



TOHOKU
UNIVERSITY

ISSN 0286-7419

東北大学
サイバーサイエンスセンター

大規模科学計算システム広報

SENAC

Vol.48 No.1 2015-1



Cyberscience
Center

Supercomputing System
Tohoku University

www.ss.cc.tohoku.ac.jp

大規模科学計算システム関連案内

<大規模科学計算システム関連業務は、サイバーサイエンスセンター本館内の情報部情報基盤課が担当しています。>

http://www.ss.cc.tohoku.ac.jp/

階	係・室名	電話番号(内線)* e-mail	主なサービス内容	サービス時間
				平日
一階	利用相談室	022-795-6153 (6153) sodan05@cc.tohoku.ac.jp 相談員不在時 022-795-3406 (3406)	計算機利用全般に関する相談 大判プリンタ、利用者端末等の利用	8:30～17:15 8:30～21:00
	利用者談話室	(3444)	各センター広報の閲覧	8:30～21:00
	展示室(分散 コンピュータ博物館)		歴代の大型計算機等の展示	9:00～17:00
	可視化機器室	(3428)	三次元可視化システムの利用	9:00～21:00
三階	総務係	022-795-3407 (3407) som@cc.tohoku.ac.jp	総務に関すること	8:30～17:15
	会計係	022-795-3405 (3405) kaikai@cc.tohoku.ac.jp	会計に関すること、負担金の請求に関すること	8:30～17:15
	共同研究支援係	022-795-6252 (6252) rs-sec@cc.tohoku.ac.jp	共同研究、計算機システムに関すること	8:30～17:15
	共同利用支援係 (受付)	022-795-3406 (3406) 022-795-6251 (6251) uketuke@cc.tohoku.ac.jp	利用手続き、利用相談、講習会、ライブラリ、アプリケーションに関すること	8:30～17:15
	ネットワーク係	022-795-6253 (6253) net-sec@cc.tohoku.ac.jp	ネットワークに関すること	8:30～17:15
四階	研究開発部	022-795-6095 (6095)		
五階	端末機室	(3445)	PC 端末機(X 端末)	

* () 内は東北大学内のみの内線電話番号です。青葉山・川内地区以外からは頭に 92 を加えます。

本誌の名称「SENAC」の由来

昭和 33 年に東北地区の最初の電子計算機として、東北大学電気通信研究所において完成されたパラメロン式計算機の名前で SENAC-1 (SENdai Automatic Computer-1) からとって命名された。

[巻頭言]

「普通の人々のためのスーパーコンピュータセンター」を目指して

サイバーサイエンスセンター長 小林広明

東北大学サイバーサイエンスセンターは、その前身の大型計算機センターが 1969 年に大学の教員、その他の研究者が学術研究等のために利用する全国共同利用施設として設置されて以来、45 年が経過しました。その間、大学の萌芽的研究から産業化を見据えた実用的な研究に至るまで、多様な分野での HPC ニーズに応えるべく、「最先端かつ世界最大級のコンピュータシステムの導入と利用環境の構築・高度化、次世代コンピュータシステムの要素技術の研究開発」に取り組んで参りました。

その一方、サイバーサイエンスセンターを含む全国共同利用型情報基盤センターを取り巻く環境は大きく変化し、特に、京コンピュータと基盤センター群のスーパーコンピュータを統一した利用環境で国内の研究者に幅広く提供する「HPCI システム」構築後は、エクサフロップス級の性能を目指す「ポスト京システム」の研究開発と併せて、「ポスト京」時代の基盤センター群の役割などが文科省の有識者会議で検討されています。

そのような背景のもと、サイバーサイエンスセンターでは、里見 進東北大総長の 2017 年度までの全学横断的な取り組みを示す「里見ビジョン」達成に向けて、サイバーサイエンスセンターの部局ビジョンを 2014 年に策定し、それに基づいた取り組みを進めております。具体的には、HPC に関して以下の 3 つの柱を基本にしています。

1. 「京」および「ポスト京」を補完する世界最先端の高性能計算基盤の整備、運用
2. 高性能計算に関する研究機能の強化、および実践的、かつ学際的な人材の育成
3. HPC を通じた産業、および社会貢献

これら 3 つについて簡単に現在の取り組みの状況をご紹介します。まず、1. については、2014 年に新スーパーコンピュータ棟の建設を完了し、2015 年 1 月から試験運用、2 月 20 日から本格運用を予定しています。新スーパーコンピュータシステム SX-ACE は現在運用の SX-9 の 25 倍以上の性能となる 707Tflop/s の演算性能を有するものですが、他のスーパーコンピュータシステムと大きな違いは総メモリデータ転送性能 655TB/s、総容量 160TB と他のセンターのペタフロップス級のスーパーコンピュータの同性能・容量を大きく超えるものとなっている点です。私どもが NEC と共同で取り組んだ SX-ACE に関する研究成果の 1 つをご紹介しますと、TOP500 で使われる LINPACK ベンチマークの役割を補い、より実際のアプリケーションの特性をとらえて開発が進められている新しいベンチマークプログラム HPCG による性能評価では、他のペタフロップス級のスーパーコンピュータがピーク性能の 1%~4%程度しか性能を引き出せないのに対し、私どもが導入する SX-ACE は 10%を超える実効性能を達成し、2014 年 11 月に米国ニューオーリンズで開催された SC14 の HPCG に関するセッションにおいて、その

潜在能力が高く評価されました。現在の多くのアプリケーションがメモリ性能制約であることから、新スーパーコンピュータ **SX-ACE** は、過度の並列処理に頼ることなく多くのシミュレーションプログラムを高い効率で高速に処理してくれるものと期待しております。

2. については、1997年9月より取り組んで来た利用者プログラムの高速化支援体制を強化すべく、サイバーサイエンスセンターに **HPC** に関する産学連携研究拠点として「高性能計算技術開発 (NEC) 共同研究部門」を2014年7月に新設し、活動を開始しました。この共同研究部門は、将来の日本のスーパーコンピューティング環境の発展や産業競争力の向上に貢献すると共に、研究活動を通じて計算機科学と計算科学の発展を担う人材の育成を目的としています。そのために、本研究部門には、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部門の教員、および情報部情報基盤課技術職員に加え、NEC からアプリとシステムを専門にする2名の客員教授と客員准教授、神戸大学から1名の客員教授を迎え、センター利用者にとって真に役に立つ学術情報基盤の整備・運用・研究開発にこれまで以上に密接に連携して取り組める体制を構築しました。

3. については、これまでも国産近距離航空機 **MRJ** の開発の支援などに取り組んで参りましたが、今後は地域企業からの **HPC** ニーズに応えられるように、スーパーコンピュータ利用環境の整備、および利用支援に取り組んで行きたいと考えております。また、東日本大震災の被災地の大学として **HPC** 技術がどのように我が国の安全安心化に貢献できるかを考え、2014年から本学災害科学国際研究所と連携し、高知市、静岡市、石巻市、東松島市などの自治体の協力を得て、サイバーサイエンスセンターの新スーパーコンピュータを活用したリアルタイム津波浸水・被害予測システムの研究開発に取り組んでいます。本システムは、大規模地震発生時の津波による沿岸都市部での浸水被害を高精度、かつ迅速に予測し、発生から20分以内に関係自治体に10メートル四方単位で浸水範囲や深さの情報を提供するものです。2015年4月からの本格運用に向けて、現在実証実験に取り組んでいるところです。

スーパーコンピュータの整備・運用・研究開発を議論する際に、かならず出てくるのが **TOP500** でのランキングやそのトレンドに基づく未来予測であり、まさに我が国も含めて各国が **TOP500** で外挿されたエクサフロップスを目指して、**HPC** の研究開発力を誇示するような状況にありますが、もはやハイエンドシステムのピークフロップス値が **LINPACK** や特定の分野のアプリでしか意味をなさない状況を考えると、本当に役に立つスーパーコンピュータシステムとはどのようなものか、改めて考える時期に来ていると思います。ピーク演算性能重視で世界最高性能のスーパーコンピュータが導入できても、それが利用分野を選ぶようになっては、目的と手段が逆になり、意味がありません。汎用のスーパーコンピュータセンターとして大学に附置され、その多様な **HPC** ニーズ、特に、長いスパンで丁寧な支援が求められる萌芽的研究や近年の産業利用に応えるべく発展してきたサイバーサイエンスセンターは、これからも「普通の人々のためのスーパーコンピュータセンター(The Supercomputer Center for the Rest of US)」の精神を見失うことなく、スーパーコンピュータシステムの整備・運用・利用支援・研究開発に取り組んで行きたいと思っております。ここに改めて関係各位のご理解、ご支援、ご協力をお願い申し上げます。

[お知らせ]

スーパーコンピュータシステムの更新に伴う今後の運用スケジュールについて

2月以降の運用スケジュールをお知らせいたします。更新情報の詳細については変更となる場合がありますので、ホームページ <http://www.ss.cc.tohoku.ac.jp/> にて最新情報をご確認いただくようお願いいたします。

1. 運用スケジュール

スーパーコンピュータシステムの更新とそれに伴う並列コンピュータシステムや、その他のシステムのサービス停止についてお知らせいたします。

○スーパーコンピュータシステム

現行のスーパーコンピュータ SX-9 の運用は平成 27 年 2 月 10 日に終了し、次期スーパーコンピュータ SX-ACE は平成 27 年 2 月 20 日から運用を開始します。システム更新作業に伴って 2 月 10 日 9:00~20 日 10:00 までの間サービスを停止します。

○並列コンピュータシステム LX 406Re-2

スーパーコンピュータの更新作業に伴って、並列コンピュータ LX 406Re-2 のサービスを 2 月 10 日~2 月 20 日まで停止します。

○三次元可視化システム及び大判カラープリンタ

三次元可視化システム及び大判カラープリンタは、並列コンピュータシステムと同じ期日（2 月 10 日~20 日）にサービスを停止します。

2. 停止期間中におけるジョブの扱い

停止期間中のジョブの扱いはシステムによって異なります。

○スーパーコンピュータシステム SX-9

システム停止日は 2 月 10 日ですが、ジョブの受付は前日の 2 月 9 日 17:00 で停止します。2 月 10 日 9:00 時点で残っている全てのジョブを強制終了します。

○並列コンピュータシステム LX 406Re-2

システム停止日は 2 月 10 日ですが、ジョブの受付は 1 月 30 日 17:00 で停止します。2 月 10 日 9:00 時点で残っているジョブを強制終了し継続処理は行いません。

3. ホームディレクトリ下のファイル移行

ホームディレクトリ下にあるファイルは、2月20日の運用再開後も利用可能です。2月10日時点のファイルを、センターにて新ストレージシステムに移行します。これ以外の大規模ファイル領域（short 領域）のファイルは移行対象外となります。

4. 試験運用

スーパーコンピュータ SX-ACE は、平成 27 年 2 月 20 日～2 月 27 日までを試験運用期間とし、2 月 27 日（金）10:00 までに終了したジョブの演算負担金については課金対象外とします。なお、試験運用期間中は、実行時間や利用資源量などに制限を設けることがありますのであらかじめご了承ください。

SX-ACE の利用方法に関する情報や利用講習会の開催については、センターホームページ <http://www.ss.cc.tohoku.ac.jp/> と「大規模科学計算システムニュース」にて、準備が整い次第お知らせする予定です。

運用スケジュール（スーパーコンピュータシステム）

	2 月						
		2/10		2/20		2/27	
現行スーパーコンピュータ SX-9		*1 運用終了					
次期スーパーコンピュータ SX-ACE				*2 試験運用		*3 通常運用	

*1 SX-9 システム停止：2/10(火)9:00、ジョブの受付は前日 17:00 に停止
 *2 SX-ACE 試験運用開始：2/20(金)10:00
 *3 SX-ACE 通常運用開始：2/27(金)10:00

運用スケジュール（その他のシステム）

	2 月			
		2/10		2/20
並列コンピュータ LX406Re-2		*システム停止		
三次元可視化システム・大判プリンタ		*システム停止		

* システム停止：2/10(火)9:00～ 2/20(金)10:00

（共同研究支援係，共同利用支援係）

[大規模科学計算システム]

スーパーコンピュータ SX-ACE のハードウェア

萩原 孝 浜口 博幸 山信田 恒

日本電気株式会社

1. はじめに

東北大学サイバーサイエンスセンターで本年2月より運用開始される、新スーパーコンピュータシステム SX-ACE のハードウェアを紹介します。

SX-ACE は、増大する科学技術計算需要に応えるためのシステム規模の拡大とシステムの省電力化/省スペース化を両立させることを目標に開発しました。これを実現するために、SX-9 の大規模共有メモリ型ノードを少数接続するマルチノードシステムから、単一 CPU ノードを多数接続するマルチノードシステムにシステム・アーキテクチャを刷新しています。本稿では、SX-ACE の本体ハードウェアのシステム概要、CPU コア、CPU ノード、マルチノードの構成、及び LSI 技術、実装技術を紹介します。

2. システム概要

SX-ACE システムは、4つの CPU コアで構成される単一 CPU ノードを総計 2560 ノード、総理論演算性能 707TFLOPS (総ベクトル演算性能 655TFLOPS)、総メモリ容量は 160TB を有したシステムです。このシステムは、512 ノードを1つのクラスタとして5クラスタで構成されます。クラスタ内のノード間は、ノードあたり 4GB/秒 × 2 (双方向) でノード間スイッチ (Inter-node X-bar Switch: IXS) で接続されます。システムの主な諸元とハードウェアの階層構成を表 2-1、図 2-1 に示します。

表 2-1 SX-ACE システム主要諸元

CPU ノード	CPU コア数	4
	CPU 数	1
	理論演算性能 ^{*1} /ベクトル演算性能 ^{*2}	276GFLOPS/256GFLOPS
	メモリ容量/メモリ帯域	64GB/256GB/秒
	ノード間通信性能	4GB/秒 × 2 (双方向)
クラスタ	CPU コア数	2048
	CPU 数	512
	理論演算性能 ^{*1} /ベクトル演算性能 ^{*2}	141TFLOPS/131TFLOPS
	メモリ容量/メモリ帯域	32TB/131TB/秒
	ノード間通信性能	2TB/秒 × 2 (双方向)
システム	CPU コア数	10240
	CPU 数	2560
	理論演算性能 ^{*1} /ベクトル演算性能 ^{*2}	707TFLOPS/655TFLOPS
	メモリ容量/メモリ帯域	160TB/655TB/秒
	ノード間通信性能	10TB/秒 × 2 (双方向)

*1) 同時に並列演算することができるすべての演算器(ベクトルユニットの乗算演算器、加算演算器、除算/平方根演算器、ならびにスカラユニットの浮動小数点演算器)における浮動小数点演算処理能力の総和。

*2) 同時に並列演算することができるベクトルユニット内の乗算演算器、加算演算器のみによる浮動小数点演算処理能力の総和。

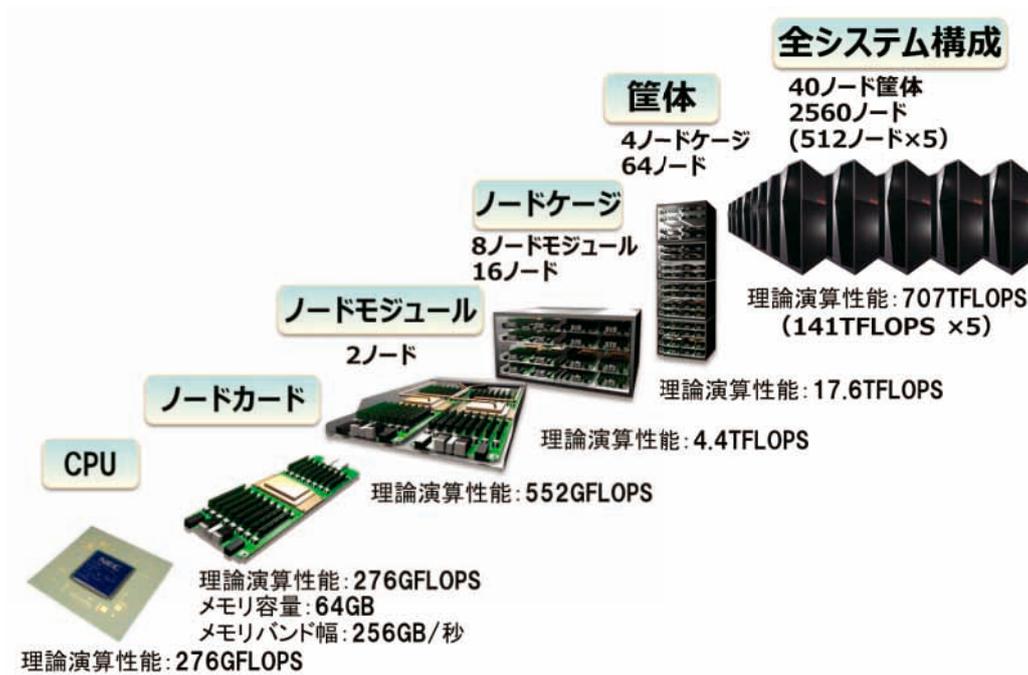


図 2-1 SX-ACE ハードウェア階層とシステム構成

3. ハードウェア・アーキテクチャ

SX-ACE は、これまでの SX シリーズと同様に高メモリバンド幅を必要とする多くの科学技術計算アプリケーションで高い実効性能を提供するスーパーコンピュータシステムです。主な特長としては、

■ **ビッグ・コア、高メモリバンド幅設計による高実効性能**

- 64GFLOPS の世界トップクラスの CPU コア性能
- コアあたり 64GB/秒の高メモリバンド幅

■ **1CPU/ノード化、先端テクノロジー採用による省電力化/省スペース化**

- 同一ピーク性能比較で前機種 SX-9 の 10 分の 1 の低電力化と、5 分の 1 の省スペース化(保守エリア含む)

が挙げられます。以下では、SX-ACE のハードウェア構成と特長を中心に紹介します。

3.1 CPU コア構成

SX-ACE の CPU コアは、従来の SX アーキテクチャを継承しつつ、さらにマイクロ・アーキテクチャレベルで改良を加え、高性能化を図っています。図 3-1 に CPU コア構成を示します。CPU コアは、スカラー処理部 (Scalar Processing Unit: SPU)、ベクトル処理部 (Vector Processing Unit: VPU)、ソフトウェア制御可能な高速バッファ部 (Assignable Data Buffer: ADB)、冗長なメモリアクセスを抑止する MSHR (Miss Status Handling Register) で構成されています。SPU は、全ての命令発行制御を行い、ベクトル命令やノード間通信命令を VPU やノード間通信制御部 (Remote access Control Unite: RCU) に発行します。SX-ACE では、SPU から VPU への命令発行専用のパスを設け、SX-9 に比べてベクトル命令の発行をより効率よく行えるようにしています。VPU は、16 本のベクトル・パイプラインから構成され、SX-9 の 2 倍のベクトル・パイプライン数となっています。これにより、1 つのベクトル演算命令を 1 サイクルあたり 16 演算のスループットで処理することで、

CPU コアのベクトル演算性能(ベクトル乗算・加算演算性能)は、64GFLOPS を実現しています。ベクトル長 (VL: Vector Length)は、SX-9 と同じ最大 256 です。各ベクトル・パイプラインの内部構成は、SX-9 と同様、乗算器×2、加算/シフト演算器×2、除算/平方根演算器×1、論理演算器×1 の 6 種のそれぞれ独立に動作可能な演算器と、マスク演算パイプライン、ロード/ストアパイプライン、マスクレジスタ、及びベクトルレジスタにより構成されています。SX-ACE では、乗算器や加算器などの主要演算器間で先行する演算結果を後続の演算器へ直接フォワーディングするパスを設け、ベクトル長の短いアプリケーションの性能向上を図っています。

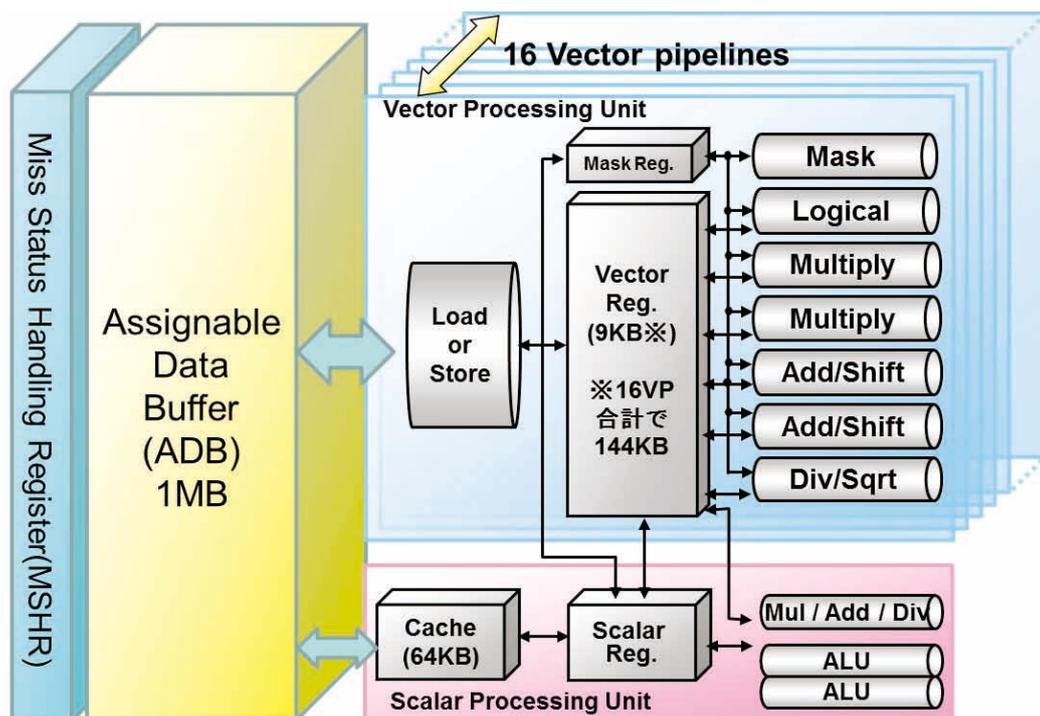


図 3-1 CPU コア構成

SX-9 から導入した ADB は、SX-ACE では SX-9 の 4 倍の 1MB の容量となります。ADB から VPU へは最大 256GB/秒(4Bytes/Flop) のデータ供給性能を備えています。ADB は、ラインサイズ単位(128B)で管理され、ADB を使用するベクトルデータの指定方法としては、コンパイラによる自動最適化、コンパイラ・オプション指定、およびプログラム中に指示行で指定する方法があります。SX-ACE では、ADB の容量増に伴い、コンパイラによる自動最適化や ADB にバッファリングするベクトルデータを指定するコンパイラ・オプションを強化し、ADB をより有効に活用しやすくしています。

また、SX-ACE では、新たに冗長なメモリアクセスを削減するために MSHR を実装しています。図 3-2 に MSHR の活用例を示します。科学技術計算プログラムでは、格子上の 1 点を計算するために周囲点をアクセス(決まった型のアクセス)するステンスル計算が多くみられます(図 3-2 a))。ベクトル機で周囲点をアクセスする場合、例に示す配列 A の参照は 4 個のベクトルロード命令でアクセスされます。このうち $A(i-1, j)$ と $A(i+1, j)$ の参照は、例えば $i=2\sim 257$ では、それぞれ $A(1:256, j)$ 、 $A(3:258, j)$ をアクセスし、254 要素が重なります(図 3-2 b))。MSHR はこの重なった要素に対する二重のメモリアクセスを抑止する機構で、この機構により、 $A(i+1, j)$ のメモリアクセスは、 $A(257:258, j)$ の 2 要素となり、CPU の外部メモリ帯域を有効活用する役割を果たします(図 3-2 c))。

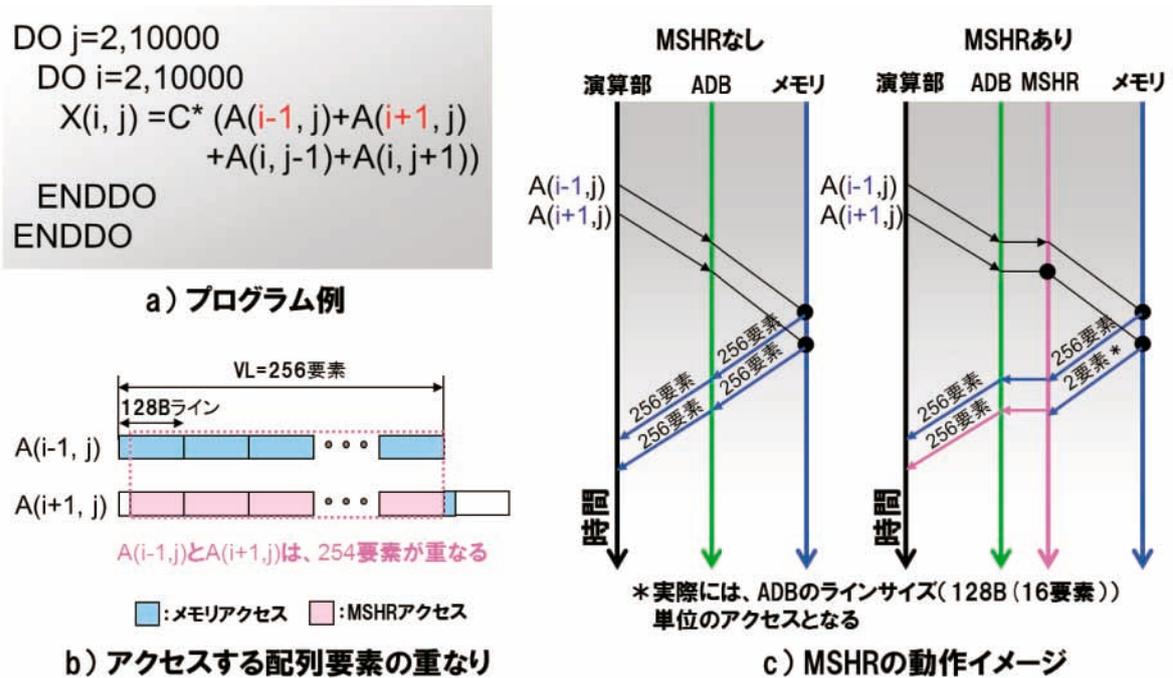
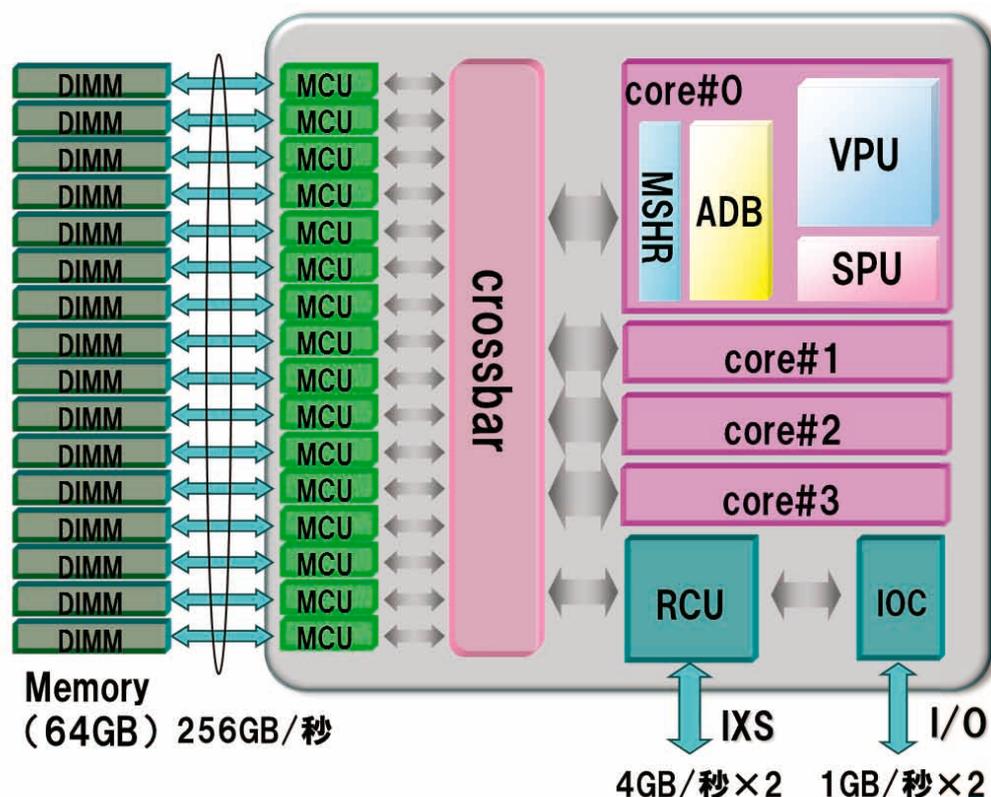


図 3-2 MSHR

3.2 CPU ノード構成

CPU は 4 つの CPU コア、16 個のメモリ制御部(Memory Control Unit: MCU)、RCU、IO 制御部(Input Output Control unit: IOC)と 4 つの CPU コア、RCU と MCU 間を結ぶクロスバー・スイッチ部から構成されます。この CPU と 16 枚の DIMM(メモリ)をカード上に実装したものを CPU ノードと呼びます。図 3-3 に CPU ノード構成を示します。4 つの CPU コアと 16 個の MCU の間のクロスバー・スイッチは、4 つの CPU コアが同時にメモリアクセスした場合は、各 CPU コアのメモリバンド幅は平均 64GB/秒となりますが、1 つの CPU コアのみを利用した場合や各 CPU コアのメモリアクセスタイミングが重ならない場合は、1CPU コアで 256GB/秒の全メモリバンド幅を使用することが可能になっています。また、このクロスバー・スイッチ内には 4CPU コア間の高速な同期処理を実現するための通信レジスタ(Communication Registers)を備えています。

SX-ACE では 1 つの CPU(LSI)と 16 枚の DIMM だけでノードを構成することで物理的に非常にコンパクトな構成となり、省電力化・省スペース化を実現するとともにメモリアクセスのレイテンシを SX-9 の約 1/2 に短縮しています。これにより、スカラ・キャッシュミス時間の削減、メモリからのベクトルデータのロードを伴うベクトル処理の立ち上がり時間の短縮を実現し、リストベクトルや短ベクトル長のアプリケーションの性能向上を図っています。



SPU: Scalar Processing Unit, VPU: Vector Processing Unit, ADB: Assignable Data Buffer
 MSHR: Miss Status Handling Register, RCU: Remote Access Control Unit,
 MCU: Memory Control Unit, IOC: Input Output Control unit

図 3-3 CPU ノード構成

流体計算などの科学技術計算アプリケーションで高い実効性能を得るためには、高い演算性能に見合うだけの高いメモリ性能が必要となります。SX-ACE の CPU に搭載された 16 個の MCU は、外部メモリに効率よくアクセスするため、メモリバンク競合で待たされている先行するロード/ストア・リクエストを後続リクエストが追い越す機能や、ストアデータをメモリアクセス単位である 128B のブロックにまとめ、バンク競合を低減して効率よくストア処理するストアデータのマージ機能などを有しています。

RCU は専用開発の IXS を介したノード間通信を制御するユニットです。RCU は 4 つの CPU コアで共有され、各 CPU コアからのノード間通信リクエストの管理機構、論理ノード番号から物理ノード番号へのアドレス変換機構、自 CPU ノード内のメモリアクセスのアドレス変換機構などを備えます。

一般的な分散メモリ型の大規模並列システムでは、IO 機能を一部のノードに集約するか、専用の IO ノードを設ける方式が取られますが、SX-ACE では、各ノードに IOC を設け、SX-ACE の高いプロセッサ性能、及びメモリ性能に見合った高速なデータ入出力転送性能を実現しています。今回のシステムではノードあたり 1GB/秒×2(双方向)の IO 性能を有しています。

3.3 マルチノード構成

SX-ACE の 1 クラスタのマルチノード構成図を図 3-3 に示します。クラスタの総理論演算性能は 141TFLOPS、総主記憶容量は 32TB になります。ひとつのクラスタは 512 ノードで構成され、各ノードは IXS を介して接続されます。IXS は専用開発の 32 ポートを持つスイッチ LSI を使って、2 段のフル・ファット・ツリー・

トポロジーを採用しています。RCU のデータ受信部、及び送信部は、それぞれ独立に動作可能であり、ノード当たり4GB/秒×2(双方向)の通信バンド幅を実現しています。また、RCUはCPUとは独立に動作するデータムーバ(DMA エンジン)を持つことにより、ノード間のデータ転送をCPU 動作とは独立して行なうことが可能です。512ノードのクラスタのバイセクション・バンド幅は、2TB/秒になります。SX-ACEのIXSは、SX-9のIXS同様に高速なハードウェア・バリア同期機構を持ち、MPI 通信の同期処理などの高速化を実現しています。

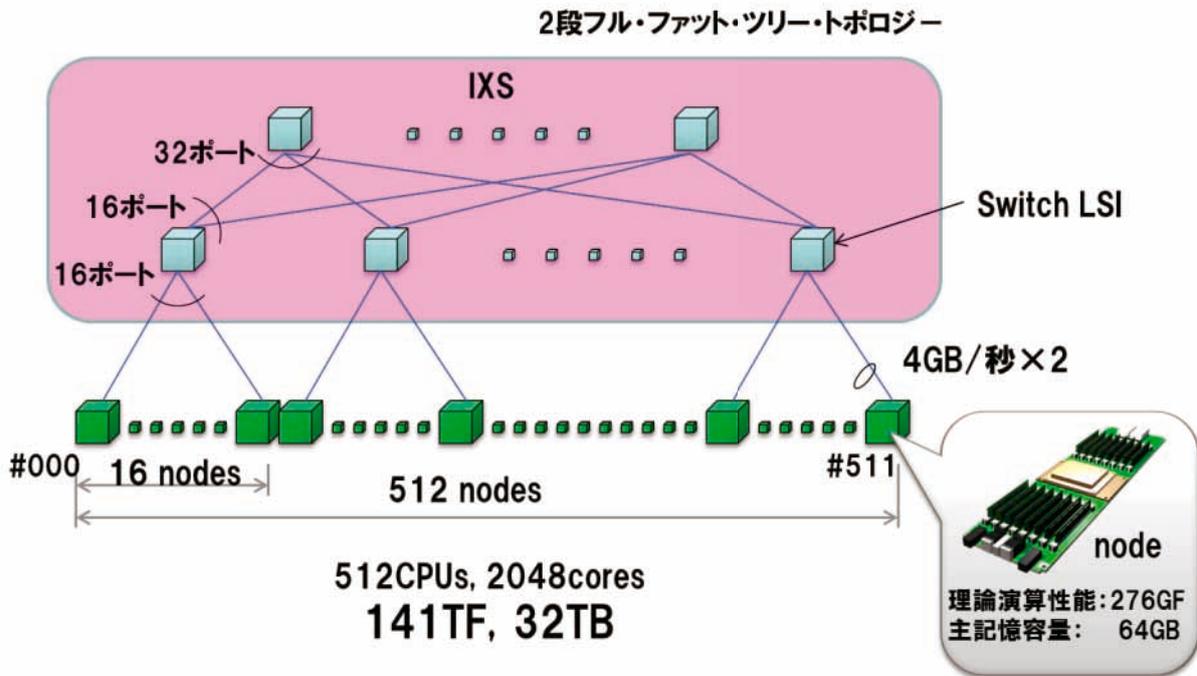


図 3-3 マルチノード構成(接続図)

4. SX-ACE のテクノロジー

4.1 LSI 技術

これまで SX シリーズでは CMOS テクノロジによる高集積化, 及びプロセッサの並列化により高性能化を実現させてきました。SX-ACE では、さらに高い性能を実現するために LSI 技術及び回路技術を発展させています。より高いメモリバンド幅を実現するために、インタフェース回路の低消費電力化、小面積化により LSI への多チャンネル搭載を実現しています。また、高速化、高集積化による消費電力の増大を抑えるために様々な工夫を施しています。表 4-1 に SX-ACE の CPU チップ諸元を示します。

表4-1 CPUチップ諸元

テクノロジーノード	28nm
搭載トランジスタ数	20億個
電源電圧	0.9V
ピン数(内信号ピン)	4,344(2,586)
配線層構成	銅11層
実装形態	BGA (Bump pitch:0.8mm)

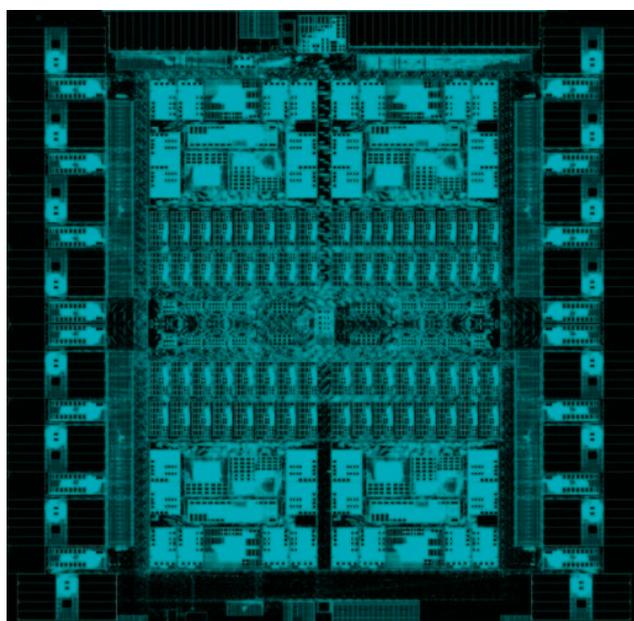


図 4-1 CPU チップレイアウト

SX-ACEに使われるLSIは、28nm CMOSプロセス、11層銅配線テクノロジーの採用による高速・高集積化、低電源電圧による低消費電力化、及び大容量オンチップキャパシタの搭載による高いノイズ耐力を実現しています。さらに、低消費電力実現のための設計上の工夫として、Multi-Vthトランジスタの最適な使い分け、レジスタ単位およびCPUコア単位でのクロックゲーティング、チップ内各種センサを用いた電源電圧の最適制御

など、最先端の低消費電力化技術を盛り込んでいます。

システムの処理能力の向上には、LSI 内信号伝送の高速化とともに、LSI 間信号伝送の高速化と並列化が必要となります。信号伝送の高速化や並列化を妨げる要因となる電源ノイズ対策も重要です。SX-ACE では高速、かつ安定した信号伝送を実現するために、伝送信号の波形を改善するエンファシス、イコライズ機能を備えた回路、及び波形ひずみの少ないコネクタなどを採用しています。また隣接する信号の配線間隔を始めた信号種に応じた配線ルールを定め、配線間のクロストークを低減しています。さらにトランジスタが高速化し、電源電流の時間変化が大きくなることにより電源ノイズが増加するため、デカップリング用コンデンサの搭載数や LSI パッケージの層構成を最適化するなどにより電源ノイズの低減を実現しています。

4.2 実装技術

SX-ACE では、実アプリケーションの高い実効性能と省スペース化のために、高密度接続技術、高密度配線基板技術、および高効率冷却技術により高いメモリバンド幅と高密度実装を実現しています。

高いメモリバンド幅を実現するために 4,000 ピン超の CPU パッケージと多ピン化対応と微細化を追求した高多層+HDI (High Density Interconnect) 基板を採用しています。また、省スペース化を実現するためにコンパクトな水冷構造と高密度ノードモジュールを採用しています。

表 4-2 に CPU パッケージ、ノード基板、およびノードモジュールの諸元を示します。また、図 4-2 から図 4-5 に CPU パッケージ、ノード基板、ノードモジュール、およびノード筐体を示します。

CPU パッケージは、他社比 2 倍以上の 4,344 (信号数:2,596) の IO 端子数を有しています。また、CPU LSI を保護する Lid を無くし、CPU LSI を直接冷却することで高い冷却効率を実現しています。

ノード基板は、スルーホールピッチが 0.8mm、34 層という多層化により信号配線収容性を高めるとともに、表裏面 1 層に Line/Space が 65/70 μm と微細な配線層を形成することによって多ピン CPU パッケージの搭載を可能としています。また、カスタム化された Mini-DIMM を搭載して高メモリバンド幅と高密度化を実現しています。

ノードモジュールには、2 つのノード基板、CPU を冷却するコールドプレート、および外部と信号接続する光モジュールが搭載されています。また、CPU 冷却にコンパクトな水冷構造を採用することでノードモジュールを高密度化しています。

ノード筐体には、ノードモジュールが 8 個で 1 ユニットとして 4 ユニット、つまり 32 ノードモジュール (64 ノード) 搭載できるようになっており、筐体あたりのベクトル演算性能は SX-9 の 10 倍の 16.4TFLOPS を実現しています。

表 4-2 SX-ACE 実装技術の諸元

項目		諸元
CPU パッケージ	材料	セラミック
	パッケージ外形	55mm×55mm×2.0mm
	IO 端子ピッチ	0.8mm ピッチ
	IO 端子数	4,344
ノード基板	構造	HDI(High Density Interconnect)
	基板サイズ	370mm×110mm t=4.9mm(層数:1-34-1)
	配線仕様	スルーホールピッチ:0.8mm、L/S = 65/70 μm
	実装物	表: CPU, DIMM, 高速伝送用コネクタ 裏: 電源部品
ノードモジュール	シャーシサイズ	253mm×420mm t=60mm
	実装物	表: CPU×2, DIMM×32(RAM×320), 光 mod×4 裏: DC-DC コン, 高速伝送用コネクタ×4
ノード筐体	筐体サイズ	1200mm×700mm
	実装物	Node モジュール×32(64 ノード), 電源
	性能	16.4TFlops/筐体



図 4-2 CPU パッケージ(TOP 面)



図 4-3 ノード基板

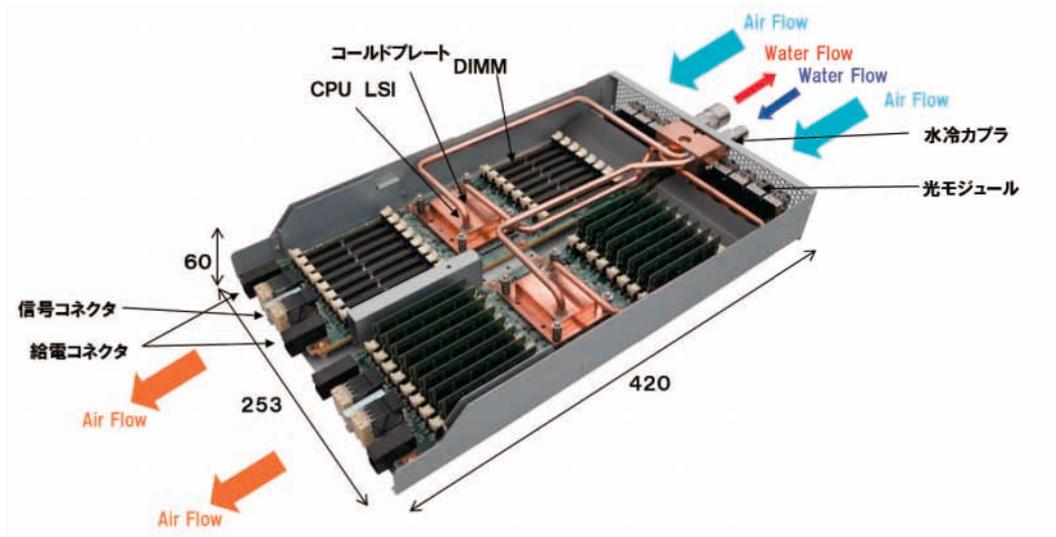


図 4-4 ノードモジュール

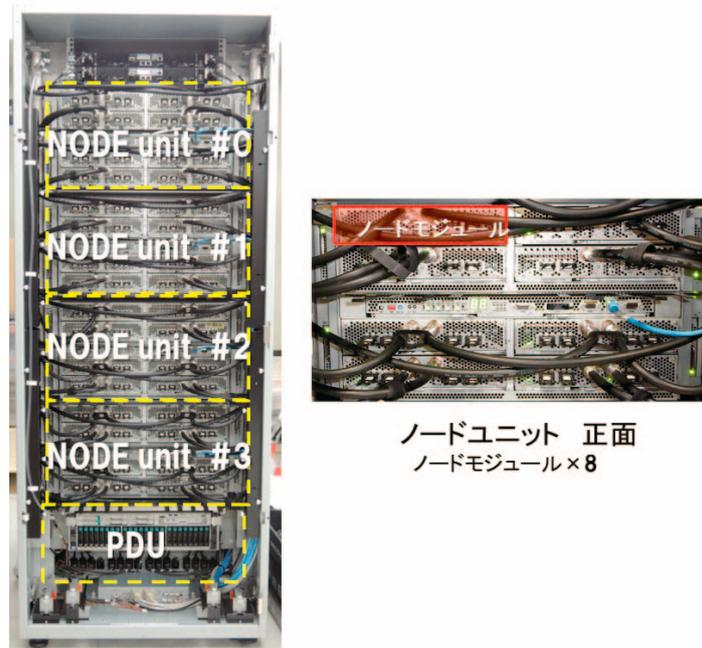


図 4-5 ノード筐体

5. おわりに

SX-ACE は、従来機種比で同一演算性能あたり、10 分の 1 の低消費電力化、5 分の 1 の省設置スペース化を達成しつつ、スーパーコンピュータの要件である高い演算性能と、それに見合うメモリからのデータ供給性能のバランスを重視した分散メモリ型のスーパーコンピュータ・システムとして開発しました。NEC は、今後も様々な研究分野の発展を支える強力なツールであるスーパーコンピュータを開発していきます。

参考文献

- [1] Shintaro Momose, Takashi Hagiwara, Yoko Isobe, Hiroshi Takahara, “The Brand-New Vector Supercomputer, SX-ACE”, Proceedings of ISC14, 2014
- [2] Shintaro Momose, “SX-ACE Processor: NEC’s Brand-New Vector Processor”, HOTCHIPS26, 2014
- [3] Akihiro Musa, Yoshiei Sato, Takashi Soga, Ryusuke Egawa, Hiroyuki Takizawa, Koki Okabe, Hiroaki Kobayashi, “Effects of MSHR Prefetch Mechanisms on an On-Chip Cache of the Vector Architecture”, Proceedings of International Symposium on Parallel and Distributed Processing with Applications, 2008.

[大規模科学計算システム]

SX-ACE でのプログラミング(ベクトル化編)

工藤 淑裕 横谷 雄司

日本電気株式会社

概要

SX-ACE システムのハードウェア性能を引き出すために重要となるプログラムのベクトル化、並列化に関わるコンパイラの機能、およびプログラミングの際にご留意頂きたい点について、「ベクトル化編」、「並列化編」の2回に分けてご紹介します。

SX-ACE システムでは主なプログラミング言語として、Fortran 95、Fortran 2003、C、C++言語が利用できます。本稿では、SX-ACE システムの Fortran 95 言語コンパイラである FORTRAN90/SX (以降、単にコンパイラと略す)を用いて、コンパイラが持つ自動ベクトル化機能の特長、ベクトル化プログラミングでの性能向上についてご紹介します。

1. SX-ACE の CPU

SX-ACE システムは、4 個の CPU コアを内蔵した CPU を持つノードを複数結合した「スケーラブル・パラレル・スーパーコンピュータ」です。各 CPU コアは一つの命令で最大 256 個のデータを処理できるベクトル演算器を 6 個(加算器 2 個、乗算器 2 個、除算器 1 個、論理演算器 1 個)ずつ含みます。

SX-ACE のベクトル命令には、一般的なスカラプロセッサで導入されている複数のデータを一度に処理できる SIMD 命令に対して、次の特長があります。

- 最大 256 個までの大量のデータを一つのベクトル命令で演算できる
- 一つのベクトル命令で処理するデータの個数を自在に変更できる

SX-ACE の CPU コアのハードウェア性能を十分に引き出すためには、個々の CPU コアの中でベクトル命令を使って効率的に計算するためのベクトル化が大変重要となります。

2. 自動ベクトル化機能

SX-ACE システムでのベクトル化技法は、SX-9 システムの場合と基本的には同じですが、今回はスーパーコンピュータを初めて使われる方にもご理解頂けるようベクトル化の基本概念からご紹介します。

2.1. ベクトル化の基本概念

通常の演算命令は、一度に一組のデータを演算できます。このような演算命令をベクトル命令と対比させるためにスカラ命令と呼びます。これに対してベクトル命令は、複数の組のデータに対する演算を一つの命令で行うことができます。

例-1 のプログラムは、配列 B と配列 C を加算し配列 A に代入、配列 E と配列 F を加算し配列 D に代入する DO ループです。

例-1

```
DO I = 1, 100
  A(I) = B(I) + C(I)
  D(I) = E(I) + F(I)
ENDDO
```

例-1 の DO ループ中の加算をスカラ命令、ベクトル命令で実行したときの実行イメージは図-1 のとおりです。一つのスカラ命令は、一度に一組のデータに対する演算処理を行います。これに対して、SX-ACE のベクトル命令では一つの命令で一度に複数(この場合は 100 個、最大 256 個)のデータに対する演算処理を行うことができます。

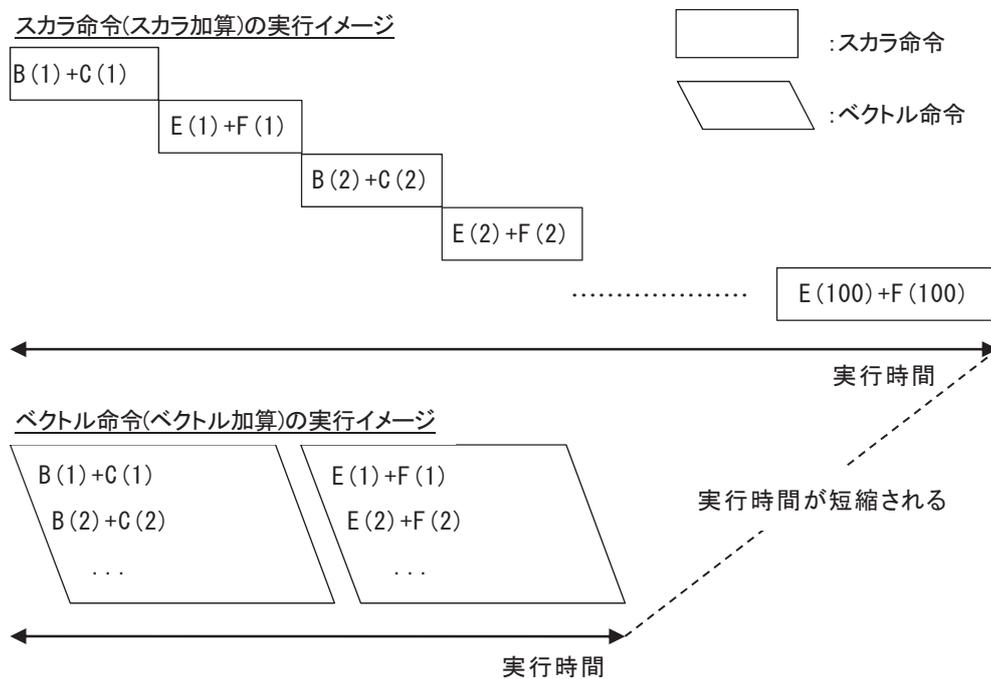


図-1 加算の実行イメージ

ループ中で計算される行列の要素など、規則的に並んだ配列データに対してベクトル命令を適用することをベクトル化(の適用)と呼び、ベクトル化することによって高速な演算が可能となります。

コンパイラの自動ベクトル化機能は、ソースプログラムを解析し、ベクトル命令で実行できる部分を自動的に検出しベクトル化を適用します。

2.2. ベクトル命令の適用例

例-2 は、DO ループが自動ベクトル化されたときのベクトル命令の適用例です。二つの配列のメモリロード、ベクトル加算、メモリストアの四つのベクトル命令で実行されます。

例-2

```
DO I = 1, 100
  C(I) = A(I) + B(I)
ENDDO
```

↓

```
VR1 ← 配列A      (配列Aからベクトルレジスタに100個のデータをロード)
VR2 ← 配列B      (配列Bからベクトルレジスタに100個のデータをロード)
VR3 ← VR1 + VR2  (100個のデータをベクトル加算)
配列C ← VR3      (配列Cに100個の演算結果をストア)
                                VRn: ベクトルレジスタ
```

2.3. ベクトル化の対象範囲

自動ベクトル化機能は、表-1 で示すループ、および、ループに含まれる文、データ、演算を対象としてベクトル化を適用します。

表-1 自動ベクトル化の対象

対象	Fortran 言語要素
ループ	配列式、DO ループ、DO WHILE ループ、FORALL ループ
文	代入文、CONTINUE 文、GOTO 文、CYCLE 文、EXIT 文、IF 文、SELECT 構文(CALL 文、入出力文等は不可)
データ型	INTEGER(KIND=4)、INTEGER(KIND=8)、REAL(KIND=4)、REAL(KIND=8)、COMPLEX(KIND=4)、COMPLEX(KIND=8)
演算	加減乗除算、べき算、論理演算、関係演算、型変換、組込み関数(利用者定義演算、ポインタ代入等は不可)

2.4. データの依存関係

ループにベクトル化を適用するには、「2.3. ベクトル化の対象範囲」で示した対象範囲に加えて、ベクトル化の適用による文、演算の実行順序が変更されても、データの定義・参照順番(データの依存関係)が変わらないことが条件となります。

例-3 は、二つの配列を定義する DO ループです。図-2 は、例-3 の DO ループのベクトル化適用前、すなわち、スカラでの文の実行順序とベクトル化後の文の実行順序のイメージを示しています。

例-3

```
DO I = 1, 100
  A(I) = B(I) + C(I)
  D(I) = E(I) + F(I)
ENDDO
```

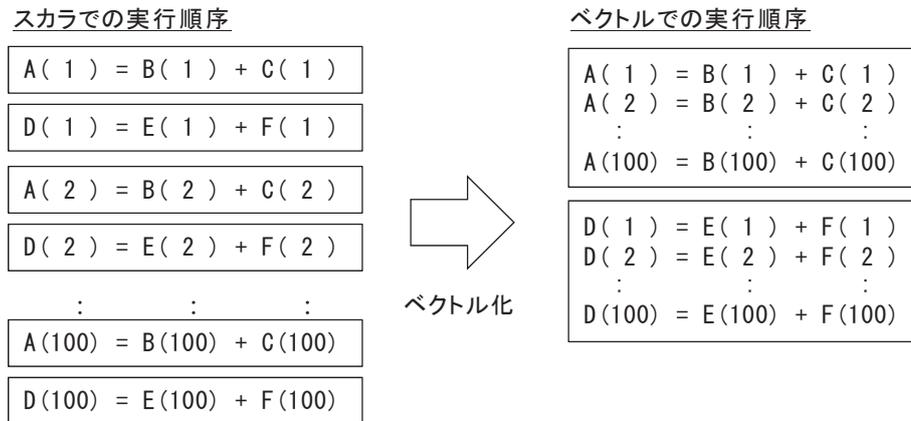


図-2 文の実行順序

スカラでの実行では、 $A(1)=B(1)+C(1)$ 、 $D(1)=E(1)+F(1)$ 、 $A(2)=B(2)+C(2)$ 、...と実行されますが、ベクトル化した場合は一つの命令で複数のデータを処理するため、 $A(1)=B(1)+C(1)$ 、 $A(2)=B(2)+C(2)$ 、...、 $A(100)=B(100)+C(100)$ 、 $D(1)=E(1)+F(1)$ 、 $D(2)=E(2)+F(2)$ 、...、 $D(100)=E(100)+F(100)$ と配列 B、配列 C の加算、配列 A への代入後、配列 E、配列 F の加算、配列 D への代入が実行されます。ベクトル化した場合、このように文、演算の実行順序が変わります。

このような文の実行順序の変更により、データの依存関係が変わってしまうことがあります。例-4 はベクトル化後にデータの依存関係が変わってしまう例です。以前の繰返して定義された配列要素や変数を後の繰返して参照するパターンのとき、図-3 のようにデータの依存関係が変わり正しい結果が得られなくなるため、コンパイラはベクトル化を適用しません。

例-4

```
DO I = 2, N
  A(I+1) = A(I) * B(I) + C(I)
ENDDO
```

図-3 は、例-4 の DO ループの文の実行順序を示すイメージです。ベクトル化すると、「2.2. ベクトル命令の適用例」で示した例-2 のように、配列 A のデータをまとめてメモリからロードするため、更新された A の値が使用されず正しい演算結果が得られません。

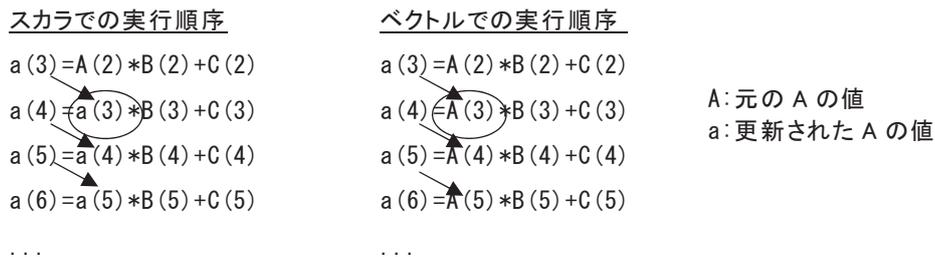


図-3 ベクトル化を阻害するデータの依存関係

例-5 はベクトル化を適用してもデータの依存関係が変わらない例です。

例-5

```
DO I = 2, N
  A(I-1) = A(I) * B(I) + C(I)
ENDDO
```

図-4 は、例-5 の DO ループの文の実行順序を示すイメージです。ベクトル化してもデータの依存関係は変わらないのでコンパイラはベクトル化を適用します。

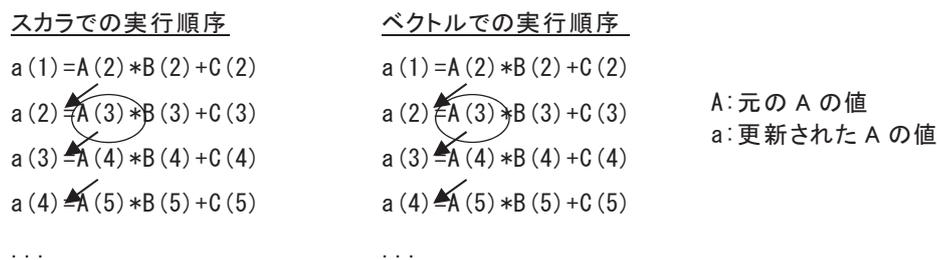


図-4 ベクトル化可能なデータの依存関係

コンパイラの自動ベクトル化機能は、ソースプログラムを解析して、ベクトル命令で実行できる部分を自動的に検出するとともに、必要ならベクトル化に適合するようにプログラムを変形して、ベクトル化を適用できる範囲を広げます。

プログラムのループを記述する際、文をループの繰返しごとに並べたとき、図-3 のような右下向きの矢印の依存関係ができないようにすると、ループにベクトル化を適用でき、ループ内の処理を高速化できる可能性が高まります。

3. 拡張ベクトル化機能

コンパイラの自動ベクトル化機能は、ベクトル化を適用できる範囲を広げるため、データの依存関係などの理由でベクトル化できないループを変形してベクトル化したり、プログラムを変形することによってベクトル化の効果をさらに高めたりします。これを拡張ベクトル化機能と呼びます。ここでは、コンパイラの持つ拡張ベクトル化機能のうち主なものをご紹介します。

3.1. 文の入れ換え

「2.4 データの依存関係」の例のようなループは、ループ中の文を入れ換えるとベクトル化できることがあります。コンパイラは、実行結果が正しく保てるのであれば、例-6 に示すように自動的に文を入れ換えてベクトル化を適用します。

例-6

(文の入れ換え前)

```

DO I = 1, 99
  A(I) = 2.0      ! 定義
  B(I) = A(I+1)  ! 参照
ENDDO

```

(文の入れ換え後のイメージ)

```

DO I = 1, 99
  B(I) = A(I+1)  ! 参照
  A(I) = 2.0    ! 定義
ENDDO

```

3.2. ループの一重化

ループの一重化は、多重ループの外側のループと内側のループを一つのループにまとめて、ループの繰返し数を大きくする最適化です。

SX-ACEのベクトル命令は一つの命令で最大256個のデータを処理できます。よって、ベクトル命令で処理するときにはできるだけ256個ずつ演算した方が命令の実行回数を減らすことができ効率が良くなります。さらに外側ループの繰返し制御のための時間も省くことができます。

自動ベクトル化では最内側ループをベクトル化の対象とします。例-7では最内側ループの繰返し数は最大でも100回と256より短いので、このままでは一度に100個ずつ処理するようベクトル化されてしまいます。このループに一重化を適用すると、繰返し数を10,000(=100×100)にでき、256個ずつ処理できるようになります。

例-7

(ループの一重化前)

```

INTEGER::M, N
PARAMETER(M=100, N=100)
REAL(KIND=8)::A(M, N), B(M, N), C(M, N)
DO I = 1, M
  DO J = 1, N
    A(J, I) = B(J, I) + C(J, I)
  ENDDO
ENDDO

```

(ループの一重化後の変形イメージ)

```

DO IJ = 1, M*N
  A(IJ, 1) = B(IJ, 1) + C(IJ, 1)
ENDDO

```

3.3. ループの入れ換え

多重ループのとき、ループを入れ換えることによりベクトル化できないデータの依存関係の問題が解消されてベクトル化できるようになる場合や、内側のループよりも外側のループの繰返し数が大きく、入れ換えた方が効率がよいと判断された場合には、例-8のようにコンパイラが自動的にループを入れ換えてベクトル化を適用します。

例-8

(ループの入れ換え前)

```

DO J=1, 1000
  DO I = 1, 999
    A(I+1, J) = A(I, J) + B(I, J)
  ENDDO
ENDDO

```

(ループの入れ換え後のイメージ)

```

DO I = 1, 999
  DO J = 1, 1000
    A(I+1, J) = A(I, J) + B(I, J)
  ENDDO
ENDDO

```

なお、SX-ACE では、ハードウェア特性を考慮し、ループの繰返し数を長くするよりも配列要素の連続アクセスを優先して最適化することがあります。すなわち、外側ループの繰返し数が内側より大きく SX-9 ではループ入れ換えされていたループが、内側ループが連続アクセスできるときは SX-ACE ではループ入れ換えの対象とならない場合があります。

3.4. 条件ベクトル化

条件ベクトル化とは、一つのループに対してあらかじめ二つ以上の命令コードを用意しておき、実行時に最も効率よくに実行できる命令コードを選択して実行するベクトル化です。

例-9 は、データの依存関係がベクトル化に適合しているかどうかコンパイル時に不明であったとき、ベクトル化した命令コードとベクトル化しない命令コードの両方を用意しておき、プログラムを実行するときそのどちらかを選択して実行する条件ベクトル化の例です。

例-9

(条件ベクトル化前)

```
DO I = N, N+10
  A(I) = A(I+K) + B(I)
ENDDO
```

(条件ベクトル化後のイメージ)

```
IF (K.GE.0.OR.K.LT.-10) THEN
!CDIR NODEP                ! ベクトルでの実行
  DO I = N, N+10
    A(I) = A(I+K) + B(I)
  ENDDO
ELSE
  DO I = N, N+10            ! スカラでの実行
    A(I) = A(I+K) + B(I)
  ENDDO
ENDIF
```

例-9 の条件ベクトル化後のイメージの IF 文が実行時にコードを選択するための判定文で、THEN ブロックはベクトルで実行するコード、ELSE ブロックはスカラで実行するコードです。IF 文の条件式に現れる「K.GE.0」が真のときは例-5 と同じようにデータの依存関係が変わりません。また、「K.LT.-10」のときには、「DO I=N,N+10」であることから常に $I \neq I+K$ が成り立ちデータの依存関係がありません。よって、条件式が真のときベクトルでの実行(THEN ブロック)となります。逆に、IF 文の条件式が偽であるとき、「2.4 データの依存関係」の例-4、図-3 で示したようなベクトル化後にデータの依存関係が変わってしまう(右下向きの矢印ができてしまう)のでスカラでの実行(ELSE ブロック)となります。

例-9 はデータの依存関係に着目した条件ベクトル化(依存関係による条件ベクトル化)です。他の条件ベクトル化として、ループの繰返し数による条件ベクトル化も行います。

4. プログラミングの際の留意点

Fortran 言語でプログラムを作成する場合、実行性能の面で注意が必要な言語機能があります。このセクションでは、使い方によって性能に大きな影響が出る可能性のある、配列代入文と配列ポインタおよび形状引継ぎ配列を使う場合の留意点について説明します。

4.1. 配列代入文

Fortran 言語では、配列データの演算を一括して行う配列式や配列代入文が利用できます。配列式や配列代入文は基本的にベクトル化され高速に実行できますが、複数の配列代入文を書く場合、同じ形状の配列代入文をなるべく連続して書くようにしてください。

同じ形状の配列代入文が連続するとき、コンパイラはそれらの配列代入文を1個のループにまとめてベクトル化するため、より効率のよい実行命令列が得られます。

例-10

```
SUBROUTINE SUB (A, B, C, D, N, X, Y, M)
  REAL :: A (N), B (N), C (N), D (N)
  REAL :: X (M), Y (M)

  A = B**2 + C**2      ①
  X = SIN (Y)          ②
  D = SQRT (A)         ③

  END SUBROUTINE SUB
```

例-10 では、配列 A、B、C の大きさが N、配列 X、Y の大きさが M となっています。すなわち、大きさ N の配列代入文①と③の間に、大きさが M の配列代入文②が割り込む形になっています。この場合、例-11 のように 3 個のループで書いた場合と同じような実行命令列が生成されます。

例-11

```
DO I = 1, N
  A (I) = B (I)**2 + C (I)**2
ENDDO
DO I = 1, M
  X (I) = SIN (Y (I))
ENDDO
DO I = 1, N
  D (I) = SQRT (A (I))
ENDDO
```

これを、例-12 のように同じ大きさの配列代入文①と③を並べて書くようにすれば、コンパイラはこの二つをまとめて一つのループとして実行命令列を作成します。

例-12

```

A = B**2 + C**2      ①
D = SQRT (A)        ③
X = SIN (Y)         ②

```

上記コードに対するコンパイラの最適化イメージ

```

DO I = 1, M
  A(I) = B(I)**2 + C(I)**2
  D(I) = SQRT (A(I))
ENDDO
DO I = 1, N
  X(I) = SIN(Y(I))
ENDDO

```

複数の配列代入文が一つのループにまとめられたかどうかは、最適化診断メッセージ、または、編集リストで確認できます。診断メッセージで確認するときは-Wf,-pvctl,fullmsg、編集リストで確認するときは-R2、または、-Wf,-L,fmllist のコンパイラオプションを指定してください。例-12 で二つの配列代入文が一つのループにまとめられた場合のメッセージとリストは例-13 のようになります。

例-13**最適化診断メッセージ**

```
f90: opt(11): ex. f90, line 5: 配列代入を融合した。:line 5 - 6
```

編集リスト

LINE	LOOP	FORTRAN STATEMENT
1:		SUBROUTINE SUB (A, B, C, D, N, X, Y, M)
2:		REAL::A (N), B (N), C (N), D (N)
3:		REAL::X (M), Y (M)
4:		
5:	V----->	A = B**2 + C**2
6:	V-----	D = SQRT (A)
7:	V=====	X = SIN (Y)
8:		
9:		END SUBROUTINE SUB

編集リスト上で、7 行目のように、「V=====」と表示されている場合は、その配列代入文が単独でループになっていることを示しています。この表示が複数並んでいるときは、同じ形状の配列代入文がないか、もしあったらそれらを連続するように並べ換えできないか確認してください。

4.2. 配列ポインタ、形状引継ぎ配列

通常、配列の各要素はメモリ上連続して配置されています。しかし、配列ポインタと形状引継ぎ配列の場合、隣接する要素がメモリ上不連続になることがあります。

例-14

```

INTERFACE
  SUBROUTINE SUB(X)
    REAL::X(:)          ! 形状引継ぎ配列
  END SUBROUTINE SUB
END INTERFACE

REAL, POINTER::P(:)    ! 配列ポインタ
REAL, TARGET::A(100)
REAL::B(5, 5)

P=>A(1:100:2)          ! PはAの奇数要素と結合
CALL SUB(B(1, :))     ! SUBの形状引継ぎ配列XはBの1次元目の添字が1の要素と結合
...

```

例-14において、ポインタ代入後の配列ポインタPと、サブルーチンSUBが呼び出されたときの形状引継ぎ配列Xの要素のメモリ上の配置は図-5のようになります。

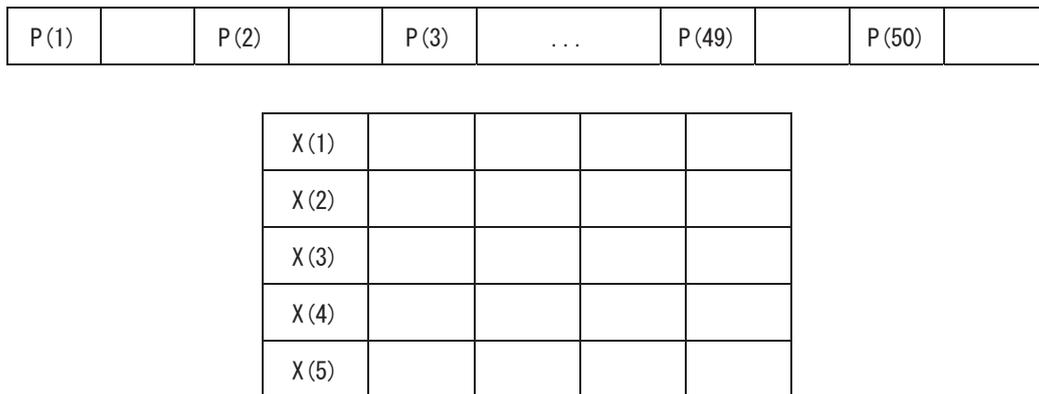


図-5 メモリ上の配置

通常は、このようにメモリ上不連続となっていることをあまり意識する必要はありませんが、これらの配列を手続の実引数に指定したときに性能上問題となることがあります。

例-15

```

INTERFACE
  SUBROUTINE SUB2(Y)
    REAL::Y(50)
  END SUBROUTINE SUB2
END INTERFACE

REAL, POINTER::P(:)    ! 配列ポインタ
REAL, TARGET::A(100)
P=>A(1:100:2)          ! PはAの奇数要素と結合
CALL SUB2(P)

```

例-15では、配列ポインタPをサブルーチンSUB2の実引数としています。Pに対応するSUB2の仮配列は、Y(50)と宣言されています。Fortran言語では、配列ポインタと形状引継ぎ配列以外の配列の要素はメモ

り上連続領域でなければならないと定められています。このため、コンパイラは SUB2 を呼び出す前に P の要素が連続領域となっているかどうかをチェックし、もし不連続なら P の全要素の値を一旦作業用の領域にコピーし、その作業領域を SUB2 に渡し、SUB2 から戻った時点で作業領域から P に値を書き戻す処理を挿入します。

コンパイラが生成する処理のイメージは、以下のようになります。

例-16

```
f_copyin(P, wk)      ! この実行時ルーチンはPが連続領域でなければ作業領域に
                    ! コピーしwkにその作業領域のアドレスを設定する。
                    ! Pが連続ならwkはPの結合先のアドレスが設定される。
CALL SUB2(wk)        ! 連続領域であることが保証されているwkをSUB2に渡す。
f_copyback(wk, P)   ! この実行時ルーチンはwkとPが同じメモリ領域なら何も
                    ! しない。そうでなければwkからPに値をコピーする。
```

P が連続領域でない場合、2 回コピー処理が行われ、その分余計な時間がかかることとなります。P が連続領域である場合でも、コピーほどではないにせよ、連続領域かどうかのチェックのためのオーバーヘッドがかかります。連続領域へのコピー処理が生成されたかどうかは、コンパイラオプション-Wf,-msg,o を指定したとき出力される以下の注意メッセージで確認できます。

例-17

注意メッセージ

```
f90: observ(1): wk.f90, line 11: 実引数の配列を連続領域にコピーするコードを生成した。
```

このようなコピー処理が生成されるのを避けるには、ポインタまたは形状引継ぎ配列を実引数として結合される可能性のある仮引数を形状引継ぎ配列として宣言します。

以下に例を示します。

例-18

```
INTERFACE
  SUBROUTINE SUB3(Z)
    REAL::Z(:)      ! Zは形状引継ぎ配列
  END SUBROUTINE SUB3
END INTERFACE

REAL, POINTER::P(:)  ! 配列ポインタ
REAL, TARGET::A(100)

P=>A(1:100:2)        ! PはAの奇数要素と結合
CALL SUB3(P)        ! Pと対応するZが形状引継ぎ配列であるためコピー処理は生成されない。
```

配列ポインタに関しては、もう一つ注意点があります。

Fortran 言語では、同一の配列が複数の異なる配列ポインタと結合することを許しています。配列ポインタが実際にどの配列と結合しているかは、コンパイル時には確定できませんので、言語仕様に忠実に解釈すると、型と次元数が同一であるすべての配列ポインタおよびそれと結合可能な配列の間には、依存がある可能性を考慮してベクトル化することになります。

たとえば、以下の例-19 において、配列ポインタ P1 と P2 が同一の配列と結合していた場合、依存関係があるものとみなされるためベクトル化できません。

例-19

```
REAL, POINTER::P1(:), P2(:)
REAL::X(100)
DO I = 1, 100
    P1(I+10) = P2(I) + X(I)  ! P1とP2が同一配列と結合しているとベクトル化できない依存関係がある
ENDDO
```

しかし、実際のプログラムでは、同一の配列を別々の配列ポインタに結合して一つのループ中でそれらを混在させて使用するようなことはほとんどありません。そのため、コンパイラの既定値の動作では、異なる配列ポインタは別の配列に結合しているものと仮定して、例-19 のようなループにも最適化・ベクトル化を適用します。

もし、上記のような仮定が成り立たないループをベクトル化すると正しい実行結果が得られません。そのようなプログラムをコンパイルする場合は、コンパイラオプション-Cvsafe を指定してください。なお、このオプションを指定すると、最適化・ベクトル化に様々な制約が課せられるため既定値と比べてベクトル化できなくなるループが増加してしまいます。したがって、一つの手続内で同一配列を複数の異なる配列ポインタに結合して使うことは避けてください。

5. ベクトルデータ・バッファリング機能

5.1. ADB

SX システムでは、ADB(Assignable Data Buffer)と呼ばれるハードウェア機構を利用したベクトルデータ・バッファリング機能を利用することで、ベクトル命令によるメモリアクセス性能を改善できます。

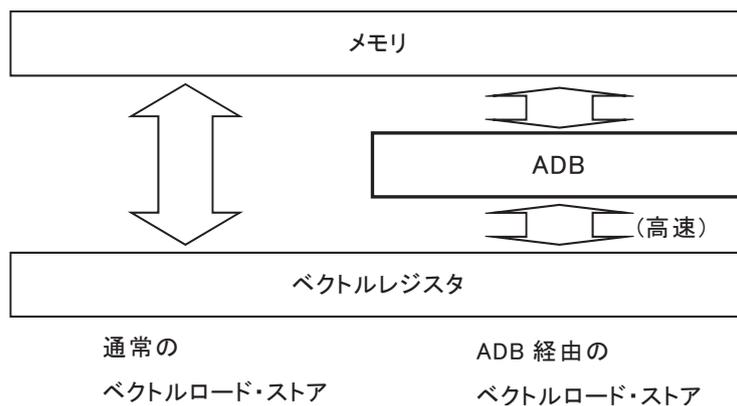


図-6 ADB (Assignable Data Buffer)

ADB を利用するとき、配列データの最初のベクトルロード、または、ベクトルストア時に、データが ADB にバッファリングされます。次にそのデータを利用するときには、ADB にバッファリングされたデータがロードされます。ADB からはメモリからロードするより高速にデータをロードできるため、2 回目以降のベクトルロードを高速化できます。

コンパイラは、配列データの再利用性、すなわち、ある配列データが 2 回以上アクセスされるかどうかに着目し性能向上が期待できる配列を自動的に選択し ADB にバッファリングされるようにします。

5.2. SX-ACE の ADB

ADB は SX-9 で初めて導入されましたが、その容量が SX-ACE では CPU コアあたり 1 メガバイトに大きく増加されました。また、ハードウェアのバッファリング機構の仕組みも改善され、バッファリングされたデータをより高速に置き換えられるようになりました。これに伴い、コンパイラでの ADB の利用ポリシーが変更されました。

表-2 SX-ACE と SX-9 の ADB

	SX-ACE	SX-9
ADB 容量/コア	1 メガバイト	256 キロバイト
利用ポリシー	再利用されることが明確なデータに加えて、リストベクトル、作業ベクトル、等間隔ベクトルなどで再利用性が不明な場合でも積極的に ADB にバッファリングし、できるだけ多くのベクトルデータをバッファリング	再利用性のあるデータをコンパイラが自動選択し、バッファリング
コンパイラオプション	<code>-pvctl,on_adb=</code> カテゴリ指定のコンパイラオプションが追加	<code>-pvctl,on_adb</code>
コンパイラ指示行	右の指示行に加えて、データをプリフェッチするための指示行が追加	<code>!CDIR ON_ADB</code>

5.3. コンパイラオプションによるカテゴリ指定

ADB の容量には限りがあるため、すべての配列データを ADB にバッファリングすることはできません。配列の大きさが ADB の容量(1メガバイト)以下でありアクセス頻度が高い配列データを選択的にバッファリングする方が性能向上を期待できます。

コンパイラは、こういった配列データをコンパイラが自動的に選択する機能の他に、コンパイラ指示行 `!CDIR ON_ADB` でユーザが配列を指定する機能、特別な属性を持つ配列データをコンパイラオプションでまとめて指定する機能(コンパイラオプションによるカテゴリ指定機能)を持ちます。コンパイラオプションによるカテゴリ指定機能は SX-ACE で新たに強化されました。

コンパイラオプションによるカテゴリ指定では、連続、リストベクトルなどのアクセスパターンごと、共通ブロック、モジュール変数、引数などのデータの種類ごと、特定の用途で利用されるものごとなど、バッファリング

する配列データをその種類によって一括指定できます。

表-3 コンパイラオプションによるカテゴリ指定

カテゴリ指定	バッファリングされるベクトルデータ
-pvctl,on_adb= arg	仮引数、実引数
common	共通ブロック要素
contiguous	連続ベクトル
indirect	リストベクトル
module	モジュール変数
reuse	コンパイラが自動選択した再利用されるベクトルデータ
stride	等間隔ベクトル
work	部分ベクトル化で利用される作業ベクトル

カテゴリを複数指定するときは、「:」(コロン)で区切って指定します。カテゴリ指定の前に「no」を付けると、そのカテゴリに属するデータを ADB にバッファリングしなくなります。カテゴリ指定を省略したとき、contiguous 以外のすべてのカテゴリが指定されたものとみなされます。例-20 にコンパイラオプションの指定例を示します。

例-20

```
% cat a.f90
SUBROUTINE SUB(C)
COMMON /COM/A, B
REAL (KIND=8) :: A(1000), B(1000), C(1000)
B = A * 2
A = B + C
END SUBROUTINE
% sxf90 -Wf, -pvctl, on_adb=arg:nocommon a.f90
```

5.4. ADB ヒット要素率

ADB にバッファリングされたデータの再利用性を確認するための性能値として「ADB ヒット要素率」がプログラム実行解析情報、簡易性能解析機能/Ftrace Viewer の解析リスト/グラフに出力されます。出力形式については、「6. 性能解析」を参照してください。

あるベクトルデータをベクトルロードする際、その配列データがあらかじめ ADB にバッファリングされていたとき ADB ヒットとみなされます。ADB ヒット要素率は、ベクトルロードした全要素のうち ADB ヒットした要素

の比率を表します。

この性能値の利用にあたっては次のことにご留意ください。

- ある配列要素が二度ベクトルロードされる時、1 回目は必ずミスヒット、2 回目は 1 回目でバッファリングされていればヒットとなります。ADB ヒット要素率の値は大きいほどよいですが、アクセスパターン、頻度によっては値が小さくても性能上問題ありません。
- 手続中に再利用されるベクトルデータがない(少ない)、すなわち、二度以上アクセスされるデータがないとき、ADB ヒット要素率の値は 0% に近づきます。

6. 性能解析

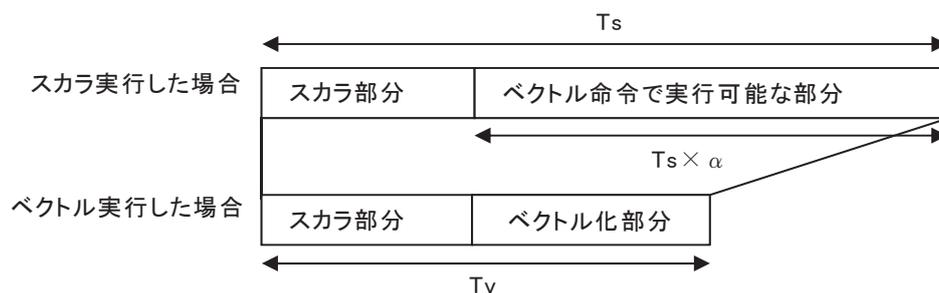
SX-ACE においては、プログラムに含まれるループをできるだけベクトル命令を使って、データを最大ベクトル長である 256 個ずつ実行できているかどうかを性能分析のポイントになります。

本セクションでは、プログラムの性能を確認するための性能値の取得方法、ベクトル化が十分できていない手続を絞り込む方法などを説明します。

6.1. ベクトル化率

プログラムをスカラ命令だけで実行させた場合の実行時間に占めるベクトル命令で実行可能な部分の時間の割合をベクトル化率と呼びます。一般にベクトル化率を正確に求めることは困難であるため、ベクトル化率の近似値としてベクトル演算率を用います。

ベクトル演算率は、プログラムで処理された全演算数に占めるベクトル命令で処理された演算数の割合を求めたものです。SX-ACE では、このベクトル演算率を大きくすることを目標にチューニングしてください。ベクトル演算率は、プログラム中のループをベクトル化することにより実行性能が向上しますが、ベクトル演算率が 50% 程度では、スカラ実行時の高々 2 倍の性能にしかならないことは、図-7 からもわかると思います。一般にはベクトル演算率が 98% 以上を目指すことになります。ベクトル演算率は、以降で示すプログラム実行解析情報、簡易性能解析機能/Ftrace Viewer で参照できます。



T_s : スカラ実行したときの実行時間 α : ベクトル化率

T_v : ベクトル実行したときの実行時間

図-7 ベクトル化率

6.2. プログラム実行解析情報

プログラム実行解析情報は、プログラムの実行時に参照される環境変数 F_PROGINF に YES、または、DETAIL が設定されているとき、プログラムの実行終了時に標準エラー出力ファイルに出力されます。この情報から、プログラムがハードウェアの性能を十分に引き出しているか否かを判断できます。環境変数 LANG に値 japan、または、ja_JP.UTF-8 が設定されているときは日本語、そうでないときは英語で出力されます。

例-21 にプログラム実行解析情報の出力例を示します。

例-21

***** Program Information *****			
Real Time (sec)	:	875. 787967	経過時間
User Time (sec)	:	3498. 747107	ユーザ時間
Sys Time (sec)	:	2. 455352	システム時間
Vector Time (sec)	:	3329. 276575	ベクトル命令実行時間
Inst. Count	:	597017752356.	全命令実行数
V. Inst. Count	:	464204772550.	ベクトル命令実行数
V. Element Count	:	100858351225030.	ベクトル命令実行要素数
V. Load Element Count	:	11548190101324.	ベクトルロード命令実行要素数
FLOP Count	:	43776313912608.	浮動小数点データ実行要素数
MOPS	:	28864. 951114	MOPS値
MFLOPS	:	12511. 997173	MFLOPS値
A. V. Length	:	217. 271250	平均ベクトル長
V. Op. Ratio (%)	:	99. 868490	ベクトル演算率
Memory Size (MB)	:	5376. 000000	メモリ使用量
MIPS	:	170. 637584	MIPS値
I-Cache (sec)	:	0. 605043	命令キャッシュミス時間
O-Cache (sec)	:	23. 921375	オペランドキャッシュミス時間
Bank Conflict Time			
CPU Port Conf. (sec)	:	35. 571349	CPUポート競合時間
Memory Network Conf. (sec)	:	1597. 203733	メモリネットワーク競合時間
ADB Hit Element Ratio (%)	:	25. 849985	ADBヒット要素率

プログラム実行解析情報のベクトル演算率が 98%未満のとき、プログラムを調べ、ベクトル化できていないループがないかなどを調査します。

6.3. 簡易性能解析機能／Ftrace Viewer

プログラムの規模が大きいとき、プログラム内のコードすべてについてループのベクトル化状況を調べるのは作業効率がよくありません。ベクトル性能を引き出すには、実行時間が長く、ベクトル演算率が低い手続に絞って調べるのが効果的です。この絞り込みには簡易性能解析機能、またはFtrace Viewerが有効です。

Ftrace Viewer では、手続ごとの実行回数、実行時間、FLOPS 値、ベクトル演算率、ADBヒット要素率などの様々な性能値に加えて、それらをグラフ (Function Metric Chart) 表示し、視覚的にプログラムの性能を分析できます。

図-8 は、実行時間とベクトル演算率を重ねて表示したグラフです。これを参照し、実行時間が長く、ベクトル演算率の低いものを探し、その中のループのベクトル化状況を調べます。

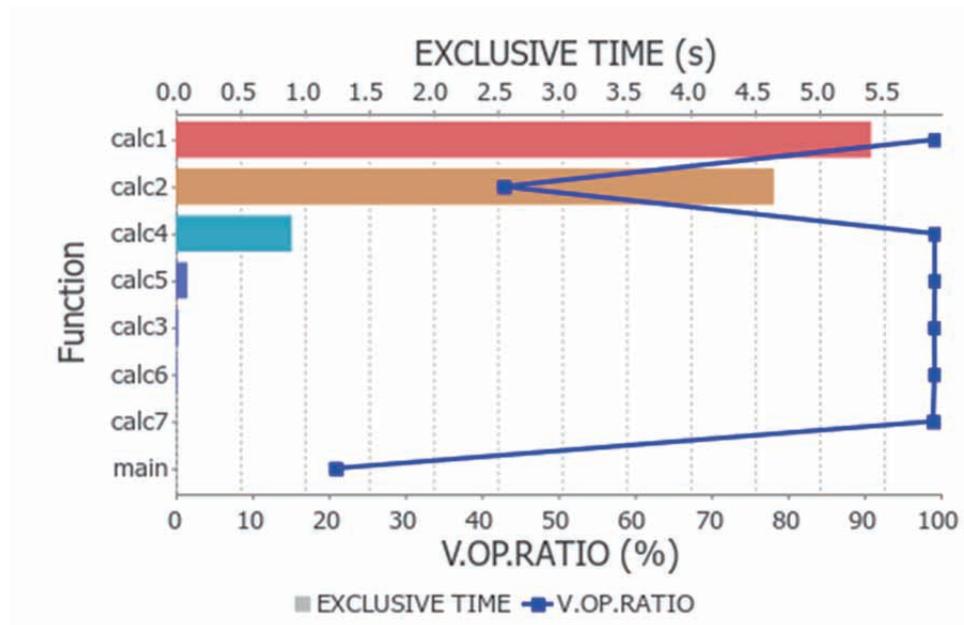


図-8 Function Metric Chart

図-8 では、縦軸が手続名、上横軸が実行時間(EXCLUSIVE TIME、単位は秒)、下横軸がベクトル演算率(V.OP.RATIO、単位は%)です。実行時間は棒グラフ、ベクトル演算率は折れ線グラフで表示されています。

このグラフから、実行時間が長くベクトル演算率が低い手続は「calc2」と分かります。「calc1」は実行時間は長いですがベクトル演算率も高く、ベクトル演算率を高めるという観点からのチューニングの余地は少なそうです。他のベクトル演算率の低い手続に main がありますが、これは実行時間が短いのでチューニングしてもプログラム全体の実行時間の短縮は望めません。

図-8 の例では、手続「calc2」に絞ってソースコードを調べる必要があると判ります。

個々のサブルーチン、手続の詳細な性能値を調べたいとき、グラフと一緒に表示されるテーブル(Profile Tree Table)を参照します。

PROC.NAM	FREQUENCY(#)	EXCLUSIVE TIME (s)	AVER.TIME (ms)	V.OP.RATIO (%)	ADB HIT ELEM.% (%)
Total	351	11.07	31.53	97.45	99.91
calc1	50	5.40	108.04	99.22	100.00
calc2	50	4.65	92.93	42.87	99.88
calc4	50	0.90	18.01	99.22	100.00
calc5	50	0.09	1.80	99.22	93.75
calc3	50	0.02	0.37	99.18	100.00
calc6	50	0.01	0.18	99.20	100.00
calc7	50	0.00	0.02	99.01	100.00
main	1	0.00	0.14	20.89	0.00

図-9 Profile Tree Table

図-9 から、手続「calc2」の実行回数(FREQUENCY)は 50 回、一回当たりの実行時間(AVER.TIME)は 92.93 ミリ秒で総実行時間は 4.65 秒だったことがわかります。さらに、ベクトル演算率(V.OP.RATIO)は 42.87%、ADB ヒット要素率は 99.88%だったこともわかります。

Ftrace Viewer で性能値を参照するには、コンパイラオプション-ftrace を指定してコンパイルされたプログラムを実行して、解析情報ファイル(ファイル名 ftrace.out.*.*)を出力することが必要です。

7. おわりに

以上、FORTRAN90/SX コンパイラの自動ベクトル化機能を中心にご紹介させて頂きました。SX-ACE のコンパイラは、他にも本稿でご紹介できなかった種々のベクトル化機能を持っています。詳細につきましては、マニュアル「FORTRAN90/SX プログラミングの手引」をご参照ください。

皆様が SX-ACE をご利用になる上で、本稿が多少なりともお役に立てれば幸いです。

参考文献

- [1] FORTRAN90/SX プログラミングの手引 日本電気株式会社 G1AF07
- [2] Ftrace Viewer 利用の手引 日本電気株式会社 G1AF33

[共同研究成果]

東北地震に伴う固有地震活動の揺らぎから推定された

摩擦特性と余効すべり伝播過程

- A trial estimation of frictional properties and afterslip process on the basis of aperiodicity off Kamaishi just after the 2011 Tohoku Earthquake-

有吉 慶介^{*1}・内田 直希^{*2}・松澤 暢^{*2}・日野 亮太^{*2}・
長谷川 昭^{*2}・堀 高峰^{*1}・金田 義行^{*3}

^{*1}独立行政法人 海洋研究開発機構 地震津波・防災研究プロジェクト

^{*2}東北大学大学院理学研究科附属 地震・噴火予知研究観測センター

^{*3}名古屋大学減災連携研究センター

釜石沖は再来間隔と規模が固有の地震であることが知られていたが、東北地方太平洋沖地震（東北地震）の直後から、それらに大きな揺らぎが生じたことが後の解析で明らかとなった。本研究では、その揺らぎを大規模な数値シミュレーションから再現することにより、釜石沖周辺での摩擦特性および余効すべり伝播方向などの推定を試みる。

1. はじめに

釜石沖（図1）では、再来間隔（5.5年）と規模（M4.8）がほぼ固有的に起きていること（図2a）が知られているが、東北地震の直後から、頻発する傾向が診られている（図2b）。

図2bを詳しく診ると、規模の大きな地震だけでなく、小さな地震も発生していることが分かる。規模の大きな地震が発生した要因として、プレート運動の速度が、東北地震による大規模なすべりによって加速された [1] ことなどが挙げられる。一方で、規模の小さな地震が発生する要因として、釜石沖の固有地震震源域の周辺にある複数の小さなアスペリティが壊れたためだという考え [2] もあるが、その場合、なぜ小さなアスペリティがごく限られた時しか起きないのかを説明するのが難しいように思われる。

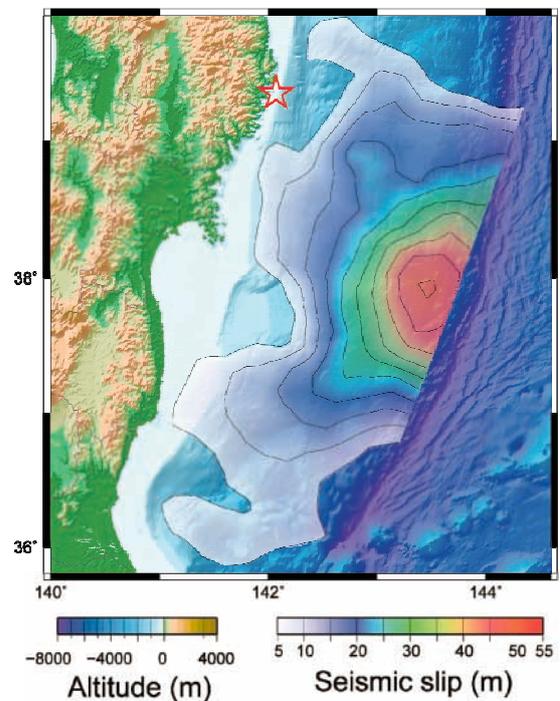


図1. 釜石沖の固有地震の震央（☆）と東北地震のすべり分布 [4] との関係。

そこで本研究では、大規模な余効すべりによる応力擾乱を受ける釜石沖地震について、大規模な数値シミュレーションでモデル化を行い、固有地震が揺らぐ過程を調べると共に、摩擦特性や余効すべりの伝播過程を調べる。

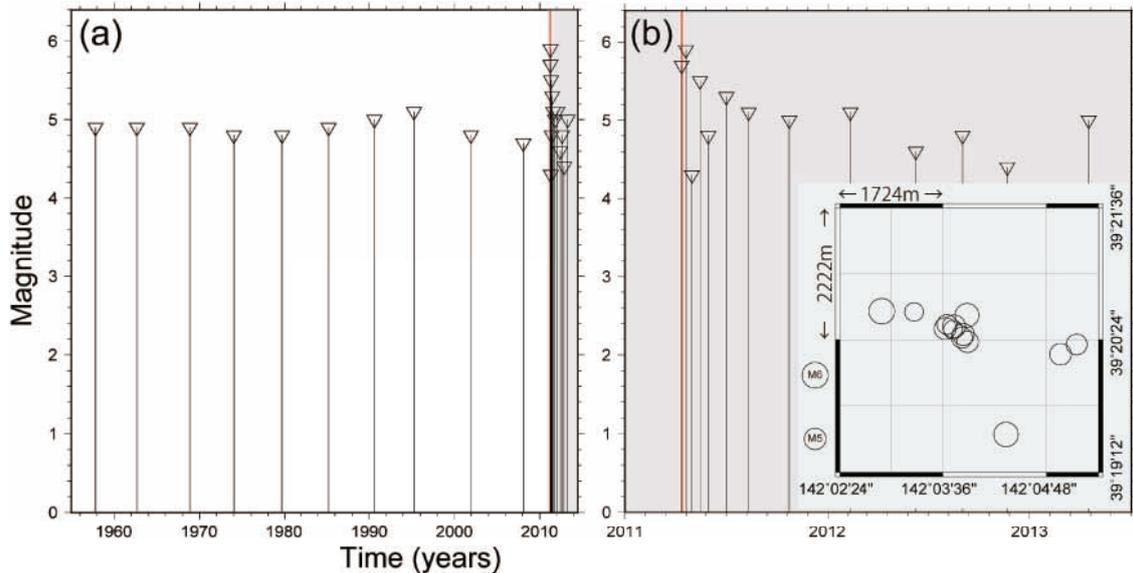


図2. 釜石沖の固有地震の活動履歴. 赤線は東北地震の発生時刻を示し、灰色の時間帯については、(b)で拡大し、震央分布を併記した。

2. 固有地震のモデル化

モデル化に当たって、東北地震のような M9 クラスの超巨大地震と釜石沖の M5 クラスの中規模地震を両立させるのは、地震のエネルギー比でおよそ 100 万倍も異なるため、時間・空間分解能の観点から必要な計算量は膨大なものとなる。そこで、本研究では、M7 クラスと M5 クラスの地震を共存させるモデルで代用する。両者の位置関係については、図 1 の情報だけでは設定することが難しいので、本研究では、図 3 のように走行方向に複数の小アスペリティを設定することにした。

プレート境界面では、速度・状態依存摩擦構成則 (aging-law と slip-law) [3] によって生じる摩擦力と断層すべりに伴うせん断応力が準動的に釣り合っている条件を課すことで、地震サイクルを再現した [4]。また、擬似的なすべり伝播を避けるため、周期境界条件は与えておらず、計算領域の外側では、プレート収束速度 ($V_{pl} = 9 \text{ cm/year}$) で動いているものとする。

本研究では、東北大学サイバーサイエンスセンターとの共同研究によって開発された、SX-9 向けに最適化された FFT の計算プログラムコードを使用した。これにより、約 5 年間隔で発生する M4 クラスの地震と 50 年以上の時間間隔で発生する M7 クラスの地震が共存する、高い時空間分解能を有する大規模シミュレーションを実施することが可能となった。

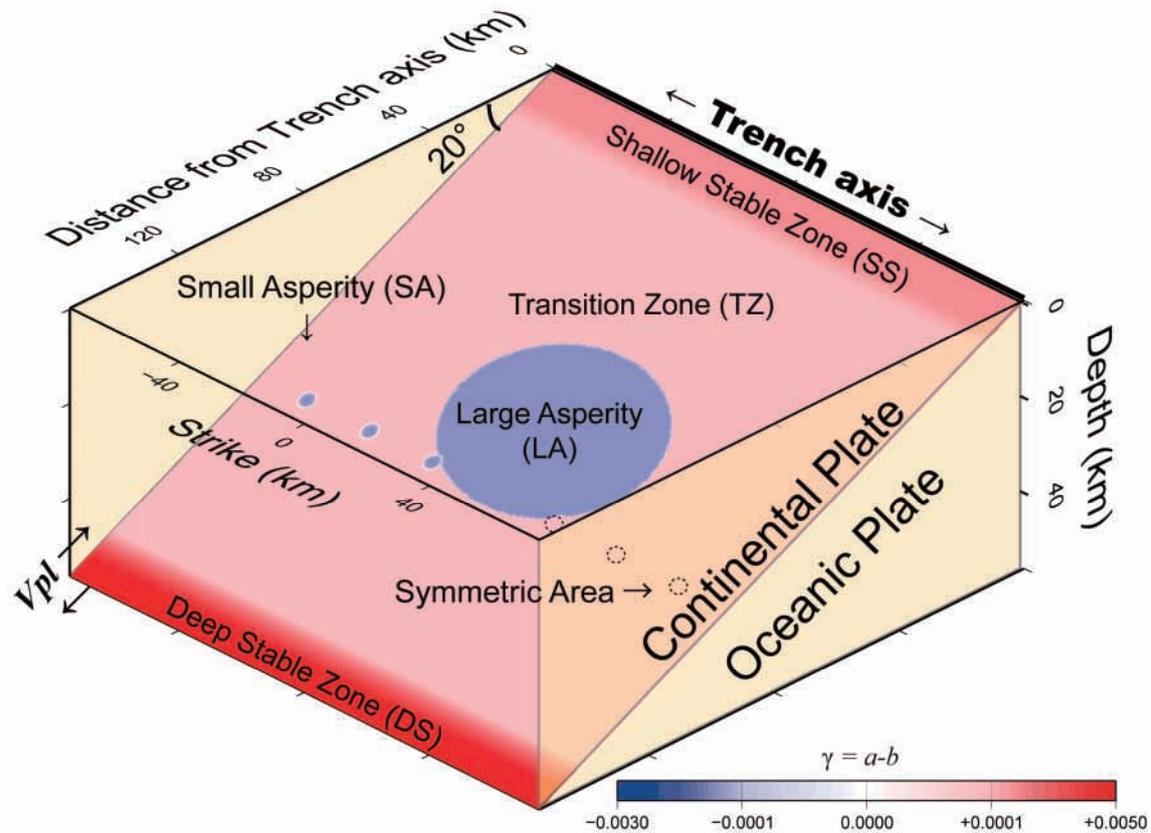


図3. 本研究で想定した沈み込みプレート境界面における摩擦安定性の空間分布図. 暖色系は摩擦安定 (粘性的なすべり特性), 寒色系は摩擦不安定 (固着と地震性すべりの繰り返し) を示す. 大きな円域内 (大アスペリティ) は海溝型巨大地震の震源域に相当し, 小さな円域内 (小アスペリティ) は通常時には固有地震が発生する. 小アスペリティは走行方向に沿って3つ配置したが, 十分な距離があるため, 小アスペリティ同士の相互作用は無視できる.

3. 数値シミュレーションから再現された固有地震の揺らぎの特徴

図3において, 大小のアスペリティが共存ではなく, 単独で存在する場合, 固有的な振る舞いをする事が知られているが, 大小アスペリティの再来間隔 (T_r, t_r) と規模 (M_w, m_w) は, $(T_r, M_w; t_r, m_w) = (62.9, 7.7; 2.5, 3.8)^{\text{aging}}, (82.5, 7.9; 6.3, 5.3)^{\text{slip}}$ となった.

両者が共存する場合の履歴を図4に示す. この場合, 大アスペリティの巨大地震は固有的な振る舞いのままであるが, 小アスペリティは応力擾乱の影響を受けて揺らいでいることが分かる.

また, この結果から, slip-law では, 大規模な余効すべりが通過しても, 固有地震は活発化するどころか静穏化する傾向がみられることも分かる.

一方で, aging-law は, 地震活動が一時的に頻発しており, 規模についても通常と比べて大きいもの(図 4c①, d)だけでなく, 小さいもの(図 4c②③, e) も発生していることが分かる. そして十分な時間が経過すると, 通常の規模に収束する(図 4c⑤, f, g). この特徴は, 図 2b の観測結果を概ね再現できているように見える.

図 4e で示された通常よりも小さな地震については, 図 5 で示された余効すべりの伝播方向と合致する. すなわち, 余効すべりの応力擾乱によって, 小アスペリティの部分破壊が促進され, 伝播方向前面側の②が起きた後に, 残りの応力蓄積の解放として後面側の③が起きたと解釈することができる.

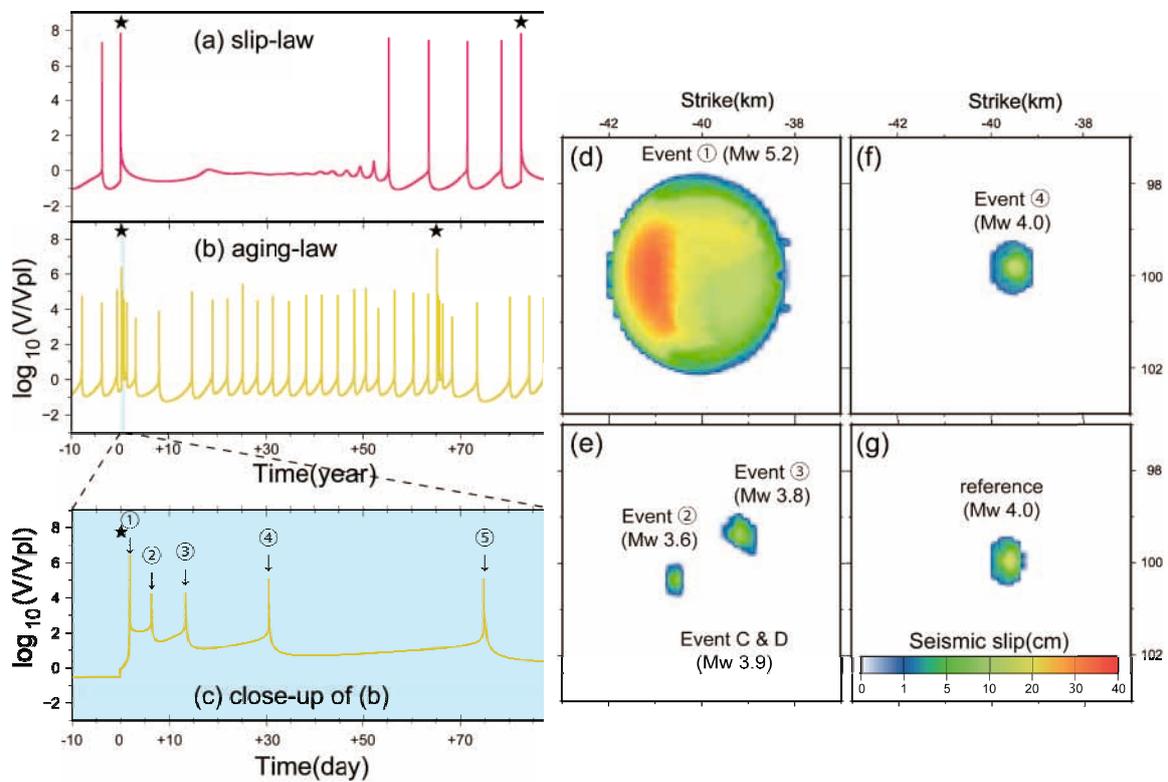


図 4. 小アスペリティが中央に位置する場合の地震速度履歴. ★は大アスペリティで巨大地震が発生した時刻 ($t=0$) を表わす. (b) において, 背景が水色の時間帯は, (c) で拡大し, 各イベントを①~⑤の番号で表記した. (d)~(f) は, ①~④の小アスペリティ域周辺における地震性すべり分布を示す. (g) は, 大アスペリティが固着している期間に小アスペリティで発生する固有地震の一例 (図 5 を参照).

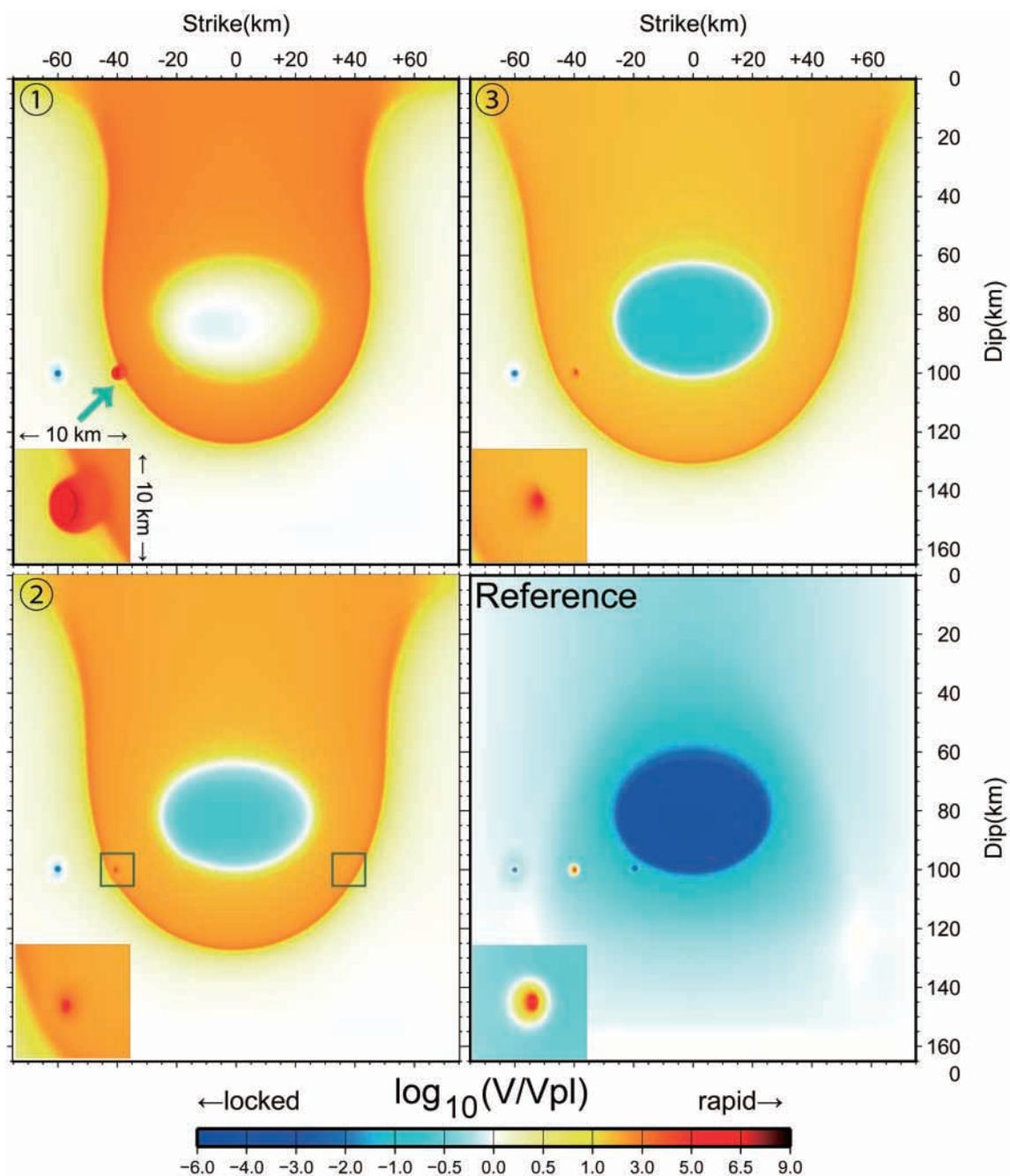


図5. プレート境界面におけるすべり速度のスナップショット. 小アスペリティ周辺の拡大図を左下に示した. 時刻については, 図4cを参照.

4. まとめ: 釜石沖における摩擦特性と余効すべり伝播過程

本研究の数値シミュレーションから示された特徴について, 釜石沖における固有地震の揺らぎに適用することで, 以下のことが結論付けられる.

- (i) 東北地震後に釜石沖で地震活動が一時的に活発化する現象は、摩擦構成則 (aging-law) から説明することが出来る。この場合、小アスペリティが単独であっても再現できるため、通常時の固有地震活動履歴とも整合する。これは、岩石実験とは独立に、摩擦構成則の条件を数値シミュレーションから絞り込むことが出来るアプローチである。
- (ii) 通常よりも小さな規模の地震が起きたのは、余効すべりに伴う応力擾乱によって、小アスペリティ内で部分破壊が生じたためだと考えられる。図 2b より、その震央分布は東南東～西北西方向に揺らいでいることから、釜石沖での東北地震の余効すべり伝播方向は、東南東→西北西であったと推定される。これは、測地観測とは独立した、新しい推定方法であるといえる。

本研究では、M9 クラスと M5 クラスの共存モデルを構築することは出来なかったが、今後さらに、スーパーコンピューターの性能向上や計算アルゴリズムの改良を重ねることで、より現実的な数値シミュレーションを実行することが、今後の課題となろう。

参考文献

- [1] Heki, K. and Y. Mitsui (2013), Accelerated pacific plate subduction following interplate thrust earthquakes at the Japan trench, *Earth Planet. Sci. Lett.*, **363**, 44-49.
- [2] Ye, L., T. Lay, and H. Kanamori (2012), The Sanriku-Oki low-seismicity region on the northern margin of the great 2011 Tohoku-Oki earthquake rupture, *J. Geophys. Res.*, **117**, B02305.
- [3] Ampuero, J.-P., and A. M. Rubin (2008), Earthquake nucleation on rate and state faults – Aging and slip laws, *J. Geophys. Res.*, **113**, B01302.
- [4] 有吉 慶介・松澤 暢・日野 亮太・長谷川 昭・堀 高峰・中田 令子・金田 義行, 2014. 超低周波地震の活動に基づくプレート間固着の推定, *SENAC* **47(1)**, 29-34.

[大学 ICT 推進協議会 2014 年度 年次大会論文集より転載]

スーパーコンピュータシステム SX-ACE の紹介

†山下毅 †森谷友映 †佐々木大輔 †齋藤敦子
 †小野敏 †大泉健治 †岡部公起 †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

yamacta@cc.tohoku.ac.jp

概要：東北大学サイバーサイエンスセンターは、全国共同利用設備として大規模科学計算システムの整備と、HPCI の資源提供機関としての役割を担っている。本稿では、2015 年初頭に運用を開始する本センターの主力計算機である新ベクトル型スーパーコンピュータ SX-ACE と、その運用方針およびユーザの利用環境について紹介する。

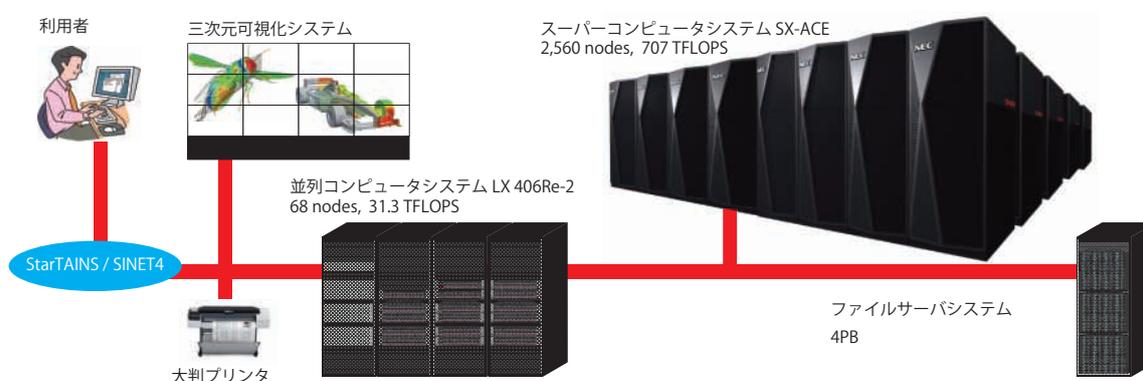


図1 大規模科学計算システムの構成 (2015 年以降)

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）では、2015 年初頭に新スーパーコンピュータシステム SX-ACE（日本電気株式会社製）の運用開始を目指し、本年 11 月に竣工の HPC 新棟（仮称）内において、現在ハードウェアおよびソフトウェアの環境構築作業を行っている。今回導入した SX-ACE のシステムは全 2,560 ノードで構成され、1 ノードあたり理論最大演算性能 276GFLOPS の世界初のマルチコア（4 コア）ベクトルプロセッサを 1 基搭載し、システム全体では約 707TFLOPS となる。主記憶は 1 ノードあたり 64GB を搭載し、256GB/s という高いメモリバンド幅でプロセッサと接続されることで、高い演

算性能とメモリ性能の最適化を実現している。

本稿では、新スーパーコンピュータシステムのハードウェアおよびソフトウェアの特徴と、新システムの導入に伴い今回新設された HPC 新棟の概要、および利用者の利便性とセキュリティの向上を考慮して構築を行った、大規模科学計算システムの運用方針について紹介する。

2 大規模科学計算システム

2.1 システムの概要

本センターの大規模科学計算システムの構成を図 1 に示す。本センターの大規模科学計算システムは、ベクトル型スーパーコンピュータを主力計算機とし、汎用アプリケーションの実行環境とし

表1 SX-9 と SX-ACE の性能比較

性能		SX-9	SX-ACE	向上比
CPU あたり	コア数	1 個	4 個	4 倍
	理論最大演算性能	118.4GFLOPS	276GFLOPS	2.3 倍
	最大ベクトル演算性能	102.4GFLOPS	256GFLOPS	2.5 倍
	メモリバンド幅	256GB/sec	256GB/sec	1 倍
	ADB	256KB	1,024KB/コア×4	16 倍
ノードあたり	CPU 数	16 個	1 個	0.06 倍
	理論最大演算性能	1,894GFLOPS	276GFLOPS	0.15 倍
	最大ベクトル演算性能	1,638GFLOPS	256GFLOPS	0.16 倍
	メモリ容量	1TB	64GB	0.06 倍
	メモリバンド幅	4TB/sec	256GB/sec	0.06 倍
	ノード間通信速度	256GB/sec	8GB/sec	0.03 倍
システムあたり	CPU 数	288 個	2,560 個	8.9 倍
	理論最大演算性能	34.1TFLOPS	706.6TFLOPS	20.7 倍
	最大ベクトル演算性能	29.5TFLOPS	655.4TFLOPS	22.8 倍
	メモリ容量	18TB	160TB	8.9 倍
	最大消費電力	590kVA	1,080kVA	1.8 倍
	計算機室床面積	293 平米	430 平米	1.5 倍

てスカラ型の並列コンピュータの運用も行っている。この二種類の計算機の運用により、利用者の幅広いニーズに応えるサービスを提供している。

SX-9 システムは、2008 年 3 月から運用を開始し本年で 7 年目を迎えているが、6 年半の平均利用率は 80% を超え、また今年度上半期の平均利用率は過去最高の 90% と高い利用率となっており、科学技術計算においてベクトル型スーパーコンピュータのニーズの高さを伺うことが出来る。(利用率：ユーザプログラムの CPU 時間合計÷システムの稼働時間合計×100【%】)

並列コンピュータシステムの LX 406Re-2 と、合計 4PB の容量を有するファイルサーバシステム、および三次元可視化システムは、2014 年 4 月に導入され運用を行っている。これらのシステムは、スーパーコンピュータシステムとの連携、および分散ファイルシステムを活用した計測データの高速な I/O により、高速かつ高精度な防災・減災シミュレーションを行い、シミュレーション結果を三次元可視化するシステムとして活用される。これらのシステムを利用した、ものづくり分野に

おける萌芽的研究、産業利用の促進も期待される。

以下では、2015 年初頭に運用を開始するスーパーコンピュータシステム SX-ACE について説明する。

2.2 SX-ACE システムの紹介

2.2.1 ハードウェアの特徴

■システム構成 SX-9 システムと SX-ACE システムの性能比較を表 1 に示す。SX-9 システム全体 18 ノードでの理論最大演算性能が 34.1TFLOPS であるのに対し、SX-ACE システム全体の 2,560 ノードでは 706.6TFLOPS となり、約 21 倍の性能向上となる。本センターの SX-ACE システムでは最大 512 ノード並列の実行環境に加え、来年度からは最大 1,024 ノード並列の大規模な実行環境が利用者に提供される。

■消費電力・設置面積 SX-ACE は SX-9 と比較して、同一性能時の LSI 数を約 1/100 へと削減したため、SX-9 と同等の性能を 1/10 の消費電力と 1/5 の設置面積で実現している。

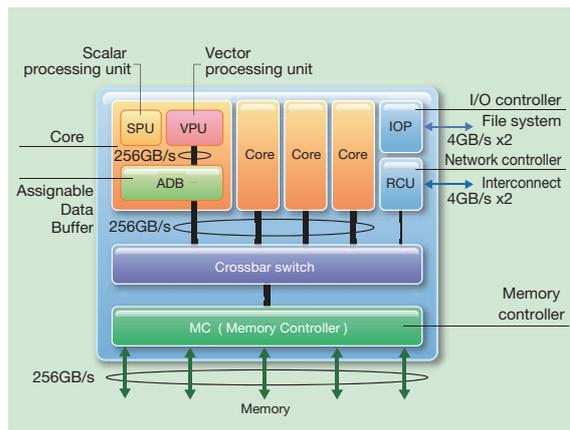


図2 SX-ACE マルチコアプロセッサ

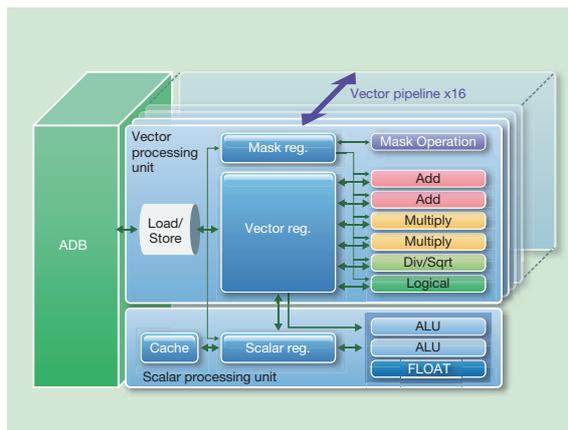


図3 ベクトルプロセッシングユニット

SX-ACE システムは SX-9 システムと比較して約 21 倍の性能向上であるが、最大消費電力および、計算機室の床面積（付帯装置および冷却装置を含む）はそれぞれ、1.8 倍および 1.5 倍として設計した。SX-ACE システムの導入によって最大理論演算性能の飛躍的向上と、省エネおよび省スペースの両立を実現することが可能となった。

■ベクトルプロセッサ SX-ACE のマルチコアプロセッサの模式図を図 2 に、ベクトルプロセッシングユニットの模式図を図 3 に示す。

SX-ACE の CPU はこれまでのベクトルプロセッサと同様の高いベクトル性能と、高いメモリバンド幅を継承し、またシリーズ初のマルチコア化を行った。1CPU は、それぞれ 64GFLOPS のベクトル演算性能を持つ計 4 コアで構成され、256GFLOPS のベクトル演算性能を持つ。

CPU 内のコア間は 256GB/sec のクロスバスイッチにより高速接続され、クロスバスイッチと主記憶間はメモリコントローラを介し 256GB/sec で高速接続されており、CPU あたり 1Byte/FLOP を達成している。

各コアには、容量が 1,024KB に拡張された HPC 専用設計キャッシュである ADB (Assignnable Data Buffer) が搭載され、ADB とコア間のメモリバンド幅は 256GB/sec を有する。データが ADB 経由でアクセスされる場合は、4Bytes/FLOP という高いバンド幅によるデータ供給性能により、

メモリ負荷の高いアプリケーションも高い実行効率での実行が可能となる。また、ノード内並列数を 1 として、単一のコアのみを演算に使用する場合も 256GB/sec のメモリバンド幅が利用でき、このときも 4Bytes/FLOP での実行が可能である。

■ノード間通信 ノード間の通信は最大 4GB/sec × 2（双方向）で接続された 2 段ファットツリーネットワークで構成される。SX-ACE は 1 ノードあたり 256GFLOPS という高いベクトル演算性能により、プログラムの実行に必要な並列度を低く抑えることが可能であり、多並列での実行時に並列性能の高いスケーラビリティが期待出来る。

以上のような SX-ACE のハードウェアの特徴により、これまでベクトル向けに開発されてきたアプリケーションは勿論のこと、学術研究者の幅広い分野のアプリケーションを高い実行効率で実行されることが期待される。

2.2.2 ソフトウェアの特徴

■オペレーティングシステム オペレーティングシステムは前システムから引き続き、POSIX 準拠の SUPER-UX を採用している。OS レベルでマルチノードをサポートし、大規模マルチノードシステムにおいても安定した利用環境を提供して

表 2 SX-ACE で利用可能なプログラミング言語と数値演算ライブラリ

言語・ライブラリ	コンパイラ名・ライブラリ名	準拠規格・機能
Fortran 90/95	FORTTRAN90/SX	ISO/IEC 1539-1:1997 準拠 自動ベクトル化、自動並列化、OpenMP 対応
Fortran 2003	NEC Fortran 2003 コンパイラ	ISO/IEC 1539-1:2004 準拠 自動ベクトル化、自動並列化、OpenMP 対応
C,C++	C++/SX	ISO/IEC 9899:1999 C 準拠 ISO/IEC 14882:2003 C++ 準拠 自動ベクトル化、自動並列化、OpenMP 対応
MPI ライブラリ	MPI/SX	MPI-3.0 準拠
科学技術計算ライブラリ	ASL ASLSTAT MathKeisan	数値計算ライブラリ 統計計算ライブラリ 数学ライブラリ集 (BLAS, LAPACK, ScaLAPACK を含む)

いる。

■言語とライブラリ SX-ACE で利用可能な言語およびライブラリを表 2 に示す。SX-ACE では新たに Fortran 2003 に対応したコンパイラを導入し、幅広い Fortran コードの実行が可能である。また、MPI-3.0 に準拠した MPI ライブラリは Fortran、C/C++ から利用可能であり、同一ノード内では共有メモリの特徴を活かした自動並列/OpenMP 並列の利用、または Flat MPI 実行による大規模並列実行が可能である。

また、SX シリーズに最適化された科学技術計算ライブラリとして、表 2 に示すライブラリが引き続き利用出来る。ライブラリを利用して SX-9 向けに作成したプログラムも、SX-ACE 用にコンパイルし直すだけで実行が可能である。BLAS、LAPACK、ScaLAPACK ライブラリを利用して記述したプログラムは、MathKeisan ライブラリをリンクすることで、ソースコードの変更なく実行が可能である。

■高速化支援ツール SX-9 で高速化支援ツールとして活用されてきた、プログラム実行解析情報 (PROGINF) と簡易性能解析情報 (FTRACE) は引き続き利用可能であり、加えて GUI でプログラムの性能を解析可能な NEC Ftrace Viewer を導

入する。これは FTRACE 機能で採取された性能解析情報をグラフィカルに表示し、ベクトル性能や OpenMP、MPI を利用した並列プログラムのスレッド・MPI プロセス毎の実行時間、MPI プロセス間の通信時間をグラフ表示することにより、性能のボトルネックやロードインバランスを把握するためのツールである。これらのツールを用いることで、実行コストの高いサブルーチンの特定や、並列実行時の演算量の均一化、ノード間通信の最適化等の高速化作業が容易になる。

■分散・並列ファイルシステム HPC システムの大規模化やデータの大容量化に対応するために、分散・並列ファイルシステムである、NEC Scalable Technology File System (ScaTeFS) を採用した。SX-ACE および並列コンピュータの各ノードとストレージシステムを ScaTeFS で接続し、データおよびメタデータを複数の I/O サーバに分散配置することで負荷分散とスケールアウトを実現し、システム全体のスループットの向上が見込まれる。

■ジョブ管理システム ジョブスケジューリング機能は前システムと同様 NQSII を採用し、計算リソースを管理して効率的なユーザのジョブ管理を行う。

NQSII によるバッチジョブ投入方法の概略を図

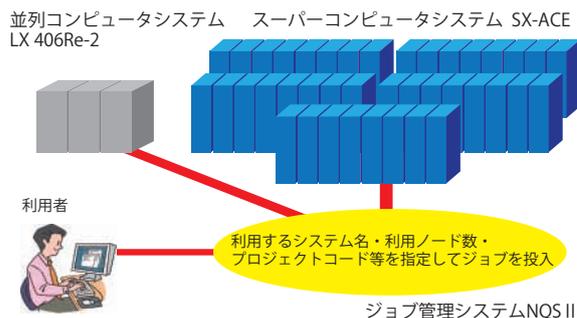


図4 NQSIIによるバッチジョブ投入方法

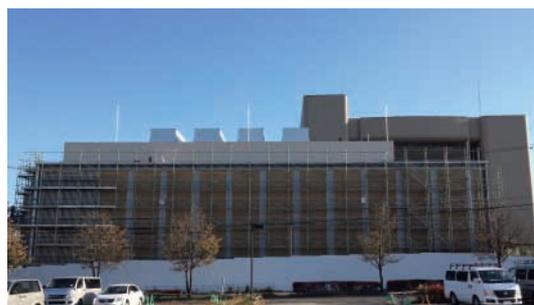


図5 HPC 新棟外観 (2014.10.31 撮影)

4に示す。利用者はジョブ投入の際に、利用するコンピュータシステム、利用ノード数、プロジェクトコード名、ジョブスクリプトファイル名、ノードあたりの実行プロセス数、ノード内並列数等を指定することにより、同一のフロントエンドサーバからSX-ACE、LX 406Re-2の両システムに対してジョブの投入が可能である。

この機能により、以下のように利用者のアプリケーションの特性に合わせて、柔軟にジョブの投入を行うことが可能となる。

- ・ロードストアを多用する、メモリ負荷の高いアプリケーションはSX-ACEを利用し、ノード内は単一コアのみを使用して高いメモリ転送速度を活用。
- ・並列実行性能が高いアプリケーションはSX-ACEを利用し、コンパイラの自動並列化機能を用いてノード内は4スレッド並列で実行。
- ・高いスカラ演算性能が要求されるアプリケーションはLX 406-Re2を利用し、単一コアで実行することでターボブースト機能を活用。

また、スケジューラマップの導入により、ジョブの実行予定がない実行ノードを検出した場合に自動的にCPUのコア縮退運転、またはノードが省電力運転へ移行することが可能であり、高いQoSを保ったまま消費電力を削減することが出来る。この機能により、計画停電などの運用予定に合わせて指定された期日に運用ノード数を調整することも可能である。

実行時に稼働していたノードに障害が発生した

場合は、他に空きノードがある場合はそれらを割り当てることにより、利用者に対して迅速に計算機環境を提供することが可能である。

3 HPC 新棟について

3.1 新棟の概要

SX-ACEシステムの導入に先立ち、本センターの正面に隣接してHPC新棟が竣工した。HPC新棟の外観写真を図5に示す。

SX-ACEシステムで必要とされる電源設備、空調機等を含めた計算機の設置スペース、およびSX-ACEの水冷方式に対応した冷却システム（水冷装置設備本体と配管設備一式）を、現在SX-9を運用している本センターの計算機室に確保することは困難であったため、新棟の建設が行われた。

今後のHPCシステムの規模拡大も見据え、継続的に情報処理基盤拠点としての役割を担うことが出来る計算機棟としての理念の下、新棟の設計が行われた。

3.2 空調・水冷設備

SX-ACEは水冷・空冷の両冷却方式で運用されるため、水冷方式に対応するための配管設備と、空冷方式に対応するための空調設備の設置が必要となる。

冷却水と空調に必要な冷水は、屋上に設置した密閉形フリークーリング方式の冷却塔（チラー）で生成される。周囲温度の低下時（中間期・冬期）

に、冷却塔のみで冷水を生成させ直接負荷側に送水することで、エネルギーの使用量を低減させることが出来る。

3.3 アイルキャッピング

SX-ACE および付帯装置が収納されるラックは、冷気の吸入側である前面側を向かい合わせてレイアウトされる。フリーアクセスフロアの底面から供給される冷気をラック前面から効率的に吸入するために、向かい合うラックの側面と上面をビニルカーテンにより仕切る、コールドアイルキャッピングの方式を採用したことで省エネ効果が期待される。

また、計算機ラック背面からの排気は天井面の吸入口から天井裏を介し、空調機に還気する方式を採用している。

4 システムの運用方針

4.1 利用者環境について

4.1.1 ログイン認証方式

■パスワード認証方式の廃止 本センターの SX-9 システムではフロントエンドサーバへのログインの際に、パスワード認証方式および公開鍵暗号方式の両方が利用可能であり、どちらの方式でログインを行うかは利用者の判断に任されている。SX-ACE システムでは、昨今の漏洩パスワードによる不正アクセスのセキュリティインシデントへの対策として、フロントエンドサーバへログインする際のパスワード認証方式を廃止し、公開鍵暗号による認証方式のみ利用可能とした。

■鍵ペアの生成 ログインに必要な公開鍵・秘密鍵ペアの生成においては、パスフレーズを設定しない、あるいは強度の低いパスフレーズにより秘密鍵が作成されることを防ぐために、フロントエンドサーバへの初回接続時には、新たに設置した SSH アクセス認証鍵生成サーバ（以下、鍵サーバ）で鍵ペアを生成する方式を採用した。

利用者は発行された利用者番号と初期パスワード

により鍵サーバへログインし、専用の鍵ペア生成プログラムにより、一定強度のパスフレーズを持つ秘密鍵を作成する。画面に表示された秘密鍵のテキストをローカル PC にコピー&ペーストにより保存し、このファイルを秘密鍵として利用する。この秘密鍵を利用して、フロントエンドサーバに公開鍵暗号方式でのログインが可能となる。なお、鍵サーバ上で鍵ペアを作成すると、鍵サーバへのログインはロックされる。

■HPCI 課題利用者のログイン方法 HPCI 課題利用者は、HPCI が提供する GSI 認証による電子証明書を用いたシングルサインオンでのログインのみが利用可能である。

4.1.2 プロジェクトコード

NQSII のジョブアカウント機能によって、ユーザが異なるプロジェクトで計算機資源を利用する際に、ジョブ単位の課金と予算管理を行うことが出来る。SX-ACE システムではこの機能を利用し、1つの利用者番号で複数の請求先の使い分けを可能とするために、利用者管理の1つとしてプロジェクトコードを導入した。プロジェクトコードの導入前と導入後の利用者と請求先の関係を図 6 に示す。プロジェクトコードの導入により、利用者には以下の様な利便性と、セキュリティ性の向上が期待出来る。

■複数の請求先の利用 従来の研究室予算での利用に加え、課題採択形式で利用されるケースが近年増加している。SX-9 システムで複数の請求先を利用する場合の模式図を図 6 左に、SX-ACE システムでの場合を図 6 右に示す。

SX-9 システムまでは、複数の請求先を利用する場合、請求先毎に支払責任者番号を発行する必要があった。そのため、利用者 A が複数の請求先を使い分けるには、請求先ごとの支払責任者番号 (u2000, u2100) のそれぞれに利用者番号 (c2200, d2300) を取得し、ログインする利用者番号を使い分ける必要があった。

SX-ACE システムからはプロジェクトコードの導入により、利用者 A は 1 つの利用者番号

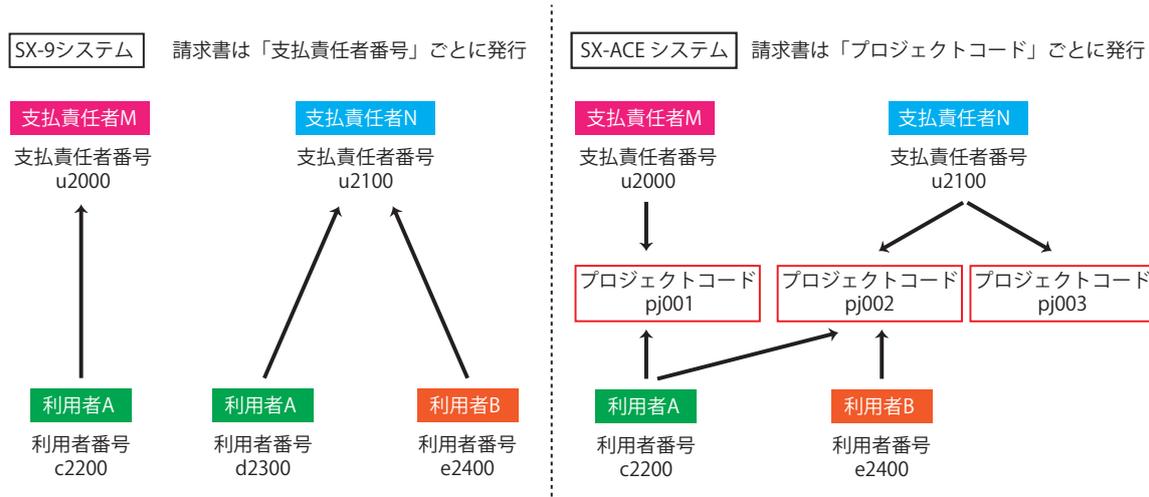


図 6 プロジェクトコードの導入による利用者と請求先の関係

(c2200) から請求先の異なる複数のプロジェクトコード (pj001, pj002) を使い分けることが可能になる。バッチジョブ投入の際に NQSII のジョブアカウント機能を用い、請求先としてプロジェクトコードを指定することで、利用者が複数の請求先を使い分けることが可能となる。また、支払責任者が複数のプロジェクトコード (pj002, pj003) を保有することも出来る。

■利用者環境 SX-9 システムでは請求先毎に利用者番号を使い分ける必要があったため、利用者がそれぞれの利用者番号に割り当てられたユーザのデータ領域内で環境の構築が必要であった。また、利用者番号・パスワードの管理も利用者番号毎に必要であった。SX-ACE システムでは、利用者は 1 つの利用者番号で、複数の請求先を利用することが可能であるので環境の構築は 1 箇所で済み、また鍵ペアの管理も簡便になるためセキュリティ性も向上する。

■課題利用期間とプロジェクトコード 採択課題の利用期間が終了したものについては、該当するプロジェクトコードを無効にすることで、利用者はジョブを投入することが不可能となる。また、利用者が利用可能な課題が追加された場合は、利用者番号に対してプロジェクトコードを追加設定することでジョブの投入が可能となるため、それ

まで利用していた環境を引き続き利用することが可能である。

4.2 利用ノード数と利用負担金について

大学・学術利用における、SX-ACE システムの利用ノード数と利用負担金を表 3 に示す。なお、民間企業利用に関しては、大学・学術利用単価の 3 倍の単価設定としている。

4.2.1 共有利用

共有利用は、他のユーザと利用するノードを共有する方式である。本センターの運用方針である大規模ジョブの長時間実行環境を提供する目的で、ジョブの実行時間はスケジューラマップ時間以内で無制限としている。待ち行列は FIFO を基本とするが、利用者がジョブの実行時間をジョブスクリプトファイルに明示することで、リソースに空きがある場合はジョブのエスカレーションが自動的に行われる。

研究室の PC、あるいは共有サーバ等で実行されているプログラムの SX-ACE への移行を支援する目的で、経過時間制限を設定した 1 ノード利用を無料としている。また、ジョブの並列化を促進させる目的で、1 ノードから 32 ノードの利用は利用負担金単価は一定とし、33 ノード以上の利用では利用するノード数が多くなると、利用負担金

表3 SX-ACE システムの利用ノード数と利用負担金 (大学・学術利用)

【共有利用】

利用ノード数	経過時間制限※	最大メモリサイズ	利用負担金単価【円/秒】
1	あり	64GB	無料
1~32	なし	2TB	0.06
33~256	なし	16TB	(利用ノード数-32) × 0.002+0.06
257~1,024	なし	64TB	(利用ノード数-256) × 0.0016+0.508

※ジョブの実行時間は、スケジューラマップ時間以内とする。

【占有利用】

利用ノード数	最大メモリサイズ	利用期間	利用負担金【円】
32	2TB	3ヶ月間	400,000
		6ヶ月間	720,000
64	4TB	3ヶ月間	720,000
		6ヶ月間	1,300,000
128	8TB	3ヶ月間	1,300,000
		6ヶ月間	2,340,000

単価の増加率が減少する2段階の単価設定として 慮した計画的な実行が可能である。
いる。

また、SX-ACE では NQSII の機能として会話リクエスト機能が追加され、クライアント環境から SX-ACE のノードに直接ログインすることなく対話型操作が可能である。会話リクエストの場合、1ノードで実行された際の課金体系(0.06円/秒)が適用される。

4.2.2 占有利用

占有利用は一定数のノードを利用者、あるいは利用者グループが占有して利用する方式である。この場合、利用者は占有利用として設定されたプロジェクトコードを指定してジョブの投入を行う。共有利用で使用されるスケジューラマップにはアサインされないため、契約されたノード数まではジョブが優先的に実行される。ジョブが利用可能なノード数を超えた場合は、実行中のジョブが終了するまでジョブは実行されないが、研究グループ内でジョブ実行までの待ち時間や、利用額を考

5 おわりに

本稿では 2015 年初頭に運用を開始する、サイバーサイエンスセンターのスーパーコンピュータシステム SX-ACE について紹介した。また、SX-ACE システムの導入に伴うシステムの運用方針の変更点について説明した。最新のベクトル型スーパーコンピュータ SX-ACE を、皆様の研究にご活用いただけたら幸いである。

謝辞

可視化画像をご提供いただきました、宇宙航空研究開発機構の中橋和博先生、スーパーコンピュータシステム SX-ACE の導入および環境構築にあたり、日本電気株式会社、NEC ソリューションイノベーション株式会社、NEC フィールドディング株式会社の皆様には多大なるご協力をいただきました。皆様に深く感謝の意を表します。

[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。

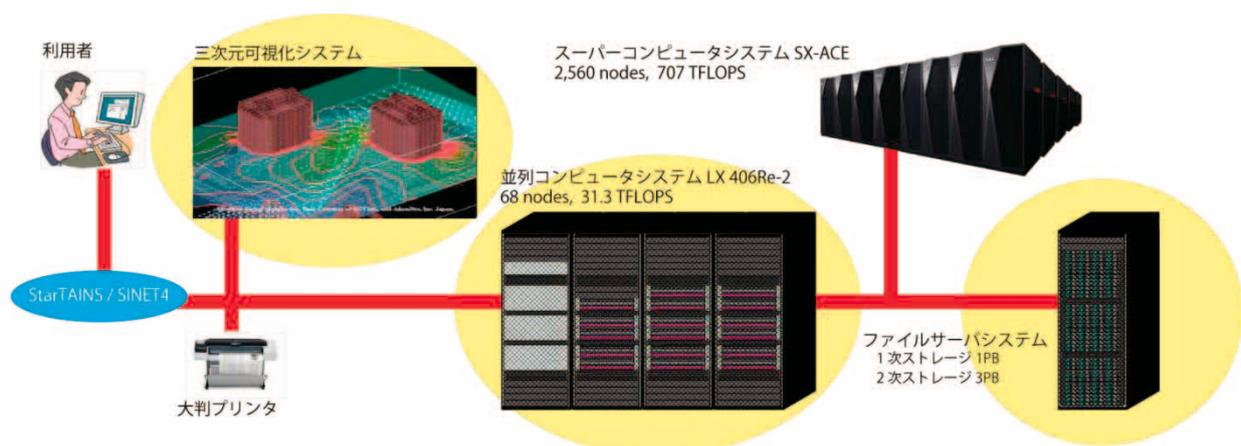


図1 大規模科学計算システムの構成

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ため、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ド並列処理も可能である。

ジョブ管理には NQS II (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。

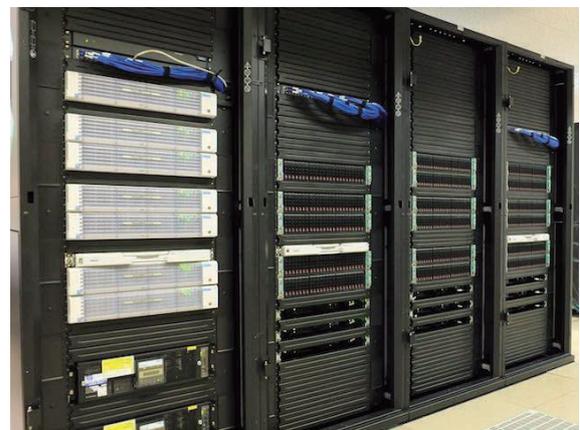


図 2 LX 406Re-2

表 1 LX 406Re-2 の諸元

システム全体		ノード性能	
総ノード数	68 ノード	理論演算性能	460.8GFLOPS (倍精度)
総理論演算性能	31.3TFLOPS (倍精度)	CPU	Intel Xeon プロセッサ E5-2695v2 (12core/2.4GHz) × 2
総メモリ容量	8.5TB	メモリ	128GB
ノード間接続	InfiniBand (4×FDR, 56Gbps)		

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	NEC NumericFactory
ライブラリ	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。

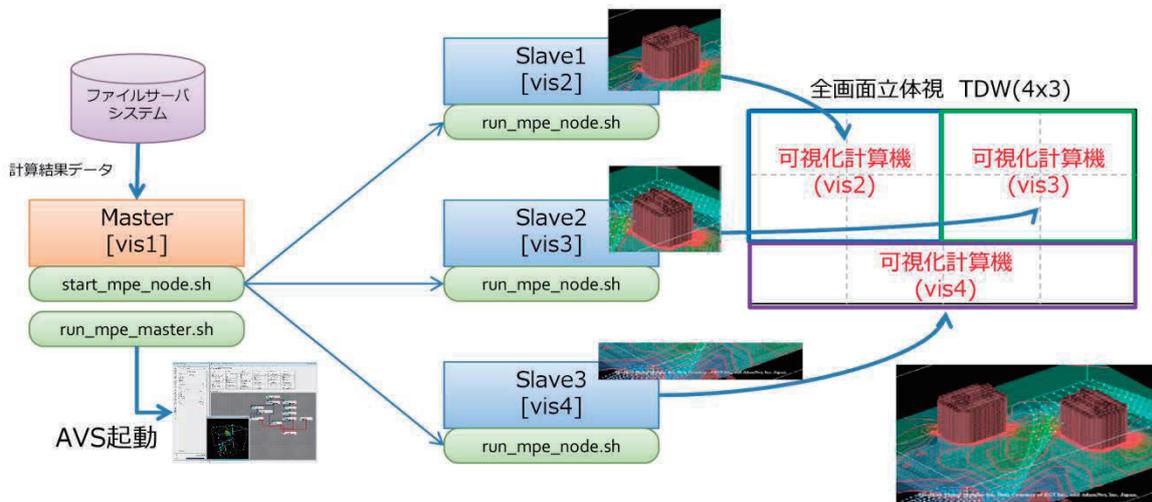


図3 三次元可視化の仕組み

● 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、入出カ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。

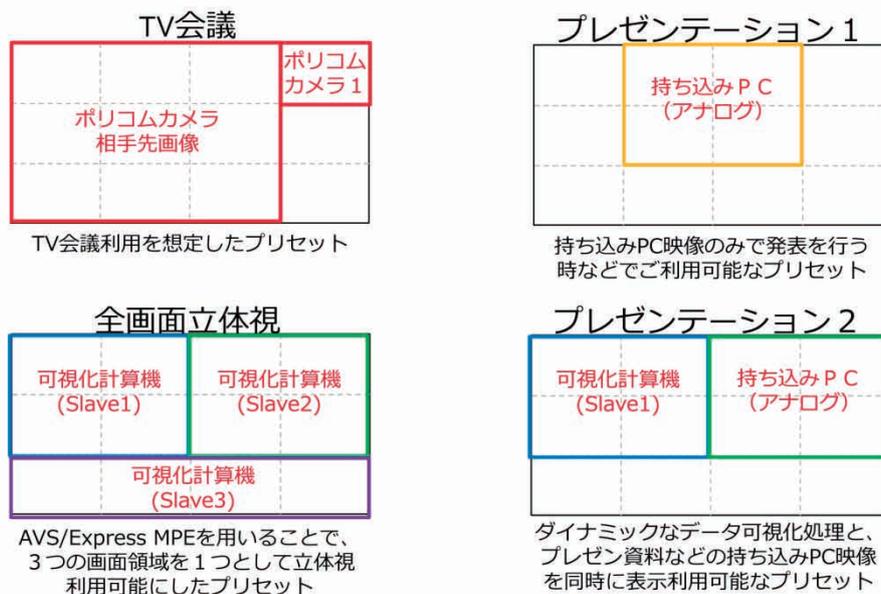


図 5 ディスプレイ表示パターンの例

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で、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。

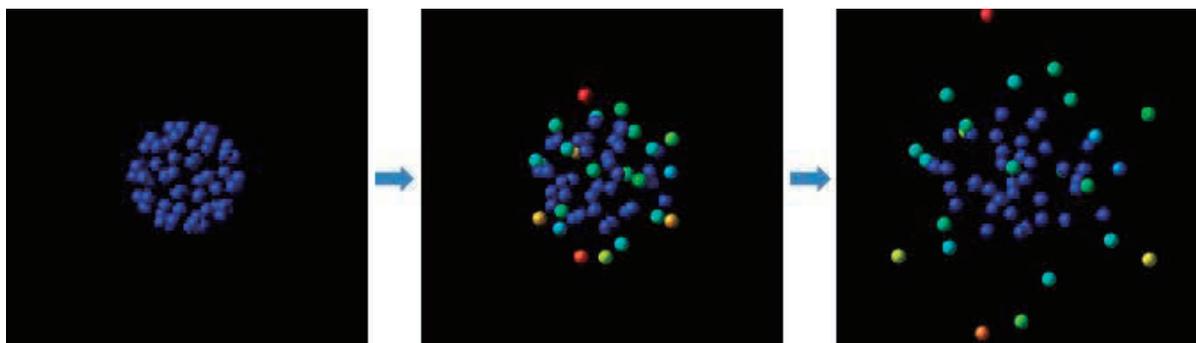


図6 フラーレン爆発シミュレーションの可視化



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の導入にあたっては、ユーザの実行環境のスムーズな移行が可能となる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 $C_{60} q + (q = 20-60)$ 」, The Journal of Chemical Physics 141, 121105 (2014)

[大学 ICT 推進協議会 2014 年度 年次大会論文集より転載]

リアルタイム津波浸水・被害予測シミュレーション システム開発の取り組み

大泉健治†, 阿部孝志*, 佐藤佳彦**, 松岡浩司*, 撫佐昭裕‡*, 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

* 日本電気株式会社, ** NEC ソリューションイノベーション株式会社

oizumi@cc.tohoku.ac.jp

概要: 東北大学サイバーサイエンスセンターは、本学災害科学国際研究所他と共同で総務省 G 空間シティ構築事業のプロジェクトに参加し、リアルタイム津波浸水・被害予測情報配信の実証に取り組んでいる。本稿では、本センターの次期スーパーコンピュータ SX-ACE 上に構築中の、地震発生後 20 分以内の高精度浸水予測および被害予測を目標とする「リアルタイム津波浸水・被害予測シミュレーションシステム」の開発について紹介する。

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、全国共同利用機関として高性能計算やネットワークなど先端学術情報基盤の整備・運用と、これら先端学術情報基盤を活用した新しい科学（サイバーサイエンス）の創造に関する教育・研究を推進している。平成 22 年は学際大規模情報基盤共同利用・共同研究拠点（JHPCN）制度のもとで運用を開始し、平成 24 年は「京」を中核とした全国主要なスーパーコンピュータを連携した革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）による運用が開始され、本センターは計算機資源提供機関として参画している。また、平成 19 年から平成 22 年は先端研究施設共用促進事業のもと、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ている。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。平成 23 年以降も、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、平成 19 年の事業開始から合わせて 7 社の利用があった。

さらに、平成 26 年度は新たな事業として、総務省の G 空間シティ構築事業のプロジェクトの一つ

である「リアルタイム津波浸水・被害予測・災害情報配信による自治体の減災力強化の実証事業」に、本学災害科学国際研究所他と共同で参加し、防災分野における社会インフラ構築の実証に取り組む。本実証では、大規模地震発生時の地震情報から、本センターの次期スーパーコンピュータ SX-ACE を用いて、津波による浸水被害の予測を地震発生から 20 分以内を目標としたリアルタイムで行うことを目指している。

本稿では、リアルタイム津波浸水・被害予測情報配信の実証事業の概要および本センターが担当するリアルタイム津波浸水・被害予測シミュレーションについて紹介する。

2 リアルタイム津波浸水・被害予測シミュレーション

G 空間シティ構築事業は、総務省の平成 25 年度「災害に強い G 空間シティの構築等新成長領域開拓のための実証事業」であり、「G 空間×ICT 推進会議」で提言された実証プロジェクトを実施するものである。この会議では、G 空間（地理情報）情報と ICT の連携により G 空間情報を高度に利活用できる G 空間社会を実現し、経済の再生や防災・減災等、我が国が直面する課題の解決に寄与する方策等を検討している。実証プロジェクトの一つに「世界最先端の G 空間防災モデルの確立に

関する事業」があり、そのモデルの一つが津波防災及び災害情報伝達に関する「リアルタイム津波浸水・被害予測・災害情報配信による自治体の減災力強化の実証事業」（事業実施代表者 災害科学国際研究所 越村俊一教授）である。

2.1 シミュレーションの概要

本センターは、災害科学国際研究所、大学院理学研究科、日本電気株式会社及び国際航業株式会社と共同で前述した津波防災及び災害情報伝達モデルの実証事業（図 1）に参加し、大規模地震発生時の津波による津波浸水・被害予測を担当する。

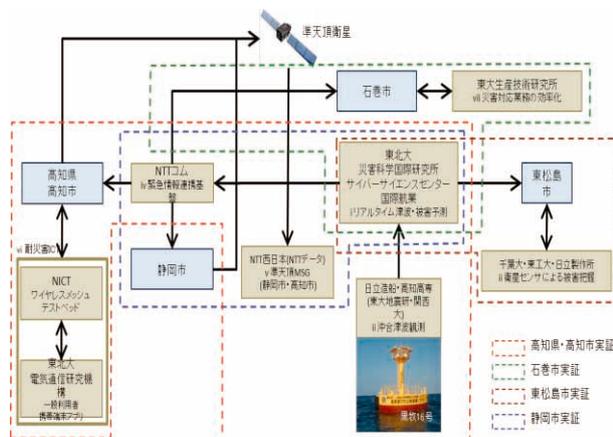


図 1：リアルタイム津波浸水・被害予測・災害情報配信による自治体の減災力強化の実証事業の概略図

大規模地震発生時の津波による浸水予測には、震源位置や地震規模、津波の対象となる地域の地形などの情報の解析が必要であり、被害の予測は、人口や建物などの情報をもとにして推計を行う。津波浸水・被害予測を実際に被害想定地域の防災・減災対策として利用するには、これら大量の情報を高精度かつ即時性をもって予測解析することが必要不可欠である。本実証では、本センター

の次期スーパーコンピュータシステム SX-ACE および運用管理システムが持つ強力な実行性能と利用・運用機能を活用することで、津波による浸水被害の予測までを地震発生から 20 分以内にリアルタイムで行う、という汎用的なサーバシステムでは困難な目標を実現しようとするものである。

2.2 リアルタイム津波浸水・被害予測シミュレーションシステム

地震発生から被害予測情報の配信までの流れが図 2 である。

まず、地震発生をきっかけに、災害科学国際研究所の専用サーバがリアルタイム津波浸水・被害予測シミュレーションに必要な震源や断層等の地震情報の即時推定を行う。つぎに、本センターの SX-ACE が、推定した地震情報や予測対象となる自治体の地形データ等を入力データとして、津波浸水・被害予測シミュレーションプログラムを実行する。プログラムの実行が完了した時点で、対象地点の浸水予測情報と被害予測情報を出力データとして災害科学国際研究所の専用サーバに転送する。最後に、出力データを受けた災害科学国際研究所の専用サーバは、自治体に被害予測情報を配信する。

地震は昼夜を問わず随時発生し得るため、この一連の処理は、人手を介さずにリアルタイムで処理することが必要である。また、本センターのスーパーコンピュータは、通常時は全国共同利用計算機として運用しているため、地震発生時には、津波浸水・被害予測シミュレーションプログラムを直ちに実行開始ができるよう、津波シミュレーション専用運用に自動的かつ迅速に切り替えて対応を行う。スーパーコンピュータを用いてリアルタイムに津波浸水・被害予測を行う実証実験は世界で初めての取り組みであるとともに、全国共同利用計算機の新たな利活用の試みである。

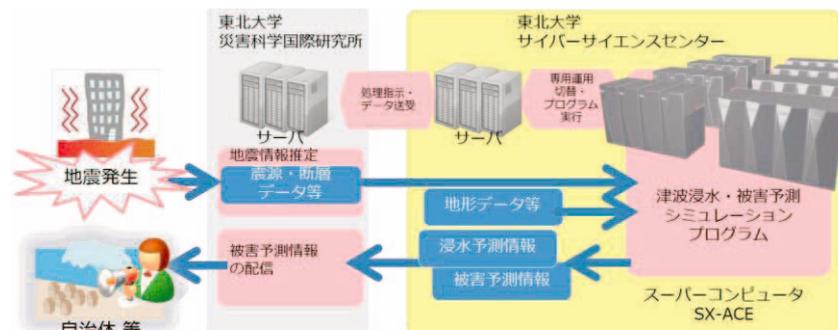


図 2：リアルタイム津波浸水・被害予測シミュレーションシステムの構成

2.3 スーパーコンピュータ SX-ACE

リアルタイム津波浸水・被害予測シミュレーションシステムは、本センターに導入予定の SX-ACE を用いて構築する。本センターのスーパーコンピュータは、従来から高速なベクトル処理能力が必要なプログラムや、高いメモリバンド幅を必要とする流体解析、気象解析及び電磁界解析の分野によく利用されている。最近はベクトル演算や自動並列化機能による高速化に加えて、複数ノードを用いた大規模計算が増加している。このため、増加する計算需要に対応し、さらなるシミュレーションモデルの大規模化と精度の向上を促進するため、現行の SX-9 を SX-ACE (図 3) に更新中で平成 27 年初頭に運用開始の予定である。

システムは、全 2,560 ノードで構成され、総理論最大演算性能は 706.6TFLOPS と大幅に増強される。1 ノードあたりは、64GB の共有メモリと 4 個のベクトルコアを有し、256GB/s の高いメモリバンド幅を持つ (表 1)。



図 3 : スーパーコンピュータ SX-ACE 外観

表 1 : スーパーコンピュータ SX-ACE の性能

CPUあたり	コア数	4 個
	理論最大演算性能	276 GFLOPS
	最大ベクトル演算性能	256 GFLOPS
	メモリバンド幅	256 GB/sec
	ADB (Assignable Data Buffer)	1,024 KB×4
ノードあたり	CPU数	1 個
	理論最大演算性能	276 GFLOPS
	最大ベクトル演算性能	256 GFLOPS
	メモリ容量	64 GB
	メモリバンド幅	256 GB/sec
	ノード間通信速度	16 GB/sec
システムあたり	CPU数	2,560 個
	理論最大演算性能	706.6 TFLOPS
	最大ベクトル演算性能	655.4 TFLOPS
	メモリ容量	160 TB

2.4 津波浸水・被害予測シミュレーションプログラム

リアルタイム津波予測・被害予測シミュレーションの中核となるシミュレーションプログラムは、災害科学国際研究所越村教授が開発したコードを、本センターの SX-ACE に最適化して実装する。本コードは、非線形長波理論にもとづき Staggered leap-frog 法を用いて、津波の伝搬と浸水を再現するものであり、津波の到達時刻、津波の波高、流量などを求めることができる。本実証では、予測対象とする宮城県石巻市・東松島市、高知県高知市及び静岡県静岡市の 4 つの自治体について、それぞれ 10m という詳細な格子サイズでの計算モデルを作成し、地震発生後 6 時間のシミュレーションを行い、津波浸水・被害予測するものである。

本プログラムは、1 つの自治体のシミュレーションにおいて、汎用的なサーバシステムの逐次処理では 80 時間以上の計算時間を必要とするが、本センターの SX-ACE の有するベクトル処理性能および大規模並列性能を用いることで、いずれの自治体も 10 分以内に予測解析が完了することを目標とする。SX-ACE は導入前のため、現行のスーパーコンピュータ SX-9 にてプログラムの最適化を進めている。プログラムの移植、目標実行時間に収めるための高速化と大規模並列化のチューニング等の開発作業を SX-9 上で行い、SX-ACE での実行時間や計算資源量を見積もっている。

プログラムの高速化では、ベクトルプロセッサの性能を引き出すためにベクトル化チューニングを実施している。本プログラムの主な処理は流体計算の「連続の式」と「運動方程式」の差分計算であり、基本的にはコンパイラの自動ベクトル化機能でベクトル化ができるものであった。しかし、一部の DO ループではベクトル化の阻害要因となる入出力処理や最適化の余地がある分岐処理があり、ソースコードの修正でベクトル化と最適化を行っている。現在、1 つの自治体のシミュレーションにおいて、SX-9 1CPU による実行時間はスカラサーバ (Intel Xeon E5-2695v2) 1 コアと比較して 5 倍以上の高速化がされている (図 4)。

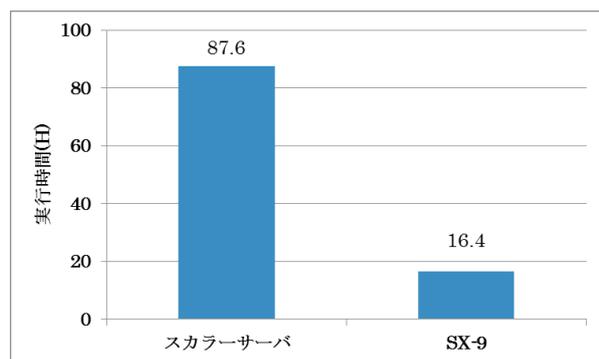


図4：シングルコアでの実行時間

また、プログラムの並列化はMPI ライブラリによる並列化を行っている。本プログラムの計算領域は、図5に示すように5つの異なる格子サイズの領域を計算しており、それぞれの領域における演算量は異なっている。そのため、並列化における演算のロードバランスが崩れないように各領域の分割数を調整できるようにした。また、ベクトル演算性能が低下しないようにベクトル長を維持した領域分割も行った。この並列化により、現在、並列化率 98%以上を達成し、1つの自治体のシミュレーションにおいて、SX-9 64CPU の実行時間は35分に短縮されている（図6）。



図5：シミュレーションの領域

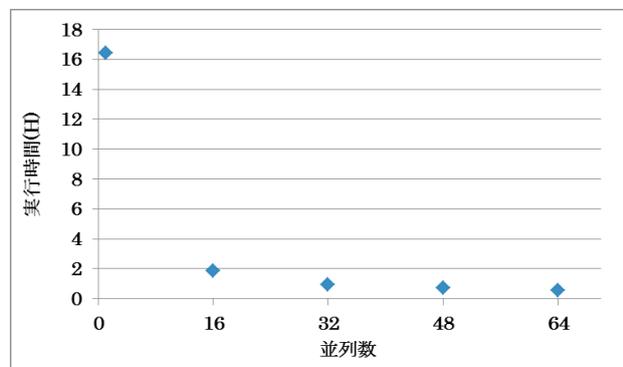


図6：並列数ごとの実行時間

る。現在行っている SX-ACE での性能見積もりでは、4つの自治体のシミュレーションにおいて、SX-ACE 192 ノードで目標の10分を達成することができる見込みである。

3 むすび

リアルタイム津波浸水・被害予測シミュレーションを概説するとともに、本センターが担当するシミュレーションコードの開発状況、そのコードを実装するスーパーコンピュータ SX-ACE のシステム構成及び運用方法について紹介した。本センターは、学術研究や産業分野への情報基盤の提供に加え、防災分野における社会インフラとしての新たな役割を、本事業での環境構築や評価を通し可能性を示していきたい。

謝辞

本システムの環境構築にあたり、災害科学国際研究所、大学院理学研究科、日本電気株式会社及び国際航業株式会社の皆様には多大なるご協力をいただいております。皆様に深く感謝の意を表します。

参考文献

- [1] G 空間シティ構築事業：
http://www.soumu.go.jp/main_sosiki/joho_tsusin/top/local_support/02ryutsu06_03000054.html

今後は SX-ACE における大規模並列実行を念頭に置き、MPI 通信のさらなる最適化を予定してい

[大学 ICT 推進協議会 2014 年度 年次大会論文集より転載]

キャンパス無線 eduroam の最新動向と国内機関向け新サービス

後藤英昭¹, 新妻 共², 中村素典³, 曾根秀昭¹¹ 東北大学サイバーサイエンスセンター² 東北大学大学院 情報科学研究科³ 国立情報学研究所

hgot@cc.tohoku.ac.jp

概要： 学術系の無線 LAN ローミング基盤である eduroam は、世界 70 か国、国内 86 機関 (2014 年 10 月現在) に成長し、大学キャンパス以外にも様々な施設でサービス展開が進んでいる。本報告では、eduroam の国内外の最新状況を概説するとともに、国内の参加機関および学術団体のための新しい認証サービスとして、(1) 国内で開催される国際会議・国内会議等のための期間限定 eduroam アカウント発行、(2) 所属機関発行の電子メールアカウントを間接的な認証に利用するオンラインサインアップシステム、(3) EAP-TLS 認証を実現するクライアント証明書発行システムの 3 種類を紹介する。

1 はじめに

国際的な学術無線 LAN ローミング基盤である eduroam (エデュローム) は、世界中の大学や研究所等において、キャンパス無線 LAN の相互利用を実現する。世界ではこの一年ほどで南米やアジア、アフリカで展開が進んでいる。[1] で既報のように、病院や空港・鉄道駅、博物館などの公共施設におけるサービスも各地で見られるようになり、eduroam サービスの厚みが増してきている。2014 年 10 月時点での国内の参加機関数は 86 であり、国内の認証基盤として、学術認証フェデレーション (学認, GakuNin)[2] と並んで近年成長が著しい。

日本国内には 1,200 を超える高等教育機関があることから、eduroam 導入の障壁を緩和し、また、利用者・管理者双方の利便性を改善することを目的として、eduroam JP では「代理認証システム」や「仮名アカウント発行システム」、SINET を利用したゲストネットワークなどを開発・提供してきた。一方、国内で eduroam サービスが利用できる大学施設や会議場が増加したのに伴い、会場の無線 LAN サービスを国際会議や学会、研究会等の参加者に提供するというニーズも増えてきた。本稿では、この需要に応えるための「会議向け期間限定 eduroam アカウントの試行」を紹介する。

eduroam JP では、eduroam インフラの利便性と耐障害性の向上のための技術開発・実証実験も継続している。本稿では、代理認証システムの拡張機能として新規開発した、「eduroam オンラインサインアップシステム」と「クライアント証明書発行システム」を紹介する。

2 国内外の eduroam の動向

2.1 国内の状況

eduroam JP の参加機関数は 2014 年 10 月時点で 86 となり、高等教育機関における普及率は約 7.2% である。機関数の推移を見ると、2006 年の日本導入以来、2011 年末 27 機関、2012 年末 43 機関、2013 年 10 月時点で 56 機関であり、この一年間で大幅に増えたことがわかる。知名度の高まりや、キャンパス無線 LAN のシステム更新時期に合わせて導入が進んでいること、教育・研究の現場における携帯端末利用の増加などが、この背景にあると考えられる。

2014 年度は、専門学校から初の問い合わせがあった。eduroam JP の正式な実施要領・運用基準はまだ策定中であるが、従来は大学・短期大学・高等専門学校が漠然と想定されていたことから、国立情報学研究所内のネットワーク運営・連携本部・認証作業部会で専門学校 (学校教育法の分類) の eduroam 参加可否について審議し、参加を認めることとした。

eduroam の利便性を向上させるために、eduroam 基地局マップのデータ提供を参加機関に呼び掛けている [1]。しかしながら、あまり協力が得られていない。現在は提出データが XML 形式のため、その編集が障害になっている可能性が高く、管理者にとって利便性の高いデータ提供手段を探っている。

2.2 世界の状況

執筆時点で世界の参加国 (地域) 数は 70 であり、大きな変化はないものの、南米やアジアの各国内での展開が進んでいる。アジア地域ではマレーシアが参加し、12 か国 (地域) に至っている。特にインドの 64 機関とタイの 18 機関 (いずれも予定を含む) は、国内展開の速さを見せている。

新興国では電源やネットワークが不安定なところもあり、そのような地域に eduroam を展開する場合、再認証を減らすことのできる安定な認証基盤が必要である。我々が開発してきた耐障害・耐災害 eduroam のアーキテクチャ[1]が応用できる可能性があるが、現時点では導入未定である。

キャンパス外の eduroam サービスとしては、ミュンヘン中心部の広場で基地局が設置されるなど、世界各地で様々な試みが見られる。スウェーデンの空港におけるサービスは今年も継続中であり、SNS 上で称賛の声を多数見ることができる。ノルウェーの空港では、2013 年に試験運用が行われていたが、2014 年 1 月に UNINETT の正式サービスとなった。

3 会議向け期間限定 eduroam アカウントの試行

国際会議では、参加者がネットワークを利用できるように、一時的に基地局を設置したり、既設の基地局のためのゲストアカウントを用意するのが通例である。このようなサービスは、外国からの参加者など、現地の携帯電話網を利用しにくい人々にとって重要である。会議情報やプロシーディングス等を閲覧する目的でも、高速な無線 LAN は利便性が高い。大規模なイベントでは、参加者が持ち込むモバイルルータによる混信が原因で、大勢がネットワークを利用できなくなる問題がある。より快適な基地局を会場で提供し、それに誘導することが望ましい。

会議ごとに基地局を立てる方法は、機材の調達や技術者の確保はもちろん、会場との調整も必要になり、実現にかかるコストが大きい。会場に大容量の基地局が既設であれば、混信等のトラブルも少なく、安定なサービスを提供しやすい。国内で eduroam サービスが利用できる大学施設や会議場が増加したのに伴って、会場の無線 LAN サービスを会議参加者に提供するというニーズが見えてきた。

会場に eduroam 対応の基地局がある場合、参加機関の利用者は各自のアカウントでそのままネットワークが利用できる。しかし、国内外ともに eduroam に未参加の機関もまだ多く、企業の研究所のように eduroam に参加できない機関からの参加者も多い。会議運営者が eduroam 参加機関の構成員であるとは限らず、参加機関であってもアカウントの発行には責任が伴い、ゲスト用に発行するのが難しいという側面もある。会議で eduroam 対応基地局を利用するためには、会議運営者が自身の責任の下でゲストアカウントを容易に取得、配布できる仕組みが必要である。eduroam JP ではこれを実現するための技術と実施要項を検討し、7月に「会議向け期間限定 eduroam アカウントの試行」を開始した。

ゲストアカウントの発行には、我々が開発して

2008 年より実証実験として eduroam 参加機関に提供している、「代理認証システム」を用いた。会議主催者を仮想的な機関とみなして、同システムの「機関」として一時的に登録し、アカウント発行権限を会議運営者に渡す仕組みである。なお、機能的な拡張は一切行っていない。代理認証システムの利用者 ID には、通常は「<大学名>.eduroam.jp」の形式のレルム名が付くが、ゲストアカウントを一般の機関のものと区別するために、会議名と年を基本とする会議略称に comf/symp/mtg などを付加したものをレルム名とするルールを課した。例えば会議略称が EMC14 の場合、レルム名を“emc14-conf.eduroam.jp”のようにする。会議主催者が事前に、会議の正式名称、主催者名、責任者名、実務担当者、会議ウェブサイト、会議期間、アカウント有効期限と発行見込み数などの情報を記入した申請書を代理認証システム管理者に送付し、承認を受けることにより、当サービスを利用できるものとした。

eduroam のアカウント発行はその主体となる機関の責任の下で行われるが、国内の eduroam 運用については eduroam JP にも部分的に責任が及ぶ。このため、当サービスの利用は国内の会議施設・大学施設などで開催される国際会議・国内会議を当面の主な対象として、以下の提供条件を定めた。

- 代理認証システムの登録と同様に、会議名(システム上は機関名)、責任者(会議代表者)、実務担当者を登録すること。
- 会議主催者は SINET 加入機関または学術団体(学振認定)であること。
- 実務担当者は SINET 加入機関(eduroam 参加機関が望ましい)の職員であること。

実務担当者は、以下に従う必要があるものとした。

- 当サービスの利用が許可された後、会議期間の前々日までに代理認証システムを操作して、アカウントのリストを得る。必要に応じてテスト用のアカウントを別途生成し、会場または最寄りの eduroam 参加機関で動作確認を行う。
- 提供条件、使用方法ガイド、アカウント情報を記したアカウント通知を利用者へ渡す。
- 不正利用などにおける責任の明確化のために、アカウントを渡す参加登録者について、妥当な身元把握をすること。例えば、事前に有料参加登録した者については、参加登録をもって身元情報として、一律に配布することができる。当日に参加登録する者や無料(および少額参加費)の会議では、身元確認をしてアカウントを渡すこと。

- 会議主催者が利用者から設定と利用の方法の問い合わせに対応する(会場の基地局運用者に負担をかけない)こと。
- 利用者対し、サービス提供条件を知らせて、順守させること。
- アカウント有効期間中および以後半年間は、利用者の特定を求める照会に回答すること。

当サービスの利用者に対しては、アカウントの利用に際し、以下の提供条件への同意を求めている。

- 会議主催者が参加登録者へ、個人を特定して無償で割り当てるアカウントであること。
- 有効期間中だけ有効であること。有効期間終了後はすみやかにこのアカウントの設定を削除すること(端末の無線 LAN プロファイル“eduroam”を削除)。
- 会場内での利用のために提供されていること。国内の会場外の eduroam 基地局でも使えるかもしれないが、保証されないこと。日本国外の eduroam 基地局で使用しないこと。
- 電気通信法規、会場のネットワーク運用規則を順守し、学術研究活動目的で利用すること。(商用利用や、著作権を侵害するコンテンツの送受信は、禁止)
- 利用に障害が生じても無保証・現状提供であること。(利用者から設定と利用の方法の問い合わせは、会議主催者で対応する)
- 会場管理者や eduroam サービス提供者が会議主催者に対し、運用上の理由で利用者の個人情報を求め得ること。
- 以上の提供条件を記したフォームを受け取り、同意・承諾したうえで、利用すること。

これまでの運用で、会議主催者が学会のような組織ではなく、実行委員会という一時的な組織である例も多いことが判っており、提供条件緩和の声も聞かれる。現在、試行期間として様子を見ながら、提供条件等について検討を続けている段階である。

国外で開催される会議については、現時点で有効なゲストアカウント配布の仕組みはない。会議用ゲストアカウントは国内独自の運用であり、eduroam の責任分界の観点でも、会議開催地で発行するのが妥当と考えられる。早期に eduroam が行き渡った欧州をはじめ、世界のほとんどの国々では、集中型アカウントサービスを持っていない。このため、ゲストアカウントを発行する仕組みをどのようにするかについて、TERENA(eduroam の開発元)の関係者などで議論が行われている段階である。

4 代理認証システムの拡張

4.1 eduroam オンラインサインアップシステム

eduroam JP では、「代理認証システム」と「仮名アカウント発行システム」の二種類の集中型アカウントサービスを提供している。後者は学認 [2] と連携しており、利用者が学認のアカウントを使ってシステムにログインし、eduroam のアカウントを取得できる仕組みである。前者の「代理認証システム」は、機関内の認証システムや学認との連携がなくても eduroam アカウントが利用できることを目標としている。2014 年 10 月現在、eduroam に参加している国内 86 機関のうち約 1/3 にあたる 29 機関が代理認証システムをメインまたは補助的な eduroam IdP (ID Provider) として利用しており、そのうち約 55% が学認に未参加である。

従来の代理認証システムでは、機関管理者がアカウントを希望数だけ発行要求し、ID とパスワードの一覧を CSV 形式などでダウンロードしてから、利用者に配布する必要があった。この方式は、小規模の大学では運用可能な範囲と考えられるが、利用者にアカウントを直接配布する手段が実現できれば、管理者の負担が減り、大人数を擁する大学でも当システムを採用しやすくなるものと考えられる。

ほとんどの機関の学生や教職員は、機関発行のメールアドレスを持っている。我々はこの点に着目し、メールアドレスを間接的に個人認証に利用するオンラインサインアップ機能を開発、付加した [3]。

eduroam の利用を希望する者(以下、利用者)は、eduroam 以外のいずれかのネットワークに接続された端末からサインアップのためのウェブサイトへアクセスし、メールアドレスや本名、メールの到達性を確認するための一時的なパスワードを入力する。メールアドレスは、機関管理者が予め設定したドメイン名(例えば“<大学名>.ac.jp”)に末尾が一致するものだけが利用できる。このように所属機関で発行されたメールアドレスに限定することで、部外者からの不正な申請を抑制している。利用者が申請ボタンをクリックすると、入力されたメールアドレスに対して、本人確認のためのメールが自動的に送られる。この様子を図 1 に示した。メールに記載のリンクをウェブブラウザで開き、先に入力したパスワードを入力することで本人確認が完了し、申請内容が機関管理者に通知される。このとき、どの機関の管理者に通知メールが送られるかは、登録されたドメイン名(部分)によってシステムが自動的に判断する。利用者はこの時点で ID・パスワードを得るが、機関管理者によってロックが解除されるまでは、実際に使うことはできない。

通知メールを受信した機関管理者は、管理用のウェブサイトにログインして申請内容を調べ、アカ

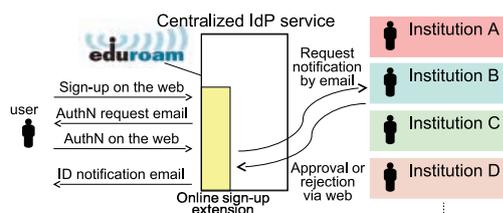


図 1: eduroam オンラインサインアップシステム

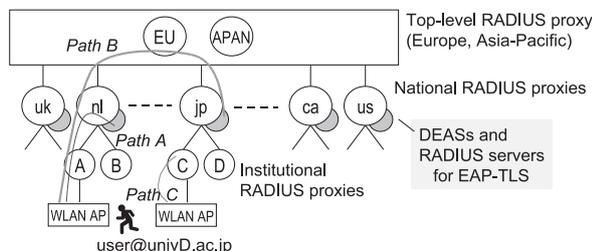


図 2: 耐障害・耐災害 eduroam のアーキテクチャ

ウントの有効期間を設定して承認するか、理由を記入して拒否の操作を行う。この承認結果は、申請者のメールアドレスに送付される。

4.2 クライアント証明書発行システム

eduroam は IEEE802.1X に基づく認証方式を採用しており、PEAP や EAP-TTLS, EAP-TLS などの様々な方式が利用できる。このうち PEAP は、Windows を始め、MacOS, Android, iOS, Linux など幅広いオペレーティングシステム (OS) が対応しており、eduroam では世界的に主流となっている。代理認証システムでも PEAP を採用している。

PEAP では ID・パスワードのペアがアカウントとして利用されるが、運用方法や利用場面によっては、電子証明書を用いる EAP-TLS 方式が適している場合がある。我々が開発している耐障害・耐災害 eduroam のアーキテクチャ [1] では、EAP-TLS 方式によって耐障害性と効率的な認証処理を実現している。例えば、図 2 において、日本 (jp) の代理認証システムの CA 証明書を機関 C が事前に取得しておけば、端末の認証を Path C の経路のみで効率的に行うことができる。将来的にこのようなシステムを実現する準備に加えて、現在の eduroam の構成のままで EAP-TLS 方式を利用したいという機関もあることから、これをサポートするための「クライアント証明書発行システム」を代理認証システムに追加した [3]。

クライアント認証のための電子証明書はデータが大きく、パスワードのような手入力では事実上困難である。そのため、利用者自身が証明書を端末にダウンロードできる手段を提供する必要がある。開発したクライアント証明書発行システムでは、eduroam アカウントとして PEAP 用に発行された ID・パスワードのペアをシステムの利用者認証に流用した。

利用者は、eduroam アカウントからレルム部分を除いた ID を用いて、同システムのウェブサイトログインする。端末に証明書をインストールするのに必要なパスフレーズを入力し、証明書発行のボタンをクリックする。システム内部で PKCS#12 形式の証明書ファイルが生成され、ダウンロードのためのリンクがウェブ画面に表示される。利用者はこの証明書ファイルを端末にダウンロードし、先に入力したパスフレーズを用いてインストールする。

クライアント証明書がインストールできることは、Windows, OS X, iOS で確認した。Android については、ベンダやバージョンによるばらつきが大きく、導入を簡略化するツールの開発が望まれる。

このクライアント証明書発行システムは、機関管理者がその利用可否を選択できる。初期状態では利用不可に設定されており、利用者は証明書を取得できない。また、万一の悪用に備えて、機関管理者は証明書失効権限を有しており、不正利用が発覚した場合は当該証明書を CRL (Certificate Revocation List) に加え、また、再発行の操作を禁止できる。

5 むすび

eduroam の国内外の最新状況を概説した。国内ではキャンパス無線 LAN の更新時に eduroam を導入する例が多く見られるようになっている。

国内で開催される会議等のために、会議向け機関限定 eduroam アカウントの試行を開始した。eduroam 基地局の整備された大学施設・会議場であれば、会議主催者が主体となってゲストアカウントを発行し、利用者の便宜を図ることができる。

また、代理認証システムの拡張として、「eduroam オンラインサインアップシステム」と「クライアント証明書発行システム」を開発した。国内機関の eduroam 参加の加速と、耐障害性・耐災害性を有する eduroam システムの開発に貢献が期待される。

参考文献

- [1] 後藤英昭, 曾根秀昭, “キャンパス無線 eduroam の国内外の最新動向 - 利便性と耐障害・耐災害性の向上 -,” 大学 ICT 推進協議会 2013 年度年次大会 論文集 W3E-5, pp.122-125, 2013.
- [2] 西村 健, 中村素典, 山地一禎, 大谷 誠, 岡部寿男, 曾根原 登, “日本における学術認証フェデレーション “学認” の展開,” 大学 ICT 推進協議会 2011 年度年次大会 論文集, 2011.
- [3] T. Niizuma and H. Goto, “Centralized Online Sign-up and Client Certificate Issuing System for eduroam,” COMPSAC Workshop MidArch 2014, pp.174-179, 2014 (Västerås, Sweden).

[大学 ICT 推進協議会 2014 年度 年次大会論文集より転載]

東北学術研究インターネットコミュニティの運用に関する報告

森 倫子†, 曾根 秀昭††, 千葉 実†, 北澤 秀倫†

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††

nmcctu-as2014@g-mail.tohoku-university.jp

概要：東北学術インターネットコミュニティ（以下 TOPIC）は、東北地区の学術研究・教育活動を支援するコンピュータネットワーク環境の発展に貢献することを目的として設立された団体で、主に参加機関へのネットワーク接続の提供と、講演会・研修会等による情報提供・情報交換を中心に活動してきた。近年解散したり下火になる地域ネットワークコミュニティが多い中、今なお活発に活動する TOPIC について、事務局として永年携わってきた立場からご報告する。

1 はじめに

TOPIC は東北学術研究インターネットコミュニティ (Tohoku OPen Internet Community) の略称である。東北地区の学術研究・教育活動を支援するコンピュータネットワーク環境の発展に貢献することを目的として 1992 年に設立された。

2 TOPIC について

2.1 TOPIC の組織

TOPIC の組織図を図 1 に示す。

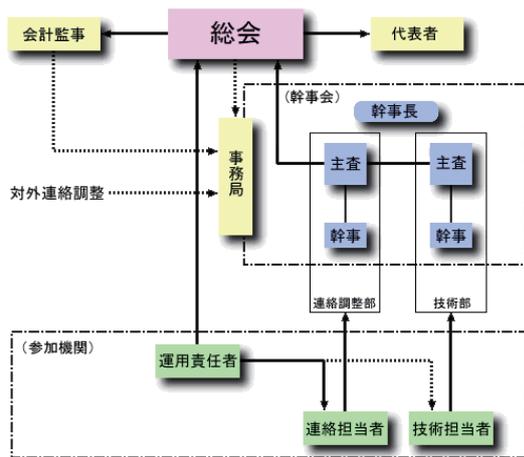


図 1 組織図

TOPIC の運営方針は幹事会により決定される。幹事会には、参加機関および幹事間の連絡調整を担当する連絡調整部、技術的検討を行う技術部が設置され、事務局がその事務を担当している。

幹事は弘前大学、岩手大学、秋田大学、東北大

学、東北学院大学、山形大学、会津大学、福島大学等から選出された十数名の教職員にて構成されている。

事務局は、東北大学情報部情報基盤課が担当している。

TOPIC の参加資格は、「東北地区における大学及び高等専門学校」、「東北地区における国又は地方公共団体に設置されている学術研究機関」である。参加機関には、運用責任者、連絡担当者、技術担当者が置かれ、運用責任者は機関の運用全般に責任を有する者であると共に、総会に出席し、事業計画や予算・決算に関すること等の重要事項を審議する。連絡担当者は連絡調整部、技術担当者は技術部との調整を行う。

2.2 TOPIC 参加機関数推移

TOPIC 参加機関数の推移を図 2 に示す。

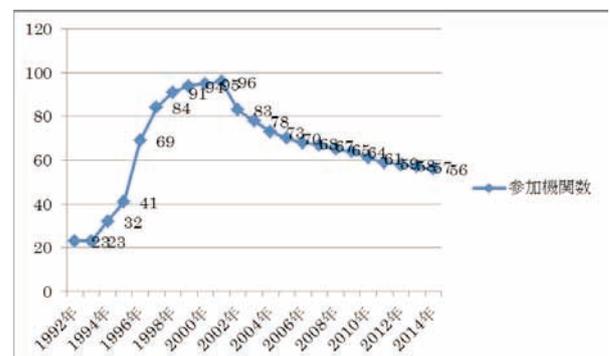


図 2 TOPIC 参加機関数推移

インターネットの普及と共に増えていった

組織数が、2001年の96機関をピークに、一般のISPやSINET直接接続へ移行することで徐々に減少していったことが見てとれる。

2.3 TOPICの活動内容

TOPICの主な活動は、(1)参加機関のネットワークを相互に接続するネットワークの運用とインターネットとの接続・運用に関する調整、(2)ネットワークを利用した学術研究・教育活動の支援である。(2)の事業として主に講演会・研修会の開催があげられる。

2.4 TOPICの運用経費

TOPICの運用は会員組織からの会費で賄われており、IPアドレス維持料やドメイン名登録更新料、総会・講演会・研修会開催経費等に使用されている。

1組織あたりの年額は当初、(1)短期大学を除く大学の会員は10万円、(2)(1)以外の会員は5万円、(3)(1)および(2)の機関のうち専用線による接続で64kbps以下の会員は5千円、未接続の会員は無料であった。しかし、TOPICの会費は、情報収集や研修などに係わる運営経費を分担するものであり、接続の提供を受けることへの対価ではないとの考えから、2002年以降は一律5万円となった。

3 TOPICのネットワーク運用

3.1 TOPICのネットワーク構成

2000年のネットワーク接続図を図3に示す。

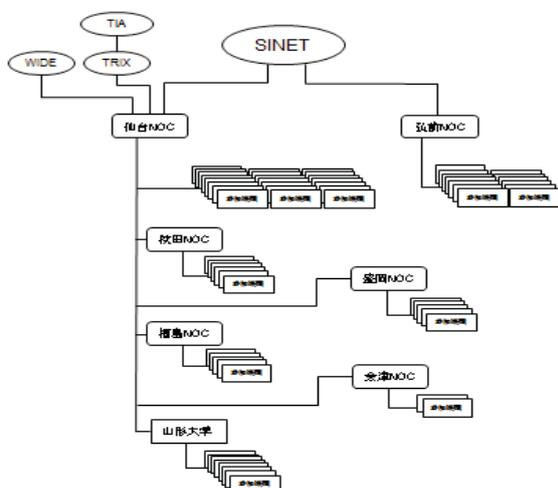


図3 2000年のネットワーク接続図

TOPICは、NOC (Network Operation Center) を東北大学、弘前大学、岩手大学、秋田大学、会津大学、福島大学に設置し、参加機関はそれぞれ回線を調達し、NOCに接続する構成になっている。各NOCは仙台NOCへと接続し、そこからSINET等へ接続することでインターネットを利用できる。

現在のネットワーク構成図を図4に示す。

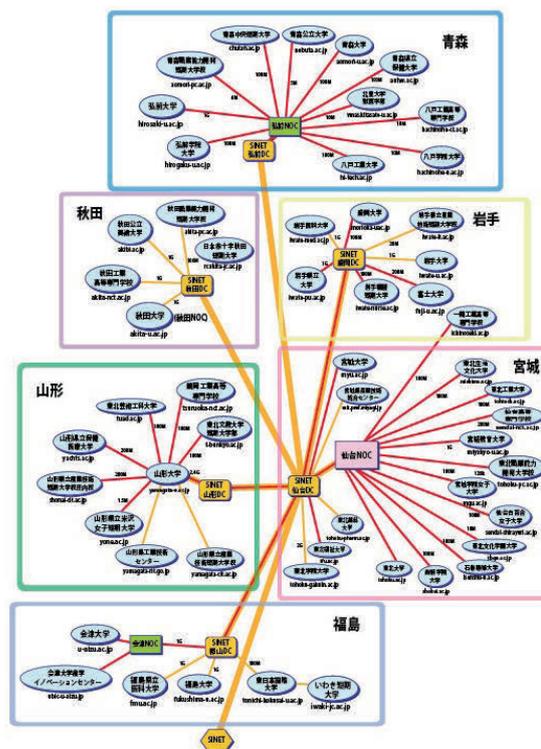


図4 現在のネットワーク接続図

2011年にSINETデータセンター(以下DC)が各県に設置されてからはSINET DCへ直接接続する組織が増えているが、TOPIC CIDRのアドレスをそのまま利用したい等の理由からVPN等によりTOPICに接続している組織もまだある。

TOPIC仙台NOC、SINETのノード室は東北大学サイバーサイエンスセンター内に設置されている。

3.2 ネットワーク運用に係る事務局の業務

事務局の業務は、東北大学サイバーサイエンスセンターネットワーク研究部(以下ネットワーク研究部)と連携して、ネットワーク機器・システムの導入・整備を行い、運用することである。例えば、機器の導入、保守契約、TOPICへの接続・変更申請の対応、JPNIC・JPRSへの

ネットワーク情報の申請，接続の際の入室対応等を行っている。

TOPIC は 202. 211. 0. 0/20 と 210. 156. 32. 0/19 の CIDR および AS2503 を保有している。また，JPNIC の非営利会員となっている。事務局は TOPIC に接続する組織への IP アドレスの割り当てや，ネットワーク情報の登録・変更申請，JPNIC 会費や IP アドレス維持料の支払い業務を行っている。

2001 年に JP ドメイン名の登録管理業務が JPNIC から JPRS へと移管された際，TOPIC は指定事業者となり TOPIC 参加機関のドメイン名（各組織につき一つ）の登録業務，ドメイン名の登録・登録更新料の支払い業務を行うこととなった。

3.3 サーバの運用

TOPIC は，運用のため topic.ad.jp のドメイン名を保有しており，事務局はその DNS・メール・ウェブのサーバを運用している。また，TOPIC の活動の一環で，東北地区における学会活動の支援として，10 程度の学会（支部）に対しメーリングリストやウェブのサービスを提供している。

4 講演会・研修会

さて，ネットワーク運用に次いで，あるいはそれ以上に重要な事業が講演会・研修会の実施である。

TOPIC 主催の講演会・研修会は年に 2 回開催されている。一つは春に開催される総会に併設のもの，もう一つは秋に東北 6 県持ち回りで開催される「ネットワーク担当者研修会」である。

総会時に開催される講演会・研修会は，総会直後の枠では，情報基盤センター長等の運用責任者も参加しているため，基調講演的に情報基盤の整備・運用の方向性に関わるような重要な話題を，2 日目午前には技術担当者向けの技術的な話題，ICT の最新の動向等の講演を行う，概ねそのような構成で開催している。

秋の研修会は一泊二日の合宿形式で，参加機関に対して参加費の半額程度を TOPIC から援助

する等，技術職員等が参加しやすくなるよう工夫している。夜のセッションでは，技術職員も事例報告を行うなど，活発な情報交換，意見交換が行われている。

その他，各県の NOC が主催の講演会・研修会も開催されている。

事務局の業務としては，幹事会で決定した内容に従い謝金や旅費等の事務手続きや，プログラムのとりまとめを行うこと等である。講師への講演依頼は担当幹事が，会場の手配，申し込みの受け付け，部屋割り，資料のコピー・配布，当日の受け付け等は開催校が行う。場合によっては事務局が講演の依頼を行うこともある。

4.1 講演会・研修会の内容の変遷

TOPIC 講演会・研修会の大きな内容の変遷を示したいと思う。

第 1 回総会時（1992 年）には，弘前大学医学部松谷氏による「インターネット（JAIN）接続，光と影」と，東北大学電気通信研究所亀山氏「東北地区のインターネットの現状と将来」の二つの講演が行われた。このように TOPIC 設立初期にはネットワーク接続環境を中心とした講演が行われた。

その後 1990 年代には，サーバの運用についての講演，またマルチメディアというキーワードと共にネットワークの利用に関する講演が行われるようになった。1997 年には「ネットワークとセキュリティ」と題する研修会も開催されており，早い段階で参加機関へのセキュリティの啓発が始まっている。2000 年代になると，認証関係，無線 LAN 関係，IPv6 関係，ソフトウェアライセンス関係等様々な分野の講演が行われた。また，SINET や学認に関する報告も定期的に行われ，NII との強い結びつきも生まれた。東日本大震災後には参加機関による被災状況・対応報告が行われ，その後継続して耐災害や BCP 関連の講演が行われるようになった。

このように，ネットワークに止まらず情報基盤の整備・運用・利用に関する数多くの講演が開催され，参加機関にとって大変有意義なものとなっている。

4.2 研修会参加者の推移

TOPIC 研修会の参加者数について記録に残っている 2006 年以降分を図 5 に示す。

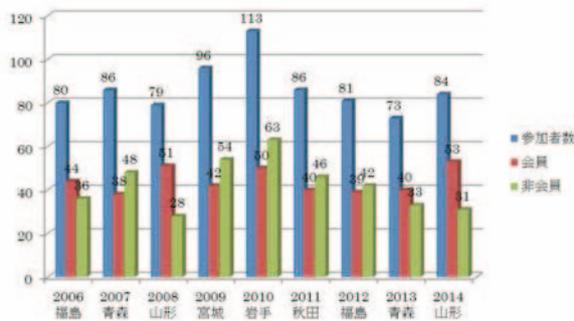


図 5 TOPIC 研修会の参加者数推移

年によってばらつきがあるが、全体で 80 名から 100 名程度の参加となっている。ここで、非会員となっている参加者の中には各参加機関の保守・運用業者も含まれる。自組織に専任の技術担当者を持たず、業者に運用を任せている組織もある。大学等のネットワーク運用の在り方について業者も一緒に学ぶことができることも TOPIC の講演会・研修会の特徴の一つと言えるかもしれない。

5 東日本大震災時の TOPIC の状況

2011 年 3 月 11 日地震発生直後にサイバーサイエンスセンターは停電になった。建物に損傷はなく、職員は一旦屋外へ避難した後、機器室の確認を行った。ラックの倒壊もなく機器の損傷もなかった。停電後 30 分程度で SINET の機器が UPS のバッテリー切れにより停止し TOPIC の接続機関は SINET への接続ができなくなった。TOPIC の機器も 2 時間半程度で UPS のバッテリーがなくなり TOPIC 参加機関同士の接続もできなくなった。

3 月 12 日には再度職員で点検を行い、建物の使用や機器の損傷がなく復電すればネットワークの運用を再開できる旨東北大学の災害対策本部へと報告を行った。

3 月 13 日 14 時半頃、サイバーサイエンスセンターが復電し、TOPIC および SINET は無事運用を再開した。

6 おわりに

本稿では、TOPIC の運用と事務局の業務について報告した。

TOPIC の主要な活動はネットワーク運用と講演会・研修会の開催であると述べてきたが、ISP 接続や SINET 直接接続の増加に伴う TOPIC NOC への接続の減少により、ネットワーク運用の比重は下がりつつある。

一方、クラウドの利用やセキュリティの問題等、新たなネットワークの利用方法、ネットワークを取り巻く問題は増加する傾向にある。

最新情報の入手が困難だったり、あるいは横のつながりがなく、ネットワークの整備・運用に悩みを抱える東北の大学・学術機関は少ないことと思う。今後も、講演会・研修会等を通じ、地域ネットワークコミュニティとして歩んでいくであろう TOPIC に、事務局の一員として引き続き貢献していければ幸いである。

謝辞

4.1 節で述べた研修参加者の推移のグラフ化にあたっては、岩手大学の庭田氏にご尽力頂いた。

参考文献

- [1] JPNIC 会員と語る, JPNIC ニュースレター No.31.
<https://www.nic.ad.jp/ja/newsletter/No31/>

[報告]

SC14 報告

東北大学サイバーサイエンスセンター 小松一彦

2014年11月16日～21日に米国ルイジアナ州ニューオーリンズ市において International Conference for High Performance Computing, Networking, Storage and Analysis (通称SC)が開催されました。SCはスーパーコンピューティングとその関連技術に関する世界最大の国際会議であり、1988年から年に1回開催されており、今回で26回目を迎えます。今回、SC14に参加し、サイバーサイエンスセンターにおける研究活動を発表してきましたので、その報告をいたします。

SCは技術論文が発表される本会議と、大学・研究所・企業等が最新の研究成果・製品を発表する展示会から構成されています。サイバーサイエンスセンターは、流体科学研究所、金属材料研究所、そして、今年から新たに参加に加わった東北メディカル・メガバンク機構と連携し、スーパーコンピューティングに関する活動について、東北大学として展示会へ参加いたしました。展示ブースでは、サイバーサイエンスセンターのスパコンシステムの紹介や次期システムの紹介をはじめ、2.5・3次元積層技術を活用したプロセッサに関する研究やスーパーコンピュータを容易に使うためのシステムソフトウェアに関する研究など、次世代の高性能計算基盤技術に関する最新の研究成果の展示や発表を行いました。また、CREST「ポストペタスケール高性能計算に資するシステムソフトウェア技術の創出」展示ブースにおいても、スパコン研究部の研究成果の一部を展示いたしました。17～20日の4日間の展示会期間中、展示ブースへの訪問者は300名以上にのぼり、展示ブースでのプレゼンテーションなどを通じて、活発な議論・情報交換を行うことができました。

今年のSC15は11月15日～20日に米国テキサス州オースティン市で行われる予定です。



SC14の会場



展示の様子



東北大学展示スタッフ

[Web 版大規模科学計算システムニュースより]

大規模科学計算システムニュースに掲載された記事の一部を転載しています。 <http://www.ss.cc.tohoku.ac.jp/tayori/>

利用負担金（1月～3月分）の請求について (No. 181)

大規模科学計算システムの平成27年1月～3月分の利用負担金の請求を次のように行います（昨年と同じです）。

《学内支払責任者》

1月から2月10日までの利用負担金を2月下旬に請求します。その後3月末までの利用負担金は、翌年度4月に請求します（平成27年度の大学運営資金で予算の調整を行います）。

なお、事前に費目の指定は必要ありません。請求金額確定後、センター会計係より各部署の会計担当を通して照会いたします。

《学外支払責任者》

1月から2月10日までの実績額と3月末までの利用見込額を合算した利用負担金を2月下旬に請求します。利用見込額は以下の mikomi コマンドで2月9日(月)までに申告してください。なお、見込額と実績が異なっても精算は行わないのでご注意ください。

また、支払費目名の入った利用負担金請求書を希望する場合や、請求書の適要欄等について不明な点がある場合は、事前に会計係（022-795-3405）へご連絡くださるようお願いいたします。

【 mikomi コマンドの使用例 】

```
ssh front.isc.tohoku.ac.jp -l 利用者番号 *並列コンピュータにログイン
利用者番号@front.isc.tohoku.ac.jp's password: パスワード して下線部を入力します。
gen% mikomi
```

2月5日 現在の見込み額は次のとおりです。

支払責任者 : ○○ ○○ (u23456)

見込み額指定者 :

見込み額 : 0円

1. 見込み額の指定 2. 削除 9. 終了

何番の処理を選びますか? 1

見込み額を入力してください(円)? 200000

登録してよろしいですか (yes/no) ? yes

.....

gen%

負担金全般に関することで不明な点がある場合は、共同利用支援係（022-795-6251）へご連絡くださるようお願いいたします。

(共同利用支援係, 会計係)

利用負担金額の表示コマンドについて (No. 181)

本センター大規模科学計算システムでは、利用者の利用額と支払責任者ごとの利用額・負担額を表示するためのコマンドとして kakin, skakin があります。これらのコマンドは、並列コンピュータ (front.isc.tohoku.ac.jp) にログインして使用します。

コマンド名	機 能
kakin	利用者ごとの利用額を各システム、月ごとに表示
skakin	支払責任者ごとに集計した利用額と負担額を表示 (負担額は割引制度に基づいた金額)

いずれも、前日までご利用いただいた金額を表示します。コマンド使用例は大規模科学計算システムウェブページをご覧ください。

負担金の確認

<http://www.ss.cc.tohoku.ac.jp/utilize/academic.html#負担金の確認>

(共同利用支援係)

研究成果リスト提出のお願い (No. 181)

平素、本センターの大規模科学計算システムをご利用いただきありがとうございます。

本センターでは、学術研究を支える世界最高水準の大規模科学計算システムの導入と利用環境の整備・拡充を行い、研究の発展に資することを心掛けてまいりました。今後もシステムの整備を進めていくには、大規模科学計算システムが多くの研究分野で必要不可欠であり、かつ研究成果が得られていることを広くアピールしていく必要があります。

つきましては、本センター大規模科学計算システムを利用して得られた研究成果を、下記により提出くださるようお願い申し上げます。なお、提出していただく研究成果は、平成 26 年度中に発表されたものとしします。

記

1. 研究成果リスト：著者名、論文名、掲載誌（巻号頁）、発表年
2. 提出方法：電子メールでお願いします。
提出先メールアドレス seika@cc.tohoku.ac.jp
3. 締切り日：平成 27 年 4 月 17 日（金）
4. 問合せ先：共同利用支援係（022-795-6251, uketuke@cc.tohoku.ac.jp）

<< 論文等への利用の明記について >>

研究成果を論文等で発表する場合には、本センターを利用した旨を明記くださるようお願いいたします。

— 記入例 —

「本研究の実験結果の一部は、東北大学サイバーサイエンスセンター大規模科学計算システムを利用して得られた。」

Part of the experimental results in this research were obtained using supercomputing resources at Cyberscience Center, Tohoku University.

(共同利用支援係)

— SENAC 執筆要項 —

1. お寄せいただきたい投稿内容

次のような内容の投稿のうち、当センターで適当と判定したものを掲載します。その際に原稿の修正をお願いすることもありますのであらかじめご了承ください。

- ・一般利用者の方々が関心をもたれる事項に関する論説
- ・センターの計算機を利用して行った研究論文の概要
- ・プログラミングの実例と解説
- ・センターに対する意見、要望
- ・利用者相互の情報交換

2. 執筆にあたってご注意いただく事項

- (1)原稿は横書きです。
- (2)術語以外は、「常用漢字」を用い、かなは「現代かなづかい」を用いるものとします。
- (3)学術あるいは技術に関する原稿の場合、200字～400字程度のアブストラクトをつけてください。
- (4)参考文献は通し番号を付し末尾に一括記載し、本文中の該当箇所に引用番号を記入ください。
 - ・雑誌：著者、タイトル、雑誌名、巻、号、ページ、発行年
 - ・書籍：著者、書名、ページ、発行所、発行年

3. 原稿の提出方法

原稿のファイル形式はWordを標準としますが、PDFでの提出も可能です。サイズ*は以下を参照してください。ファイルは電子メールで提出してください。

—Wordの場合—

- ・用紙サイズ：A4
- ・余白：上=30mm 下=25mm 左右=25mm 綴じ代=0
- ・標準の文字数（45文字 47行）

<文字サイズ等の目安>

- ・表題=ゴシック体 14pt 中央
- ・副題=明朝体 12pt 中央
- ・氏名=明朝体 10.5pt 中央
- ・所属=明朝体 10.5pt 中央
- ・本文=明朝体 10.5pt
- ・章・見出し番号=ゴシック体 11pt～12pt

*余白サイズ、文字数、文字サイズは目安とお考えください。

4. その他

- (1)執筆者には、希望があれば本誌*（10部以内の希望部数）と本誌PDF版を進呈します。
*2014年末で、別刷の進呈は終了しました。
- (2)投稿予定の原稿が15ページを超す場合は共同利用支援係まで前もってご連絡ください。
- (3)初回の校正は、執筆者が行って、誤植の防止をはかるものとします。
- (4)原稿の提出先は次のとおりです。

東北大学サイバーサイエンスセンター内 情報部情報基盤課共同利用支援係

e-mail uketuke@cc.tohoku.ac.jp

TEL 022-795-3406

スタッフ便り

今年も宜しく願いいたします。さて、新年を迎え我々センター職員が一丸となって、設計・開発・導入に取り組んで来た新大規模科学計算システムの運用が開始されました。これまで以上パワフル、かつ使いやすい計算環境を利用者の皆さんに提供出来るよう微力ながら頑張っていきたいと思っております。新棟に導入された新システムはお申し込みいただければ、見学することが可能ですので、ご興味のある方は是非いらしてください！

新年にあたり個人な目標としては、

- ・食を楽しみ、週2回のジョギングをこなしながら、体重の増加を抑えつつ健康を保つこと
- ・できることはできるうちにこなして、積み残し・持ち越しの無い日々を過ごすこと
- ・家族と過ごす時間をふやすこと

の3つを掲げて、素敵、いや無敵の一年にしたいと考えております。(R.E)



写真上:センター新棟(2号館) 正面

写真下:センター新棟(2号館) 新キャンパス側から



新年あけましておめでとうございます。この度の年末年始休暇は大型連休であり、皆様は有意義な時間をお過ごしできたでしょうか。私は有馬記念で惨敗でした。

センターに異動してきて半年、オープンキャンパスや日々の業務を通じて少しずつですが、センターのことが理解出来てきました。当初は、「スパコン？初耳です。」という状態でしたが、「スパコン知らないの?!」と上から目線で振る舞える知識は得られたと感じております。

今年はモンテディオ山形がJ1に昇格したため、ベガルタ仙台との“みちのくダービー”が久々に行われます。しかも開幕戦として開催される可能性もあり、アツイ戦いとなりそうです。私は山形出身のためモンテディオサポであり、久々に仙台でモンテディオを応援できることに嬉しさで一杯です。もちろんスタジアムで観戦したいですが、開幕戦はまだ寒いので、家でまったりと応援でいいかなと考えております。ぜひJ1で旋風を巻き起こしてくれることを願っています。

いよいよ新棟が完成となり、SX-ACEの本格起動も間近に迫ってきました。センターの一員として力になれるよう日々成長していければと考えております。(T.E)



整備中の青葉山新キャンパス

SENAC 編集部会

小林広明 曾根秀昭 水木敬明 後藤英昭
江川隆輔 佐藤恵美子 高杉佳奈 大泉健治
小野 敏 斉藤くみ子

平成 27 年 1 月 発行

編集・発行 東北大学
サイバーサイエンスセンター
仙台市青葉区荒巻字青葉 6-3
郵便番号 980-8578

印刷 東北大学生協同組合
プリントコープ

索引

SENAC Vol. 47 (No. 1~No. 4)

2014. 1~2014. 10

項目	No	p.	項目	No	p.
巻頭言			平成26年度サイバーサイエンスセンター講習会案内)	2	48
2014年東北大学及びサイバーサイエンスセンターへの期待 渡邊國彦	1	1	平成26年度(4月~9月)の負担金制度について	2	49
共同研究成果			新棟建築工事期間中の出入口等の変更について	2	52
A building-resolving simulation of sea breeze over Sendai downtown with a parallelized CFD model Guixing Chen 他	1	7	ベクトル型スーパーコンピュータ「SX-ACE」の導入を決定	3	1
高次のCBFMを用いた誘電体近傍アンテナの数値解析 今野佳祐 他	1	13	スーパーコンピュータシステムの更新について(1)	3	3
数値タービンシステム(NTS)の大規模並列計算 山本悟 他	1	19	東北大学とNEC、次世代スーパーコンピュータ技術の共同研究部門を開設	3	5
超低周波地震の活動に基づくプレート間固着の推定 有吉慶介 他	1	29	サイバーサイエンスセンターの新ロゴマーク決定	4	1
Building-Cube Methodを用いたエンジンナセルインレットからの騒音伝播解析 福島 裕馬他	1	35	ドメイン名変更のお知らせ	4	2
超高速第一原理電子状態計算コードの開発と応用 柳澤将 他	1	47	報告		
グリーン関数法を用いた有機分子における紫外光電子スペクトルの理論計算月と火星の全球地震波伝播シミュレーション 山崎馨他	2	53	SC13 報告 小松 一彦	1	65
気候モデルデータのダウンスケーリングによるヤマセの将来変化 島田照久他	2	59	平成25年度全国共同利用情報基盤センター顕彰について	2	77
航空機エンジン排気ジェットと後流渦の相互作用の解析 三坂孝志他	3	27	平成25年度東北大学サイバーサイエンスセンター顕彰について	2	78
起電力法を用いた低姿勢な大規模リフレクトアレーの設計 今野佳祐他	4	3	八巻研究員と渡辺前研究員と木下前研究員が電子情報通信学会学術奨励賞を受賞しました	2	79
Computational prediction of niobium-silicide nanopowder growth in thermal plasma synthesis Masaya Shigeta	4	11	サイバーサイエンスセンターとドイツジーゲン大学情報メディア技術センターが研究協定締結	2	80
民間企業利用サービス利用成果			第19回高性能シミュレーションに関するワークショップ(WSSP)報告	2	81
民間航空機開発における大規模空力弾性解析シミュレーションの適用 森野裕行	2	67	第12回シナジー研究会報告	2	83
大規模科学計算システム			EMC'14 報告 江川隆輔	3	59
並列コンピュータ LX 406Re-2 の利用法	2	1	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」発足式及び祝賀会を開催	3	60
アプリケーションサービスの紹介	2	25	JHPCN 学際大規模情報基盤共同利用・共同研究拠点 第6回シンポジウム報告 江川隆輔	4	17
三次元可視化システムの紹介	2	47	富岡文部科学大臣政務官がサイバーサイエンスセンターを訪問	4	25
LX 406Re-2 のハードウェア	3	7	プレスリリース報告 一東北大・NEC・国際航業、大規模地震発生時の津波浸水被害をリアルタイムに予測する実証事業を推進-	4	27
三次元可視化システムの利用法	3	15	菅沼教授らの研究グループが、KDDI 研究所等と共同で、ビッグデータに関する国際連携プロジェクト研究を開始	4	29
資料			TDK(株)が東北大学サイバーサイエンスセンター大規模科学計算システム(NEC SX-9)を活用してハードディスクの容量を2倍以上にできる基幹部品を開発	4	29
MATLAB Graphical User Interface 開発環境 GUIDE について 陳国躍他	3	37	利用相談室便り		
大学 ICT 推進協議会 2013 年度 年次大会論文集より転載			平成26年度の利用相談について	3	57
東北大学サイバーサイエンスセンターにおける分子動力学シミュレーションコードの高速化支援について 森谷友映他	1	51	展示室便り		
キャンパス無線 eduroam の国内外の最新動向 -利便性と耐障害・耐災害性の向上 後藤英昭他	1	57	①地球シミュレータ 板倉憲一	1	67
東北地区の大学・高専等のウェブサーバにおけるサーバ証明書の利用状況 水木敬明他	1	61	②Classic Macintosh 佐藤大	2	84
お知らせ			③ネットワーク	4	30
並列コンピュータシステムの更新について(1)	1	3	「Web 版大規模科学計算システムニュース」より		
新棟建築工事期間中の出入口等の変更について	1	5	スーパーコンピュータ SX-9 への直接ログインの停止について (No. 159)	1	70
計算科学・計算機科学人材育成のためのスーパーコンピュータ無償提供制度について	1	6	センター講習会資料の公開について (No. 159)	1	70
			SAS のサービス終了について (No. 161)	1	70
			利用負担金 (1月~3月分) の請求について (No. 161)	1	71
			利用負担金額の表示コマンドについて (No. 161)	1	71
			研究成果リスト提出のお願い (No. 161)	1	72
			計算科学・計算機科学人材育成のためのスーパーコンピュータ無償提供制度について (No. 163)	2	88

民間企業利用サービスについて (No. 163)	2	88		
並列コンピュータシステムの更新について(3)				
(No. 165)	2	89		
共同利用支援係 (受付窓口) の移動について				
(No. 165)	2	90		
大規模ファイル領域の利用について (No. 166)	2	90		
新並列コンピュータでのアプリケーションサービスについて (No. 166)	2	91		
平成 26 年度の共同研究について (No. 167)	2	92		
新棟工事期間中の出入口等の変更について[再掲載・一部変更その 2] (No. 169)	3	62		
MATLAB のバージョンアップについて (No. 171)	3	63		
Gaussian 09, GaussView 5 の媒体貸出について				
(No. 171)	3	63		
平成 26 年度講習会案内 [8 月、9 月開催分]				
(No. 171)	3	64		
平成 27 年度の「京」を含む HPCI システム共用計算資源の利用研究課題の募集について (No. 175)	4	38		
負担金の支払い費目について (No. 176)	4	39		
利用負担金額の表示コマンドについて (No. 176)	4	39		
新棟工事期間中の出入口等の変更について[再掲載・一部変更その 3] (No. 176)	4	40		
平成 26 年度(10 月～)の負担金制度について				
(No. 177)	4	43		

システム一覧

計算機システム	機種
スーパーコンピュータ	SX-9
並列コンピュータ	LX 406Re-2

フロントエンドサーバ

ホスト名	front.isc.tohoku.ac.jp
------	------------------------

サービス時間

利用システム名等	利用時間帯
スーパーコンピュータ	連続運転
並列コンピュータ	連続運転
可視化機器室	平日 9:00~21:00
館内利用	平日 8:30~21:00

スーパーコンピュータのジョブクラスと制限値

計算機システム	処理方法	キュー名	CPU 時間制限 [時間]	メモリ容量 [GB]
スーパー コンピュータ SX-9	会話型処理	- (4cpu)	1	8
	バッチ処理	ss (4cpu)	1	256
		s (4cpu)	無制限	32
		p8 (8cpu)	〃	512
		p16 (16cpu)	〃	1024
		p32 (32cpu)	〃	1024×2
		p64 (64cpu)	〃	1024×4

並列コンピュータのジョブクラスと制限値

計算機システム	処理方法	キュー名	利用ノード数 (コア数)	時間制限 [時間]	メモリ容量 [GB]
並列コンピュータ LX 406Re-2	会話型処理	-	1 (6)	1 ※2	8
	バッチ処理 ※1	ns	1 (1)	無制限	5
		nh	1 (24)	1 ※3	128
		n1	1 (24)	無制限	128
		n6	6 (144)	〃	128×6
		n12	12 (288)	〃	128×12
		n24	24 (576)	〃	128×24
		mg (アプリケーション専用)	1 (24)	〃	128

※ 1. 2ノード以上を利用した並列実行にはMPIの利用が必須

※ 2. CPU時間の合計時間

※ 3. 経過時間

目次

東北大学サイバーサイエンスセンター

大規模科学計算システム広報 Vol.48 No.1 2015-1

[巻頭言]				
「普通の人々のためのスーパーコンピュータセンター」を目指して	小林 広明	1		
[お知らせ]				
スーパーコンピュータシステムの更新に伴う今後のスケジュールについて		3		
[大規模科学計算システム]				
スーパーコンピュータシステム SX-ACE のハードウェア	萩原 孝・浜口 博幸・山信田 恒	5		
SX-ACE でのプログラミング (ベクトル化編)	工藤 淑裕・横谷 雄司	15		
[共同研究成果]				
東北地震に伴う固有地震活動の揺らぎから推定された摩擦特性と余効すべり伝播過程	有吉 慶介・内田 直希・松澤 暢 日野 亮太・長谷川 昭・堀 高峰 金田 義行	33		
[大学 ICT 推進協議会 2014 年度 年次大会論文集より転載]				
スーパーコンピュータシステム SX-ACE の紹介	山下 毅・森谷 友映・佐々木大輔 齋藤 敦子・小野 敏・大泉 健治 岡部 公起・江川 隆輔・小林 広明	39		
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	47		
リアルタイム津波浸水・被害予測シミュレーションシステム開発の取り組み	大泉 健治・阿部 孝志・佐藤 佳彦 松岡 浩司・撫佐 昭裕・小林 広明	54		
キャンパス無線 eduroam の最新動向と国内機関向け新サービス	後藤 英昭・新妻 共・中村 素典 曾根 秀昭	58		
東北学術研究インターネットコミュニティの運用に関する報告	森 倫子・曾根 秀昭・千葉 実 北澤 秀倫	62		
[報告]				
SC14 報告	小松 一彦	66		
[Web 版大規模科学計算システムニュースより]				
利用負担金(1月~3月分)の請求について(No.181)		67		
利用負担金額の表示コマンドについて(No.181)		67		
研究成果リスト提出のお願い(No.181)		68		
執筆要項		69		
スタッフ便り		70		