クラスタ型スーパーコンピュータ、密結合並列演算加速機構実証システム「HA-PACS」（Highly Accelerated Parallel Advanced system for Computational Sciences）の導入を平成 23 年度から進めて来ましたが、平成 24 年 2 月 1 日に稼動を開始しました。

本システムは、各計算ノードに従来以上の演算加速装置を搭載し、コンパクトながら極めて高い演算性能を実現する超並列クラスタ型スーパーコンピュータで、今後のエクサスケールまでの展開を視野に入れたアプリケーション開発と計算科学による成果獲得を目指します。

HA-PACS は、米インテル社製の最新 CPU を 2 基と米エヌビディア社製の最新 GPU を 4 基搭載したコンパクトで先進的な計算ノードを 268 台結合した超並列システムです。ノード単体のピーク演算性能は 2.99 テラフロップス（毎秒 2 兆 9900 億演算）で、これは GPU を搭載した超並列クラスタ型スーパーコンピュータとして世界最高性能となります。システム全体としての総ピーク演算性能は 802 テラフロップス（毎秒 802 兆演算）に達します。

現在、高性能計算システム分野では GPU を用いたクラスタ型計算機が注目されており、日本国内でも東京工業大学の TSUBAME2.0 が国内最高性能の GPU クラスタとして稼働中です。しかし、これらの GPU クラスタでは GPU と CPU 間の通信チャネル性能に限界があり、ノード内の GPU 数の制限やこの通信チャネル部分が性能ボトルネックとなるケースがありました。HA-PACS に搭載される最先端 CPU では、従来機の 4 倍に相当する高性能な PCI Express[*3] チャネルが提供され、4 基の GPU を通信ボトルネックなしに CPU と結合しています。これにより、GPU の持つ本来の性能を最大限に活かすことが可能になりました。

計算科学研究センターでは HA-PACS を使って、様々な計算科学アプリケーションの開発と演算加速装置向けアルゴリズムの開発を進めていきます。先端計算科学推進室を中心として、素粒子・宇宙・原子核・物質・生命・地球環境の各分野におけるブレイクスルー達成のために、分野間連携および学外連携のもと、主要アプリケーションのホットスポット解析と GPU 化を進めています。これらのアプリケーション開発は、HA-PACS の大規模並列資源を長時間占有使用することで加速され、加えてセンター内の計算機科学研究者との協業により、システム特性を活かした次世代演算加速システムにつながる成果が得られるものと期待されます。

また、GPU に代表される演算加速装置を用いた並列処理において、演算加速装置間の通信には大きな問題があり、現状では CPU の助けを借りた間接的通信のみが可能です。計算科学研究センターでは「密結合並列演算加速機構」と呼ばれる新たなノード間通信機構を開発中であり、これにより従来不可能だった計算ノードをまたいだ GPU 間の直接通信を可能とします。現在この機構のハードウエア及びソフトウエアのプロトタイプ開発が進めら

れており、今回、稼働開始する HA-PACS の機能拡張として、密結合並列演算加速機構を実装する計画です。これにより、GPU 間の並列処理が一層加速され、幅広い科学技術計算の性能が加速されることが見込まれます。

2. 背景

10 ペタフロップス級のスーパーコンピュータが京速コンピュータ「京」によって実現された現在、演算性能をエクサフロップス級まで高めるための研究がすでに始まっています。しかし、1 台の計算機で使用可能な電力や設置面積の制限から、このような超高性能を実現することはますます難しくなっており、何らかの演算加速装置を持つシステムが不可欠です。これらのシステムには、演算加速装置と CPU の間の通信や、並列演算加速装置間の通信における様々なボトルネックが存在します。加えて、超並列規模の演算加速装置を用いた大規模プログラムの開発には、アルゴリズムレベルからの改良など大きな人的コストと時間がかかります。

筑波大学計算科学研究センターでは、高密度超並列 GPU クラスタを、最先端コモディティ技術と我々独自の技術の組合せにより実現し、これらの問題に挑戦します。このための研究基盤が HA-PACS です。最先端 CPU と GPU の組み合わせによる超並列 GPU クラスタを従来にない規模で定常的に並列利用することにより、エクサスケール時代につながる演算加速型アプリケーションの開発と、我々が提唱する密結合並列演算加速機構アーキテクチャに基づく次世代 GPU クラスタを実現します。ここで培われたハードウエア及びソフトウエアのシステム開発技術をエクサスケールシステム実現への基盤技術として熟成していきます。

3. 開発経緯

計算科学研究センターは、平成 23 年度から文部科学省の国立大学法人運営費交付金特別経費を受け、3 カ年計画で「エクサスケール計算技術開拓による先端学際計算科学教育研究拠点の充実」事業（責任者 センター長 佐藤三久）を推進しています。

この事業は、超並列演算加速型クラスタ計算機「HA-PACS」を開発・製作し、これを用いて宇宙・素粒子・生命の先端的な研究を推進し、さらに次世代の演算加速型並列システムの要素技術となる密結合並列演算加速機構の技術開発を行うものです。HA-PACS の基本部分となる超並列 GPU クラスタは最先端コモディティ技術に基づく CPU と GPU を搭載したシステムとして調達します。密結合並列演算加速機構については、計算科学研究センターにおいてハードウエアからアプリケーションまでの開発を行い、HA-PACS の拡張部分として実装していきます。

4. 成果

システムの特徴

HA-PACS は、268 台の計算ノードを 2 本の並列 QDR InfiniBand ネットワーク*4 で Fat Tree 結合した超並列型の GPU クラスタ計算機です。全体で 802 テラフロップス（毎秒 802 兆回）のピーク計算性能、34 テラバイトのメモリ、504 テラバイトの共有ディスクを持っています。計算科学の大規模計算を実現可能とする特徴は次のとおりです。

1）豊富な PCI Express チャネル数を持つ米インテル社の最新 CPU である E5（SandyBridge-EP）プロセッサを 2 基搭載することにより、4 基の最新型 GPU（米エヌビディア社製 Tesla M2090）をストレスなく CPU と結合させることを可能にした。これにより、GPU への通信性能を損なうことなく、2.99 テラフロップスという世界最高のノード単体性能を 2U 相当のコンパクトな構成で実現した。

2）最新 GPU 技術と CPU 技術を最大限に利用した結果、802 テラフロップスのピーク演算性能をわずか 26 台のラックにコンパクトに実装し、総電力も 428kW に抑えた。

3）2 系統の Fat Tree 構成の QDR InfiniBand ネットワークにより、2.1 テラバイト/秒のバイセクションバンド幅を持つ超高性能並列ネットワークで全ノードを結合し、ノード間に偏りのない並列通信性能と共有ファイルシステムへのアクセスを実現した。

4）InfiniBand ネットワークを介して全ノードと結合される 504 テラバイトの Lustre ファイルシステムによる共有ファイルシステムを提供し、全ノードに均質な I/O 機能と性能を提供した。

5. 用語解説

＊1　エクサ

10 の 18 乗。ペタ（10 の 15 乗）の 1000 倍。エクサフロップスとは、現在、京速コンピュータ「京」が持つ 10 ペタフロップスの性能の 100 倍、すなわち毎秒 100 京回の演算性能に相当する。

＊2　GPU

Graphics Processing Unit の略。本来 PC サーバにおけるグラフィクス処理を目的として作られた専用プロセッサだが、近年はその高い演算性能とメモリバンド幅を利用した高性能計算への転用が活発化している。

＊3　PCI Express

PC サーバにおいて CPU とネットワーク、ハードディスク、GPU などあらゆる周辺機器を接続するための標準バス。米インテル社製の最新 CPU である SandyBridge-EP では、PCI Express の最先端規格である Generation 3 を標準サポートし、さらに 1 CPU あたり 40 本もの PCI Express I/O チャネルを提供する。これにより、CPU あたり 2 基の GPU をストレスなしに接続可能となっている。

＊4　QDR InfiniBand ネットワーク
高性能クラスタ型計算機で多用される高性能ネットワーク。Ethernet などに比べて数倍～数十倍の通信性能を持ち、さらに数百～数千ノード規模のシステムを Fat Tree と呼ばれるネットワーク構成で結合可能。

6. 関連情報

筑波大学計算科学研究センター ウエブサイト
http://www.ccs.tsukuba.ac.jp/CCS
「HA-PACS」プロジェクト特設ページ
http://www.ccs.tsukuba.ac.jp/CCS/research/project/ha-pacs

＜問い合わせ先＞
筑波大学計算科学研究センター広報室
TEL：029-853-6260　E-mail：pr@ccs.tsukuba.ac.jp