［大学 ICT 推進協議会 2014 年度 年次大会論文集より転載］

新並列コンピュータシステムと活用事例の紹介

齋藤 敦子† 森谷 友映† 佐々木 大輔† 山下 毅† 小野 敏† 大泉 健治† 江川 隆輔‡ 小林 広明‡

†東北大学情報部情報基盤課

‡東北大学サイバーサイエンスセンター スーパーコンピューティング研究部

a-saito@cc.tohoku.ac.jp

概要：東北大学サイバーサイエンスセンターでは、2014 年 4 月、防災・減災分野、ものづくり分野における研究、産業利用の促進及び HPCI システムに提供する計算機資源の拡充を目的に、並列コンピュータシステムの更新を行った。新システムは、並列コンピュータシステム LX 406Re-2、ファイルサーバシステム、そして新たに導入した三次元可視化システムからなる。本稿では、新システムの構成や運用、これらの資源の活用事例を紹介する。

1. はじめに

東北大学サイバーサイエンスセンター（以下、本センター）は、全国共同利用施設として先端的大規模科学計算環境を提供するため、常に最新鋭・高性能コンピュータシステムを導入し、先端分野の研究を強力に支援している。

2014 年 4 月、防災・減災分野をはじめとするシミュレーション研究、ものづくり分野における研究、産業利用の促進及び HPCI システム（High Performance Computing Infrastructure）に提供する計算機資源の拡充を目的に、並列コンピュータシステムの更新を行った。新システムは、並列コンピュータシステム LX 406Re-2、ファイルサーバシステム、そして新たに導入した三次元可視化システムからなる。総合演算性能の向上、ストレージの増強はもとより、三次元可視化システムの導入により、本センター内で大規模科学計算からその結果の可視化までが可能となり、より幅広いサービスが提供できるようになった。

本稿では、新システムの構成、性能と運用、そして、三次元可視化システムの活用事例、高速化支援について紹介する。

2. 新並列コンピュータシステムの紹介

2.1. システム構成

本センターのシステム構成を図 1 に示す。今回更新したシステムは、並列コンピュータシステム LX 406Re-2、ファイルサーバシステム、三次元可視化システムである。なお、スーパーコンピュータシステム SX-ACE は、2015 年初頭の運用開始を予定している。

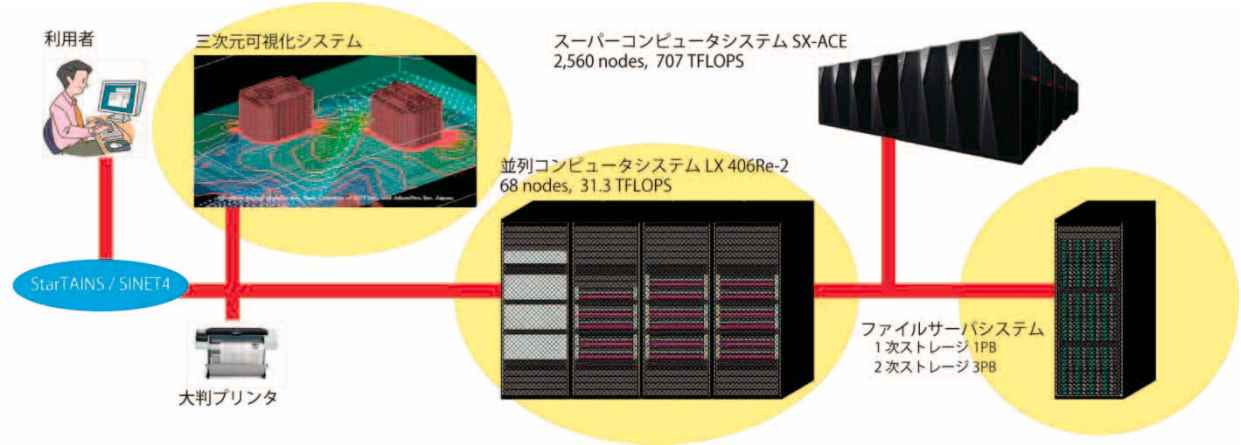

図 1　大規模科学計算システムの構成

2.2.　LX 406Re-2 の性能と運用

並列コンピュータシステム LX 406Re-2 の諸元を表 1 に示す。LX 406Re-2 は、1 ノードにインテル Xeon プロセッサ E5-2695v2（12 コア）を 2 基と 128GB の主記憶装置を搭載し、合計 68 ノードで構成される。自動並列化・OpenMP・MPI を利用したノード内の並列処理は 24 並列まで可能であり、ノードあたりの理論最大演算性能は 460.8GFLOPS となる。

LX 406Re-2 で利用可能なプログラミング言語および科学技術計算用ライブラリを表 2 に示す。コンパイラは自動並列化機能を有しているので、既存の逐次処理プログラムを修正することなく並列実行が可能である。その他、OpenMP によるノード内並列化、MPI による複数ノードを使用した並列実行、MPI と自動並列／OpenMP を組み合わせたハイブリット並列処理も可能である。

ジョブ管理には NQSⅡ（Network Queuing System）を採用している。ジョブの一元管理が可能であり、利便性の高いジョブ投入環境となっている。ジョブクラスは並列数やメモリサイズの違いにより複数用意している。並列コンピュータシステムで提供しているジョブクラスを表 3 に示す。従来同様、ジョブの大規模化・長時間化に対応するため、nh クラスを除き、すべて CPU 時間制限を無制限としている。複数のノードを使用した並列処理は、MPI の利用により最大 576 並列まで実行可能であり、ベクトル演算には向いていないプログラムも高速な実行が可能である。また、LX 406Re-2 はアプリケーションサーバとしての役割も担っており、高速ディスクアクセスが可能な SSD ドライブを搭載する専用ノードにより、Gaussian 等のアプリケーションプログラムを高速に実行することができる。

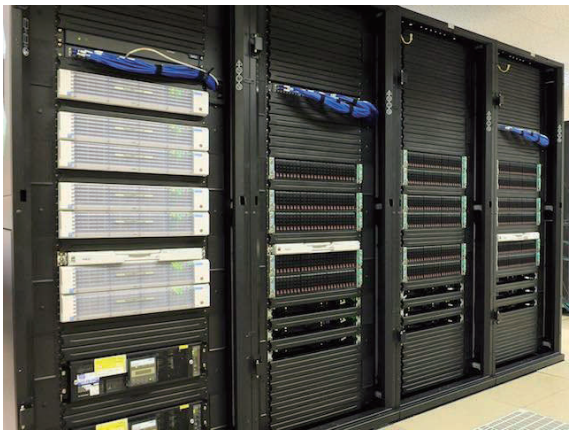

図 2　LX 406Re-2

表 1　LX 406Re-2 の諸元

システム全体	
総ノード数	68 ノード
総理論演算性能	31.3TFLOPS（倍精度）
総メモリ容量	8.5TB
ノード間接続	InfiniBand（4×FDR、56Gbps）

ノード性能	
理論演算性能	460.8GFLOPS（倍精度）
CPU	Intel Xeon プロセッサ E5-2695v2（12core／2.4GHz）×2
メモリ	128GB

表 2　プログラミング言語およびライブラリ

Fortran	Intel Fortran Composer XE
C/C++	Intel C++ Composer XE
MPI	Intel MPI ライブラリ
数値計算ライブラリ	NEC NumericFactory Intel MKL 他

表 3　ジョブクラス

ジョブクラス	利用ノード数 (コア数)	CPU 時間制限	メモリ容量 [GB]
ns	1 (1)	無制限	5
nh	1 (24)	1 時間	128
n1	1 (24)	無制限	128
n6	6 (144)	〃	128×6
n12	12 (288)	〃	128×12
n24	24 (576)	〃	128×24
mg	1 (24)	〃	128

（mg：アプリケーション専用）

2.3. ファイルサーバシステムの性能と運用

ファイルサーバシステムは、1PB の一次ストレージ領域と 3PB の二次ストレージ領域からなる。計 4PB のストレージ容量を有し、大規模なデータを扱うことができる。これらはデータ転送サーバを介して高速に相互利用可能となっている。

● 一次ストレージ

一次ストレージは主に本センターの HPCI 資源利用者に提供している。DDN 社製の lustre ファイルシステムと Gfarm ファイルシステムで構成し、1PB のディスク容量を持つ。HPCI では Gfarm ファイルシステムで構築した複数のストレージ拠点を持ち、広域に分散する大規模ストレージに対して、透過的なアクセス、簡便なファイル複製、GSI 認証による通信内容の暗号化およびデータの耐災害性の向上を図っている。

また、一次ストレージは並列コンピュータシステムと 56Gbps の Infini Band で接続しており、並列コンピュータとの I/O 性能に優れている。並列コンピュータで大規模な入出力ファイルを必要とする利用者に向けて、大規模ファイル領域としての提供もしている。

● 二次ストレージ

二次ストレージは、本センター利用者のホームディレクトリ環境として用意している。NEC 製の分散・並列ファイルシステムである NEC Scalable Technology File System（ScaTeFS）で構成され、3PB のディスク容量を持つ。ScaTeFS は、NEC 独自のプロトコルによる高効率のデータ転送方式が用いられており、多数のサーバと高速なファイルシステム共有が可能なシステムである。また、二次ストレージとスーパーコンピュータシステム SX-ACE は、最大 40Gbps の転送性能を持つ Juiper 社製 QFabric システムを介して接続する予定であり、スーパーコンピュータシステムとより高速な入出力が可能となる。

● データ転送サーバ

本センターでは、フロントエンドサーバの他にデータ転送専用のサーバも提供している。データ転送サーバは 10Gbps の Ethernet で各システムおよびネットワークに接続しており、高速なデータ転送が可能である。

2.4. 三次元可視化システムの性能と運用

三次元可視化システムは、3D 対応 50 インチ LED モニタを 12 面配置した大画面ディスプレイと、演算結果の可視化処理およびディスプレイへの描画を行う可視化サーバ 4 ノードから構成される。可視化アプリケーションは AVS/Express MPE を備えている。可視化サーバからもファイルサーバシステム上のホームディレクトリにアクセス可能であり、本センターの計算機で得られたデータを、別環境にコピーすることなく三次元可視化システムで利用可能である。また、大画面ディスプレイはテレビ会議システムとしても利用できる。

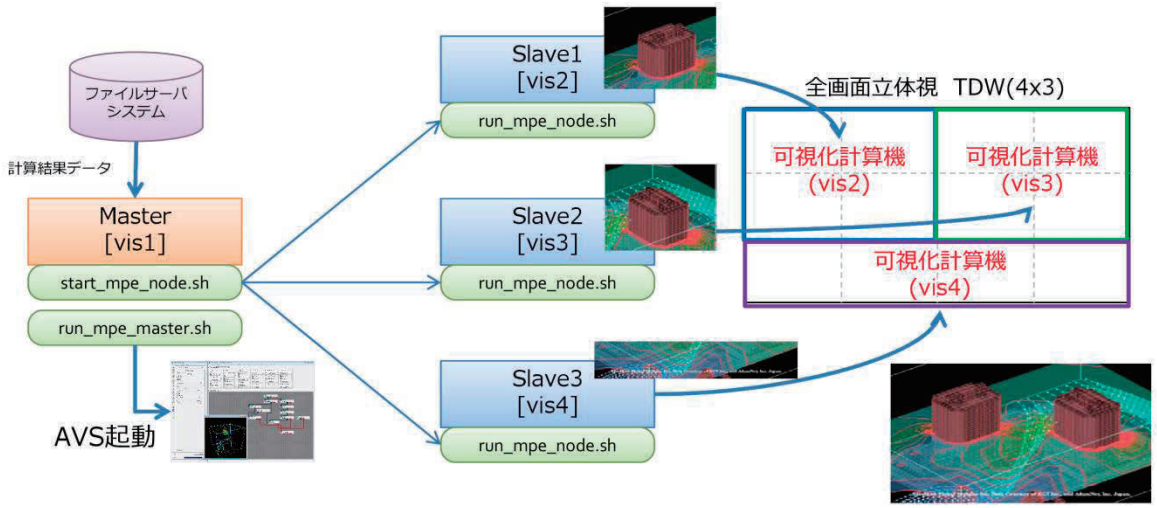

図 3　三次元可視化の仕組み

● ディスプレイ

2D／3D 表示に対応した、フル HD（1,920×1,080画素）50インチLEDモニタを12面設置し、最大 7,680×3,240 画素の高精細表示が可能である。

● 可視化サーバ

1 ノードにインテル Xeon プロセッサ E5-2670 を 2 基、メモリを 64GB、グラフィックスボード Quadro K5000 を搭載し、全 4 ノードで構成される。Master／Slave のクラスタシステム構成となっており、図 3 に示すように、3 つの SlaveNode が 12 面の大画面の映像を分担して描画する仕組みとなっている。

● 三次元可視化ソフトウェア

AVS/Express MPE を採用し、可視化コンテンツの作成および複数画面での三次元立体視表示が可能である。

● テレビ会議システム

Polycom HDX8000-1080 を採用し、フルハイビジョン（1080p）での映像接続が可能である。また、入出力インターフェースを利用してユーザの PC 画面、ビデオ映像を送信することができる。自局を含め最大 4 地点からの接続が可能である。

図 4　可視化機器室

三次元可視化システムは、本センター1F の可視化機器室に設置している（図 4）。大画面ディスプレイとほぼ同等の大きさの部屋に設置することで、より没入感のある三次元立体視が可能である。なお、三次元可視化ソフトウェア（可視化コンテンツ作成）は、可視化機器室での利用の他、リモート接続で利用することも可能である。

ディスプレイ表示パターンの例を図 5 に示す。12 画面全てを使用した全画面立体視の他、3×3 画面、2×2 画面などさまざまな表示パターンが可能であり、ユーザの多様な要求に応えることができる。

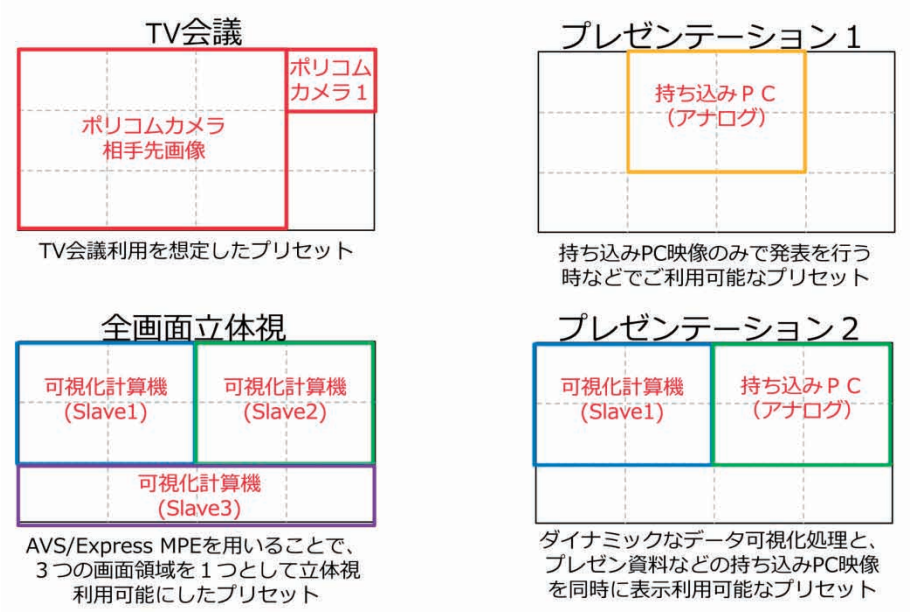

図 5　ディスプレイ表示パターンの例

2.5.　民間企業利用制度

2007 年に先端研究施設共有促進事業のもと、大学で開発された応用ソフトウェアと計算機資源であるスーパーコンピュータの民間企業への提供を開始した。この事業は、産学官の横断的な研究開発活動を推進し、大学の持つ知と施設によって我が国の経済発展に貢献することを目指している。2011 年以降は、本センターの自主事業として民間企業利用サービス制度のもと民間企業へのサービスを継続し、2007 年の事業開始からあわせてこれまでに 7 社の利用があった。

新並列コンピュータシステムもこれまで同様、民間企業での利用が可能である。一定期間の計算資源の占有利用など、柔軟なサービス提供を目指し、計算環境構築に取り組んでいる。

3.　三次元可視化システムの活用事例

本センターでの三次元可視化システムの活用事例を紹介する。

● **事例1　：シミュレーション結果の可視化**

本センターの計算機で計算された「フラーレンの爆発シミュレーション」の可視化を行った。作成した立体映像の一部を図 6 に示す。本シミュレーションは、X 線照射した際のフラーレンが爆発する様子をシミュレートしたものである。タンパク質の構造を決定する実験では、X 線の照射により構造がフェムト秒で変わるため、これを実際に観測することは難しい。数値シミュレーションにより構造変化の過程を追跡し、それを可視化することによって、実験では観測が難しい反応機構の解明が可能となる。

作成した立体映像を大画面ディスプレイに映し出し、本シミュレーションを行っている研究者に三次元立体視を体感してもらった。「奥行き情報の視覚的な認知が可能となり、二次元画像よりも時間経過による構造の変化を詳細に観測できるので、より深く理解することができる」「直感的に構造の正当性を検証することが可能になると期待される」との感想が得られ、三次元立体視の有意性を感じてもらうことができた。

また、可視化することで、本センターの来訪者にも、スーパーコンピュータ／並列コンピュータの計算結果をわかりやすい形で伝えられるようになり、本センターの広報活動にも役立っている。（図 7）。

● **事例2　：講義の遠隔配信**

組込みシステム産業振興機構主催の人材育成プログラム「組込み適塾」が関西と東北で遠隔開催され、東北会場からの中継には本センターのテレビ会議システムが用いられた。

東北での遠隔開催は数年前から行われていたが、設備上の制約から、配信される座学形式の授業を聴講するのみであり、双方のディスカッションやグループ実習ができる環境ではなかった。この解決策として、大画面ディスプレイにより臨場感あふれる双方向のビデオ通信が可能な、本センターのテレビ会議システムが利用されることとなった。入塾式にはじまり、本センターでは計 6 回の遠隔講義が開催された。講義当日は、テレビ会議システムにより、講義資料や両会場の様子が大画面で共有され、活発なディスカッションが行われた。

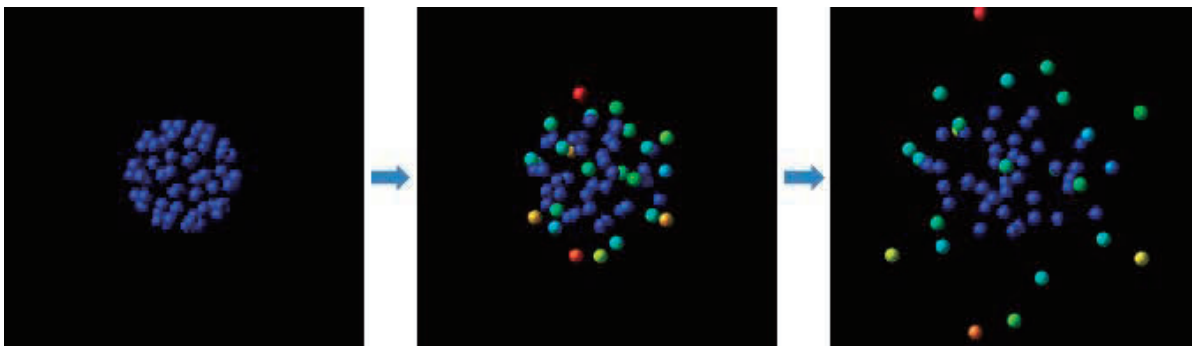

図 6　フラーレン爆発シミュレーションの可視化

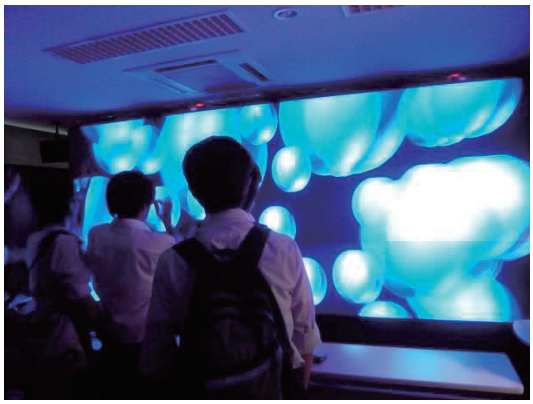

図 7　来訪者による見学の様子

4.　高速化支援

4.1.　高速化支援活動実績

本センターでは 1997 年から計算科学分野の利用者との共同研究を通じて、さまざまな分野における実アプリケーションの最適化や並列化の高速化支援を行っている。高速化支援活動の実績を表 4 に示す。センター独自の共同研究に加え、全国の情報基盤センター等と連携して JHPCN（学際大規模情報基盤共同利用・共同研究拠点）や HPCI を構成し、共同研究・高速化支援を実施している。利用者である計算科学者と本センターの計算機科学の専門家・技術職員・計算機ベンダーが密に連携し、科学・工学の恒常的な進歩を支える高速化支援活動を推進している。

4.2.　今後の高速化支援活動について

現スーパーコンピュータシステム SX-9 は 1 ノードに 16 個の高速 CPU と 1TB の大規模共有メモリを有する構成であり、コンパイラの自動並列機能により SMP 並列でこの構成を利用可能であった。次期スーパーコンピュータシステム SX-ACE は本センターの運用構成としては最大 4,096 コアと 64TB のメモリが利用可能となる予定だが、1 ノードは 4 コア、64GB のメモリで構成されており、大規模なプログラムの実行には MPI ライブラリによるプログラムの並列化が必須となる。本センターでは以前より、コンパイラの自動並列機能または OpenMP 並列のみを利用していたユーザプログラムの MPI 並列化による高速化支援も積極的に実施しており、SX-ACE の導入にあたっても、ユーザの実行環境のスムーズな移行が可能なように万全を期している。また本センターは、新規に大型計算機システムの利用を始めるユーザに対しても利用についての支援や高速化支援を行い、計算科学の研究を推進させることを継続的な目的としている。

5.　おわりに

本稿では、新並列コンピュータシステムの各システムの性能と運用および三次元可視化システムの活用事例、高速化支援活動について紹介した。今回の更新で、総合演算性能は旧システムの約 20 倍、ストレージ容量は 40 倍以上にそれぞれ増強された。また、三次元可視化システムの導入により、シミュレーション結果の高速かつ高品質な立体映像化が可能となった。ユーザ支援活動においては、従来の高速化支援に加え、可視化のフェーズまで幅広い支援が行えるようになった。今後もなお、高度化する利用者のニーズに対応できるサービスの提供を目指し、システムとサービスの強化を図っていきたい。

表 4　高速化支援活動の実績

年度	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
件数	2	9	8	9	10	7	18	20	8	29
単体性能向上比	1.9	46.7	4.5	2.5	1.6	2.2	6.7	2.9	1.5	3.1
並列性能向上比	11.1	18.4	31.7	8.6	4.9	2.8	18.6	4.5	4.1	8.0

年度	2007	2008	2009	2010	2011	2012	2013
件数	10	15	8	8	13	6	11
単体性能向上比	33.0	9.3	47.0	47.2	16.2	19.7	16.7
並列性能向上比	1.9	5.1	3.6	48.5	17.2	15.3	12.9

謝辞

本稿を執筆するにあたり、東北大学大学院理学研究科 河野研究室、日本電気株式会社、NEC ソリューションイノベータ株式会社、NEC フィールディング株式会社、日本 SGI 株式会社の皆様をはじめ、多くの方々にご協力ご支援をいただきました。心より感謝申し上げます。

参考文献

[1] 東北大学情報部情報基盤課 共同利用支援係，共同研究支援係，東北大学サイバーサイエンスセンター スーパーコンピューティング研究部,「並列コンピュータ LX 406Re-2 の利用法」，SENAC Vol.47 No.2（2014.4），p1-24，2014

[2] 日本電気株式会社 島本浩樹，小林公雄，長沢富人，「LX 406Re-2 のハードウェア」，SENAC Vol.47 No.3（2014.7），p7-14，2014

[3] 日本 SGI 株式会社 桐山智文，朝倉博紀，庄司岳史,「三次元可視化システムの利用法」，SENAC Vol.47 No.3（2014.7），p15-25，2014

[4] Kaoru Yamazaki, Takashi Nakamura, Naoyuki Niitsu, Manabu Kanno, Kiyoshi Ueda, and Hirohiko Kono ,「Two-step explosion processes of highly charged fullerene cations C60 q + (q = 20-60)」，The Journal of Chemical Physics 141, 121105 (2014)