

## SX-9による大規模並列シミュレーション

曾我 隆<sup>1,2</sup> 下村 陽一<sup>3</sup> 撫佐 昭裕<sup>1,4</sup> 江川 隆輔<sup>5</sup> 滝沢 寛之<sup>1</sup>  
岡部 公起<sup>5</sup> 小林 広明<sup>5</sup> 高橋 俊<sup>6</sup> 中橋 和博<sup>6</sup>

1 東北大学大学院情報科学研究科

2 NEC システムテクノロジー

3 NEC ソフトウェア東北

4 日本電気株式会社

5 東北大学サイバーサイエンスセンター

6 東北大学大学院工学研究科

あらまし NEC SX-9 は、Addressable Data Buffer(ADB) と呼ばれるソフトウェアから制御可能なオンチップキャッシュメモリを具備するプロセッサを搭載する新世代ベクトルスーパーコンピュータである。本稿では、次世代の流体計算アルゴリズムである Building Cube 法に基づく大規模シミュレーションを実行し、実アプリケーションに対する SX-9 の実効性能を評価する。本評価により、ADB やスケーラビリティを高めるための最適化技法が実効性能に与える影響が明らかになった。また、本アプリケーションの開発マシンであるスカラ並列システム TX7/i9610 では 36 日間必要であったシミュレーション実行時間を、SX-9 の 16 ノードでは約半日に短縮できることを明らかにした。

## Large Scale Simulation using SX-9 Supercomputer System

Takashi SOGA<sup>1,2</sup>, Youichi SIMOMURA<sup>3</sup>, Akihiro MUSA<sup>1,4</sup>, Ryusuke EGAWA<sup>5</sup>,  
Hiroyuki TAKIZAWA<sup>1</sup>, Koki OKABE<sup>5</sup>, Hiroaki KOBAYASHI<sup>5</sup>, Shun TAKAHASHI<sup>6</sup>, and  
Kazuhiro NAKAHASHI<sup>6</sup>

1 Graduate School of Information Sciences, Tohoku University

2 NEC System Technologies

3 NEC Software Tohoku

4 NEC Corporation

5 Cyberscience Center, Tohoku University

6 Graduate School of Engineering, Tohoku University

**Abstract** NEC SX-9 is a new-generation vector parallel supercomputer whose processor has a software-controllable on-chip cache memory, called an addressable data buffer (ADB). In this paper, the sustained performance of an SX-9 system for real applications is evaluated using a large-scale CFD simulation based on a next-generation algorithm named the Building Cube method. The evaluation results quantitatively clarify the effects of ADB and some scalability optimization techniques on sustained performance. The scalar parallel system, TX7/i9610, needs 36 days to execute the CFD simulation, even though it was originally developed for the system. On the other hand, a 16-node SX-9 system requires only half a day to complete it.


**SX-9による大規模並列シミュレーション**  
 平成21年 2月13日  
 曾我 隆(東北大学大学院情報科学研究科/NECシステムテクノロジー)  
 下村陽一(NECソフトウェア東北)  
 撫佐昭裕(東北大学大学院情報科学研究科/日本電気株式会社)  
 江川隆輔, 滝沢寛之, 岡部公起, 小林広明(東北大学サイバーサイエンスセンター)  
 高橋 後, 中橋和博(東北大学大学院工学研究科)


**Outline**  
 • 背景と目的  
 • 評価システム  
 • 評価アプリケーション  
 • 最適化  
 • 測定結果  
 • まとめ

No.2

**NEC**


**背景**  
 • 2008年3月, SX-9導入・稼働  
 • SX-9は、総演算性能26.2TFlop/sを有する大規模ベクトル並列システム

No.3

**NEC**


**目的**  
 • SX-9を利用した、ユーザの実アプリケーションを用いた大規模シミュレーションの実施  
 • SX-9による大規模シミュレーションの手段と可能性について検証

No.4

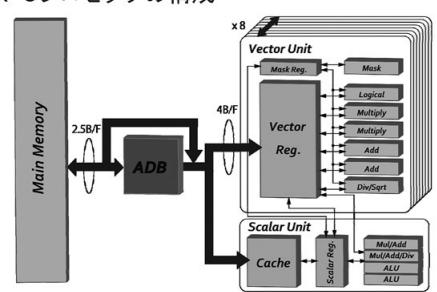
**NEC**


**評価システム**  
 • 評価システム概要  
 SX-9 16ノード(256CPU)  
 総合ベクトル演算性能 : 26.2TFlop/s  
 (1 CPUの演算性能は102.4GFlop/s)  
 メモリ容量 : 16TB  
 (ノードあたり1TB, メモリバンド幅4TB/s)  
 ノード間ネットワーク性能: 256GB/s(双方向)

No.5

**NEC**


**評価システム**  
 • SX-9プロセッサの構成



Main Memory  
 ADB  
 Vector Unit (x8): Mask Req., Mask, Logical, Multiply, Add, Div/Sqrt  
 Vector Reg.  
 Scalar Unit: Cache, Scalar Reg., Scalar ALU, MultiAdd, MultiAddInv, ALU, ALU  
 Note: B/F = メモリ/バンド幅 (GB/s) / 浮動小数点演算 (GFlop/s)

No.6

**NEC**

**評価システム**

### ADBの効果の検証

→再利用性のあるデータをADBの載せることでメモリからのアクセス時間を削減することが可能になる

**差分式の例**

Relative performance

ADB Status	Relative Performance
ADB Off	~0.8
ADB On	~1.2

約15%の実効性能向上  
(演算で隠蔽できないメモリアクセス時間37%削減)

No.7

```

do 10 k=0,Nz
do 10 i=0,Nx
  ldir on_<db>(i,k)
  ldir on_<db>(i,k)
  ldir on_<db>(i,k)
  do 10 j=0,Ny
    E_x(i,j,k)=
    & C_x,i,j,k)+E_x(i,j,k)
    & -C_x,j,i,j,k)*(H_x(i,j,k)-H_x(i,j-1,k))/dy
    & -(0,y(i,j,k)-U_y(i,j,k-1))/dz-E_x,Current(i,j,k)
    E_x,j,i,j,k)=
    & C_y,i,j,k)+E_y(i,j,k)
    & -C_y,j,i,j,k)*(H_y(i,j,k)-H_y(i-1,j,k))/dz
    & -(0,z(i,j,k)-U_z(i-1,j,k))/dx-E_y,Current(i,j,k)
    E_z(i,j,k)=
    & C_z,i,j,k)+E_z(i,j,k)
    & -C_z,j,i,j,k)*(H_z(i,j,k)-H_z(i,j-1,k))/dx
    & -(0,x(i,j,k)-U_x(i,j-1,k))/dy-E_z,Current(i,j,k)
  10 CONTINUE
  
```

NEC

**評価アプリケーション**

### 評価アプリケーション

Building-Cube法による非圧縮性流体解析コード

- Building-Cube法とは
  - 計算空間に様々な大きさの立体Cubeを積み上げる
  - Cube内に等間隔直交格子Cellを内包
  - Cellを用いることで、高精度化とアルゴリズムの単純さを確保

No.8

計算格子の全体像  
物体の密度状表現  
Cubeの分布図  
Cellの分布図

NEC

**評価アプリケーション**

### 評価モデル: レーシングカーの風の流れ

Cube数: 5,930  
CubeあたりのCell数: 32x32x32 (=32,768)  
総格子数: 約2億格子  
計算ステップ: 20

TX7/i9610 1ノード(64コア)を使用した  
20ステップの実行時間約1,230秒

No.9

NEC

**評価アプリケーション**

### 非圧縮性Navier-Stokes計算

```

graph TD
    A[計算格子の読み込み・初期条件の設定] --> B[仮の速度場 u'を計算]
    B --> C[Cube 間で u'を情報交換]
    C --> D[圧力場 p を計算]
    D --> E[Cube 間で p を情報交換]
    E --> F[真の速度場 u を計算]
    F --> G[Cube 間で u を情報交換]
    G --> H[タイムステップの更新]
    H --> I[空力係数の計算]
    I --> A
    
```

Dual time-stepping  
熱解法の場合には内部反復  
SOR 法による反復計算

No.10

NEC

**SX-9のための最適化**

### 本アプリケーションはTX7/i9610向けに作成される のでSX-9向けの最適化が必要

- ベクトルチューニング  
→SX-9の高いベクトル性能を演算性能を得るために、ベクトル演算率の向上、平均ベクトル長の拡大が必要
- 分散並列化  
→OpenMPでノード内並列のみ行っていたが、複数のノード間の並列化を行ないのでMPI(Message Passing Interface)による分散並列化が必要

No.11

NEC

**最適化 -ベクトルチューニング-**

- 最も計算コストが大きいSOR法部分にベクトル化を阻害する依存関係あり→平面法を用いてベクトル化(ベクトル演算率99.8%)

平面法

j  
6 7 8 9 10 11  
5 6 7 8 9 10  
4 6 6 7 8 9  
3 4 5 6 7 8  
2 3 4 5 6 7  
1 2 3 4 5 6  
i

② 平均ベクトル長が約32(Cellの各次元の大きさ)→ループ融合によりベクトル長の拡大(平均194.6)

No.12

NEC



## 最適化 -ベクトルチューニング-

- ③ SX-9の新しい機能であるADBの適用を検証  
→平面法の間接参照を行うアドレス・テーブルに対してADBの適用を検討。テーブルのサイズが512KBあり、全体をADBに載せることが出来ず、効果は数%程度

No.13

NEC



## 最適化 -MPIによる分散並列化-

- MPI(Message Passing Interface)による分散並列化  
→できるだけデータ転送回数と転送量を小さくする  
→各プロセスに割り当てられる演算量を均一に
- I. 最適なデータ分割を行う方向の選択  
cube単位に計算の独立性、均一な演算量
  - II. 最適なデータ転送方法の選択  
単方向通信(MPI\_GET)により、必要なデータのみ転送

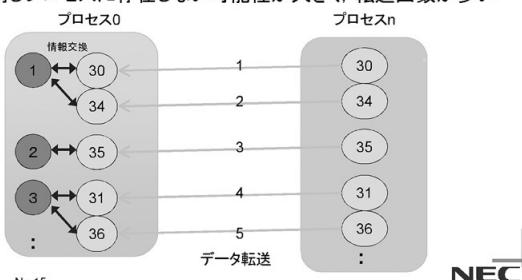
No.14

NEC



## 最適化 -MPIによる分散並列化-

- ・データ転送回数の問題  
Cubeが座標値と関係なく並べられており、隣接するCubeが同じプロセスに存在しない可能性が大きく、転送回数が多い



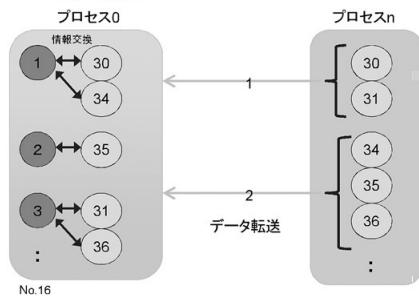
No.15

NEC



## 最適化 -MPIによる分散並列化-

- 対策1...連続するCubeを一度に転送することで転送回数の削減



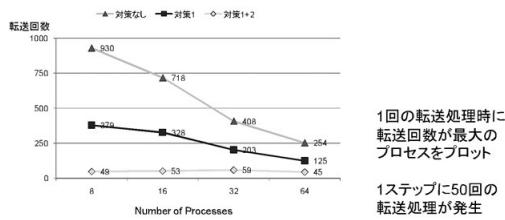
No.16

NEC



## 最適化 -MPIによる分散並列化-

- 対策2 ...更に入力データを座標値で並び替え、Cubeの局所性を向上



No.17

NEC



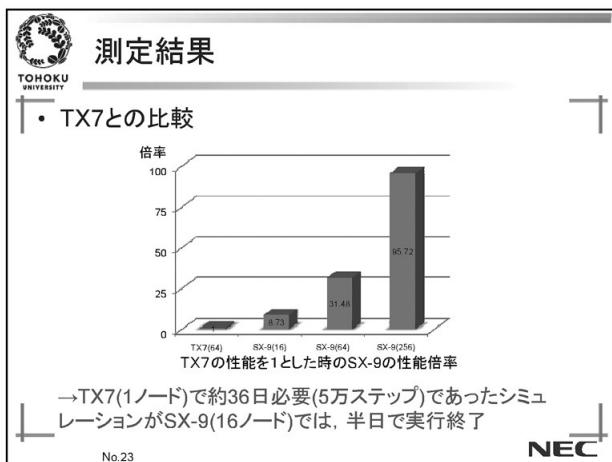
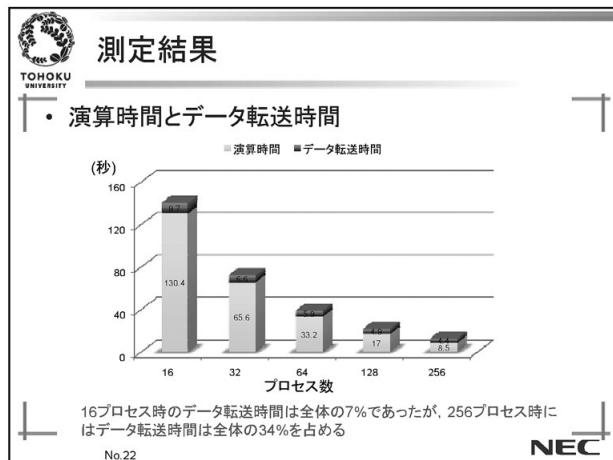
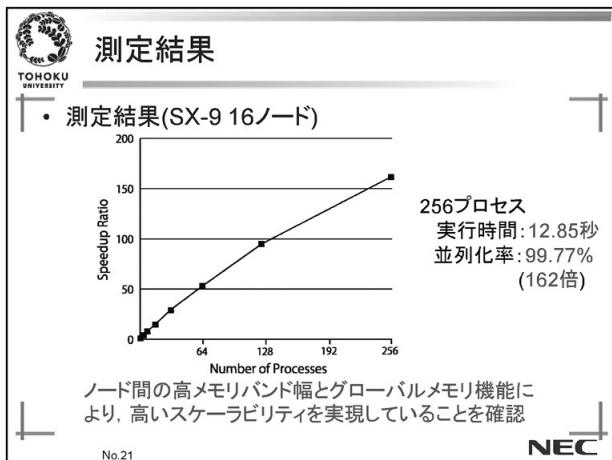
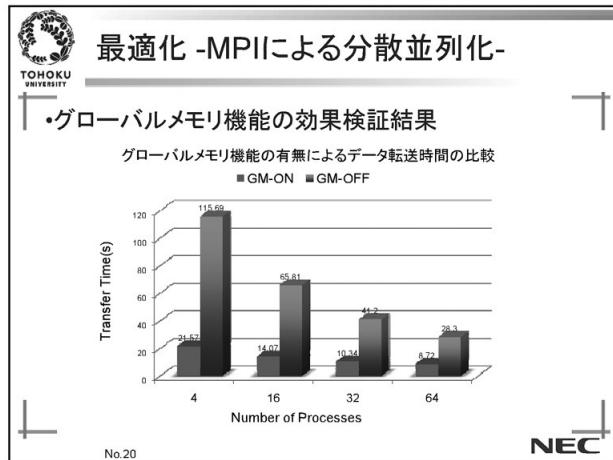
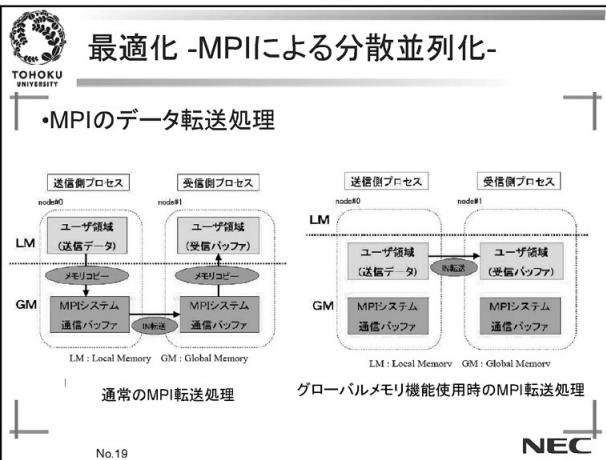
## 最適化 -MPIによる分散並列化-

- ・グローバルメモリ機能

- 1) MPIの通信バッファへのコピーを行わず、グローバルメモリから直接データ転送を行う
- 2) 条件: 転送に使用する配列の定義がallocatableである  
-gmallocオプションを使用してオブジェクトの作成

No.18

NEC



**まとめ**

- SX-9の高いベクトル性能を発揮するためには、ベクトル化率の向上と平均ベクトル長の拡大が必要
- 分散並列化では、最適なデータ分割とデータ転送方法の選択が重要
  - ユーザによる最適化
- グローバルメモリ機能により、データ転送時間の削減が可能
- 高いネットワーク bandwidth(ノード内/ノード間)により、高いスケーラビリティを達成
  - 実アプリケーションにおいて、大規模シミュレーションを可能にすることが示された

No.24

NEC



## まとめ

### 今後の課題

1)ノード内のOpenMP並列とのハイブリッド並列によるスケーラビリティ向上の研究

- ・ハイブリッド並列化により、使用メモリの削減
- ・転送回数／量の削減による並列効率の向上

2)他の大規模シミュレーションへの適用

- ・実現不可能であった大規模シミュレーションを可能に

No.25

NEC