

東北大学サイバーサイエンスセンターの取り組みとSX-9の性能評価

小林 広明・江川 隆輔・岡部 公起
伊藤 英一・大泉 健治

要 旨

東北大学サイバーサイエンスセンターは、全国共同利用施設として、高性能計算を行えるスーパーコンピュータシステムの整備・運用を行うとともに、先端学術情報基盤の研究開発を行っています。本センターでは、スーパーコンピュータの性能を最大限に引き出すため、利用者に対するプログラム高速化支援と利用者との高速化に関する共同研究を行っています。本稿では、2008年3月に稼働を開始したSX-9の中核とした大規模科学計算システムと高速化に関する支援活動について紹介を行い、更に、SX-9の性能について実シミュレーションプログラムを用いて評価した結果を示します。

キーワード

●スーパーコンピュータ ●ベクトル型 ●大規模科学計算システム ●高速ネットワーク
●大規模シミュレーション ●ADB ●Land Mine ●Antenna ●クロスバースイッチ ●Building Cube Method

1. まえがき

2008年は、東北大学とNECとの共同研究で開発を行ったパラメトロン式電子計算機SENAC-1 (NEAC1102) が完成してから50年を迎える記念の年です。そして、この記念すべき年に最新鋭のスーパーコンピュータシステムSX-9の1号機を導入し、運用を開始することになりました。これは、SENAC-1開発時に培われた、本学研究者とNEC技術者との共同研究の歩みがあったからこそ迎えることができたと感じています。今後も継続的に共同研究を行い、世界をリードするシステムの研究開発を行っていきたくと考えています。

東北大学サイバーサイエンスセンターは、SENAC-1の開発精神を引き継ぎ、1969年に大型計算機センターとして設置され、その後、2001年に情報シナジーセンター、そして2008年4月にサイバーサイエンスセンターと発展的に改組されてきました。本センターの目的は、全国共同利用施設として大規模科学計算システムや高速ネットワークなど先端学術情報基盤の整備・運用を行うとともに、これら先端学術情報基盤を活用した新しい科学の創造に関する教育・研究を推進することにあります。そして世界最高速のスーパーコンピュータシステムを導入・運用し、国内の学術研究者及び学生に高速な計算資源を提供することにより、最先端の学術研究を支援・推進しています。更に、システム利用者との共同研究を通して、

スーパーコンピュータシステムの性能を最大限に引き出すための、プログラムの高速化・並列化及びシミュレーションの高度化に関する研究開発を行っています。

本稿では、本センターの大規模科学計算システムとプログラム高速化支援活動を紹介するとともに、2008年3月に稼働を開始したSX-9の性能評価について説明を行います。そして最後に利用者との共同研究によるSX-9・16ノードでの大規模シミュレーションを紹介します。

2. 大規模科学計算システム

本センターでは、一般のサーバやPCでは実現不可能な大規模シミュレーションを実現するため、大規模科学計算システムを導入・運用しています。図1が現在運用を行っているシステムの概念図です。2008年3月に主力計算システムであったベクトル型スーパーコンピュータSX-7を、最新鋭のSX-9にリプレースしました。SX-9の導入に伴い、システム構成は、ベクトル型スーパーコンピュータSX-9・16ノード、SX-7C (SX-8相当)・5ノード、そしてスカラー型コンピュータTX7/i9610・3ノードから成り、総演算能力は28TFLOPSに達しています。

本センターの大規模科学計算システムの特徴の1つとして、ベクトル型とスカラー型の2つの方式のコンピュータを導入していることが挙げられます。ベクトル型スーパーコンピュー

タは、高いメモリバンド幅と効率よく動作するベクトル演算パイプラインを備え、更にノード間を高速クロスバースイッチによって結合することにより、最先端科学分野における大規模・高精度のシミュレーションを高効率に実行することが可能になっています。また、ベクトル型スーパーコンピュータを必要としないプログラムの実行や汎用アプリケーションのサービスのために、スカラー型コンピュータを備え、学術研究者が求める様々な演算要求に対応できる計算環境を実現しています。

現在、本システムは、学内外1,500名の学術研究者に利用さ

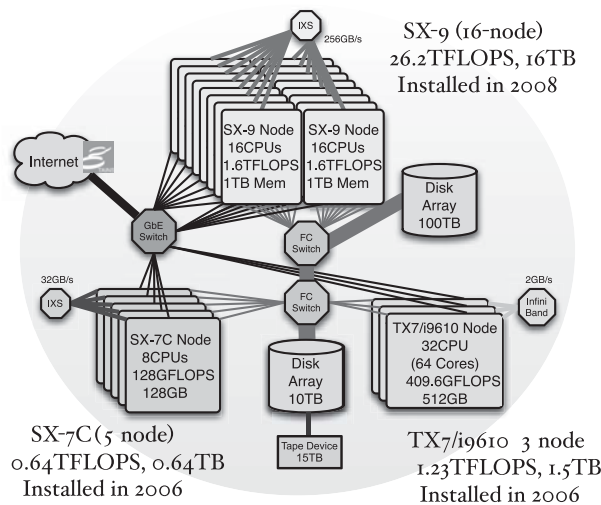


図1 大規模科学計算システム概念図

れ、図2に示した多岐にわたる科学技術分野のシミュレーションが実行されています。

3. 高速化支援活動

スーパーコンピュータシステムの性能を最大限に引き出すためには、スーパーコンピュータに関する専門的知識を用いてプログラムを色々な角度から解析を行い、最適化する必要があります。利用者のプログラムに対する高速化支援は、高性能なスーパーコンピュータを整備・運用することと並んで、本センターにおける重要なサービスの1つとなっています。本センターでは、図3に示すように、センターの教員、技術職員、メーカーの技術者が一体となり、利用者に対するプログラムの高速化支援体制を整え、利用者プログラムの解析、プログラム最適化のアドバイス、高速化技術の蓄積、講習会の実施や出版などによる高速化技術の利用者への還元に取り組んでいます。

毎年、本センターは大規模科学計算システムの利用者と共同研究を行い、プログラムの高度化、ベクトル化及び並列化を行っています。2007年度は、学内外の研究者と11件の共同研究を行い、SX-9での大規模シミュレーションを実現するための研究開発を行いました。その成果の一例を第5章で紹介します。

また、利用者プログラムの解析・高速化を通して得た技術や研究開発成果は、利用者、センターの教員、技術職員と共有し、図4に示すような研究・開発サイクルを通して、研究成果・知見を次期システムの設計・開発へと還元させ、常に

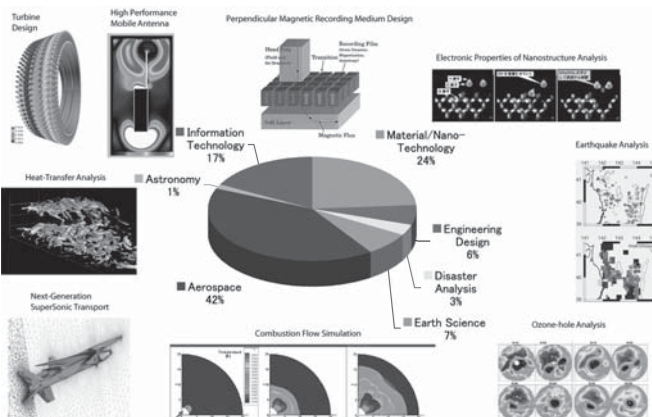


図2 科学技術分野における利用状況

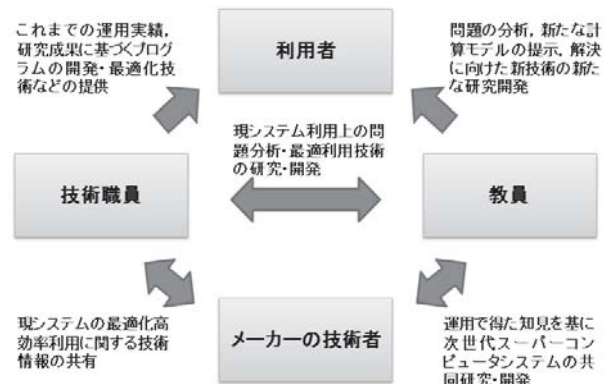


図3 利用者プログラムの解析・高速化支援体制

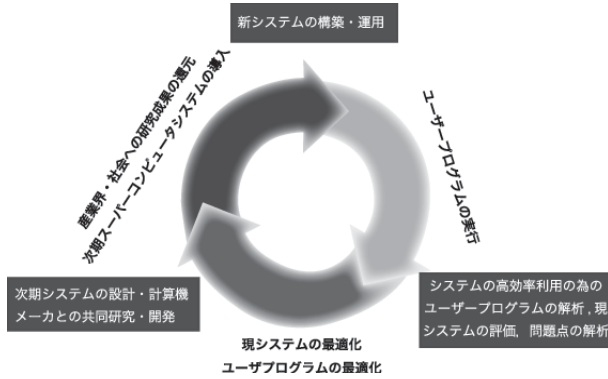


図4 システムの導入・整備・研究・開発サイクル

高性能・高効率なスーパーコンピュータの整備・運用へとつなげています。

4. SX-9の性能

2008年3月に導入した新システムSX-9の性能評価について紹介します。評価に利用したプログラムは、大規模科学計算システム利用者のプログラム6本で、図2における地球科学、航空宇宙、エンジニアリングデザイン、宇宙物理の代表的なプログラムです。そして、SX-9、SX-8、SX-8R、SX-7及びTX7/i9610との性能比較を行いました。

表1に各システムの性能諸元を示します。この表において、B/flopはメモリバンド幅を演算性能で割った値であり、CPUへの演算データの供給割合を表しています。本センタースーパーコンピュータ研究部では、B/flopと演算性能の定量的な研究を行っており、4 B/flopがCPU性能を維持するために必要な値であることを明らかにしています²⁾。また、SX-8RやSX-9のような2B/flop程度のシステムでは、SX-9で導入されたADB (Assigned Data Buffer) のようなベクトル演算用キャッシュ機構が実効性能を向上させるのに有効であることも示しています。本章では、各システムの1 CPU (1コア) での性能とノード内の並列性能について比較を行い、次にSX-9の新機能であるADBによる性能向上の例を紹介します。

4.1 CPU性能の評価

図5は、各システムの1 CPU (1コア) での性能値を示しています。ここでは、ADBの機能は利用していません。SX-9は、

表1 システム諸元

システム名	CPU数 (コア数)	クロック 周波数 (GHz)	CPU 当たり			
			演算性能 (GFLOPS)	メモリバンド幅 (GB/s)	B/flop	キャッシュ ADB
SX-9	16	3.2	102.4	256	2.5	256KB
SX-8R	8	2.2	35.2	70.4	2	
SX-8	8	2	16	64	4	
SX-7	32	1.1	8.83	35.3	4	
TX7/i9610	32 (64)	1.6	12.8 (6.4)	8.5 (4.25)	0.66	L1: 16KB L2: 256KB L3: 12MB

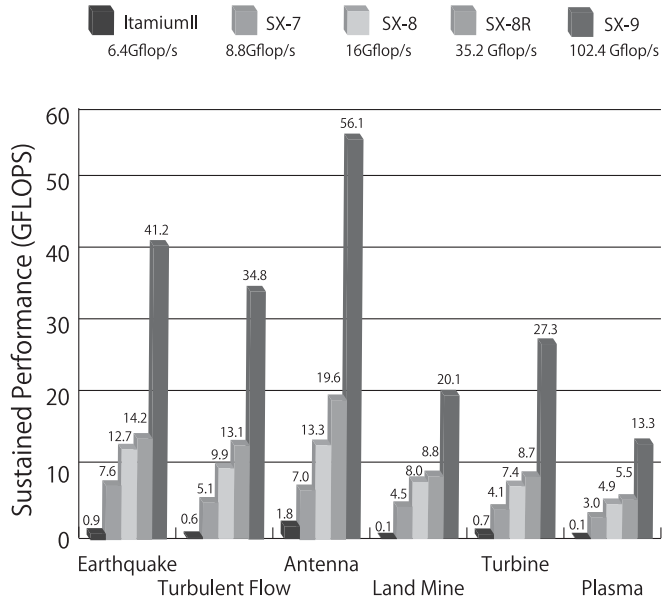


図5 1CPUあたりの性能

すべてのプログラムにおいて他のシステムに比べ、絶対性能が高いことが分かります。特に、スカラー型コンピュータであるTX7との性能比較では、SX-9は30倍から155倍の性能になっています。TX7とSX-9の理論性能比は16倍ですから、すべてのプログラムにおいてTX7よりSX-9の方がプログラムを効率的に実行していることが分かります。この中で155倍の性能比があったプログラムはLand Mineです。このプログラムはデータの局所性が低く、TX7でのキャッシュヒット率は77%程度でした。このため、TX7では130 MFLOPS程度の性能しかならず、SX-9との性能比が155倍となっています。TX7のようなスカラー型コンピュータでは、SXよりメモリバンド幅が低い場合、キャッシュを十分に利用できない場合には、SXとは

大きな性能差が発生します。また、データの局所性が高いAntennaでは、TX7のキャッシュヒット率が99.4%であるにもかかわらず実効性能は1.8 GFLOPS程度であり、SX-9との実行性能比は30倍にもなりました。キャッシュを十分に利用できるプログラムの場合でも、メモリバンド幅が高いSX-9の方が効率的にプログラムを実行できることが分かります。

一方、SX-7とSX-8に対するSX-9のプログラム性能比をみると、対SX-7で4.0倍から8.0倍、対SX-8で2.5倍から4.2倍になっています。これは、SX-9のB/flopがSX-7とSX-8の4から2.5と低くなり、各プログラムの演算密度（演算命令数÷メモリアクセス命令数）によるメモリアクセス頻度によって実効性能が大きく変わることと、各プログラムがSX-8Rから追加された2本目の加算パイプラインと乗算パイプラインを効率よく利用できるかによっています。例えばAntennaは、SX-9で56.1 GFLOPSと高い実効性能を示しています。このプログラムは演算密度が2.3と高く、メモリアクセスが少ないプログラムです。そのため、B/flopが2.5の場合と4の場合とで、ほとんど性能への影響はありません。しかし、このプログラムでは、ループ内の演算の関係で2本目のパイプラインを利用しない場合があり、そのため実行効率が54.8%にとどまっています（SX-7の場合、実行効率は79%になっています）。また、Plasmaにおいては、SX-9の実効性能が13.3GFLOPSと6本のプログラムの中で一番低い性能になっています。これは、このプログラムがリストベクトルを多用しており、ベクトルgather命令のメモリレーテンシーが陽に見えてしまい、実行時間が延びてしまっているからです。その他のプログラムについては演算密度が1.0前後でメモリアクセスが多く、SX-9の性能はB/flopが2.5である影響とパイプラインの利用の仕方によって実効性能が決まっています。

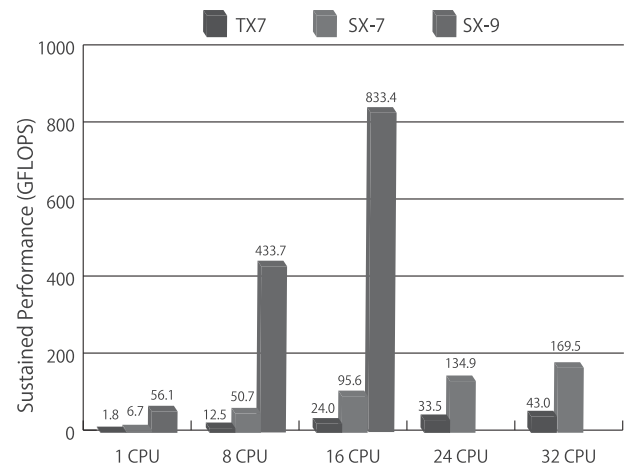
SX-9では、メモリアクセスの仕方や追加されたパイプラインの利用によって、今までのSXとは異なった実行効率になります。しかしながら、SX-9はこれらのプログラムにおいて、1 CPU当たり10 GFLOPSを超える高い実効性能を得ることができ、本センターでは、今まで以上に高速なベクトル演算環境をユーザに提供できるシステムだと考えています。

4.2 ノード性能の評価

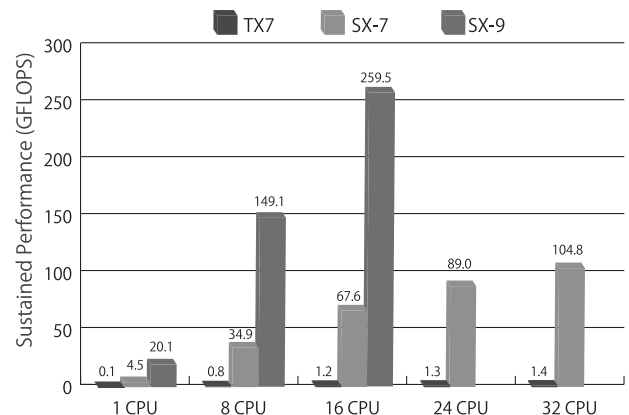
ノード性能の評価としてTX7、SX-7及びSX-9を使用し、プログラムの並列性能を比較しました。TX7は、CPU当たり2つ

のコアを有したシステムですが、本評価ではCPU当たり1つのコアのみを使用し、32CPU（32コア）までの計測を行いました。ここではメモリアクセスが少ないAntennaと多いLand Mineの結果について紹介します。プログラムの並列化には、各システムのFortranコンパイラの自動並列化機能を利用しました。

図6にAntennaとLand Mineの並列性能を示します。AntennaはSX-9・16CPUで833GFLOPSの性能に達しています。これは、TX7・1CPUの460倍、SX-7・1CPUの124倍、また、TX7・32コアの19倍、SX-7・32CPUの5倍の性能に当たります。またLand Mineでは、SX-9・16CPUで260 GFLOPSの性能であり、Antennaに比べ低い性能になっています。これは、1CPU



(a) Antenna



(b) Land Mine

図6 並列性能

における性能がAntennaの約半分であり、また、後ほど説明するスケラビリティの関係でこの性能になっています。しかしながら、SX-9・16CPUはTX7・1CPUの2,000倍、SX-7・1CPUの58倍、また、TX7・32コアの187倍、SX-7・32CPUの2.5倍の性能に当たります。このようにSX-9は、前システムのSX-7より半分のCPU数になっていますが、SX-9の1CPUの性能が高いことによって、SX-7のノード性能より高い性能になっていることが分かります。

また、並列化による性能向上の割合については、メモリアクセスの少ないAntennaはTX7、SX-7及びSX-9共に、ほぼアムダールの法則に従った向上を示しています。しかし、Land MineはCPUの増加に伴って性能向上の割合が低下しています。図7にLand Mineの並列化による性能向上の割合を示します。ここでは、各システムの1CPUの性能を1にして正規化しています。特にTX7では、16CPU以上でほとんど性能向上がありません。これは、Land Mineがメモリアクセスの多いプログラムのためです。TX7では4つのCPUを1つのセルボードに実装し、各セルボードをクロスバースイッチで接続する構成になっています。TX7の性能向上の割合の低下は、セルボードに跨ったデータ転送を各CPUが同時に行い、データ転送が律速になったためです。また、SX-7とSX-9においては、それぞれSX-7では16CPU、SX-9では8CPUまでアムダールの法則に従った性能向上を示しますが、それ以上のCPUでは性能向上の割合が低下しています。これは、各CPUからのメモリアクセスによりバンクコンフリクトが増加し、メモリアクセス時間が延びたことによります。このようにSX-9は、従来のSX

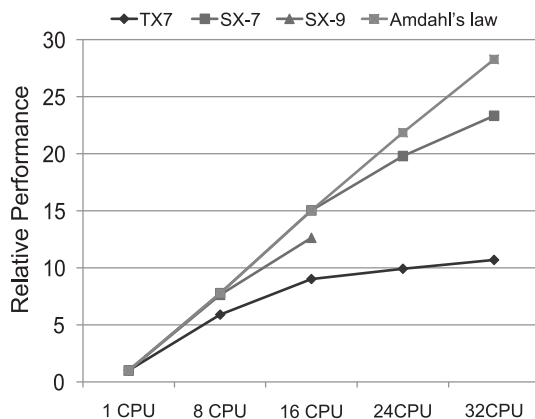


図7 Land Mineの性能向上の割合

シリーズと同様にメモリ負荷が並列性能に影響する傾向を有しています。しかしながら、図6で示すようにSX-9は従来から利用している自動並列化機能によって、前システムSX-7では得られなかった高いノード性能を実現できるシステムとなっています。

4.3 新機能ADBの評価

ADBは、SX-9のベクトルプロセッサに新たに導入された256 KBのオンチップメモリです。このADBを利用することによって、実効性能を向上させることができることを示します。SX-9ではB/flopが2.5になっていますが、ADBとメモリとの両方から演算データを供給することによって、B/flopが4相当になり実効性能が向上します。ここでは、EarthquakeとLand Mineの例を示します。

図8はEarthquakeの主ループになります。コンパイラ指示行ON_ADBで配列を指定することにより、指定された配列だけがADBをキャッシュのように利用することができます。図8のループはインデックス j でベクトル化され、配列 gd_dip と $wary$ がベクトルロードされます。配列 $wary$ の容量は250 KBで、すべてのデータをADBに載せることが可能です。このことによって、配列 gd_dip がメモリから、配列 $wary$ がADBから供給されることになり、ベクトル演算パイプラインに対して4 B/flopのデータ供給率になります。図9がADBを利用した場合の性能向上比ですが、ADBを利用しない場合より実効性能が1.7倍高くなります。

次にLand Mineの例を示します。図10がLand Mineの主ループになります。このループはFDTD法のカーネル部でほとんどのループが同じ構造になっています。このプログラムでは、インデックス i のループ長が50と小さいため、コンパイラはインデックス j のループをベクトル化します。評価に用いたモデルでは各配列の大きさは858 MBで、256 KBのADBにはすべてのデータを載せることはできません。しかし、このループは

```
do i=1,ncells
!CDIR ON_ADB(wary)
do j=1,ncells
wf_dip(i)=wf_dip(i)+gd_dip(j,i)*wary(j)
end do
end do
```

図8 Earthquakeの主ループ

差分式になっていますので、例えば、図10の下から3行目の $H_y(i, j, k) - H_y(i-1, j, k)$ に着目すると、 $H_y(i, j, k)$ の

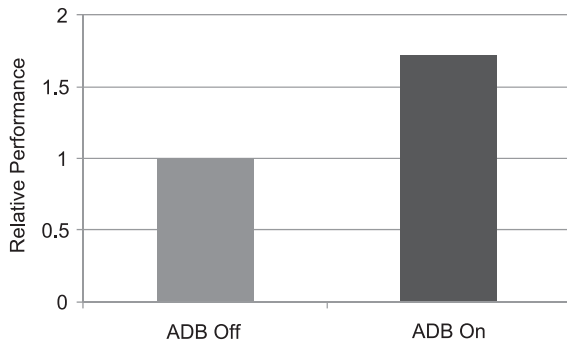


図9 ADBによる性能比 (Earthquake)

```

DO 10 k=0,Nz
DO 10 i=0,Nx
!CDIR ON_ADB(H_y,H_z)
DO 10 j=0,Ny
  E_x(i,j,k) = C_x_a(i,j,k)*E_x(i,j,k)
  & + C_x_b(i,j,k) * ((H_z(i,j,k)-H_z(i,j-1,k))/dy
  & -(H_y(i,j,k)-H_y(i,j,k-1))/dz -E_x_Current(i,j,k)
  E_y(i,j,k) = C_y_a(i,j,k)*E_y(i,j,k)
  & + C_y_b(i,j,k) * ((H_x(i,j,k)-H_x(i,j,k-1) )/dz
  & -(H_z(i,j,k)-H_z(i-1,j,k))/dx -E_y_Current(i,j,k)
  E_z(i,j,k) = C_z_a(i,j,k)*E_z(i,j,k)
  & + C_z_b(i,j,k) * ((H_y(i,j,k)-H_y(i-1,j,k) )/dx
  & -(H_x(i,j,k)-H_x(i,j-1,k) )/dy -E_z_Current(i,j,k)
10 CONTINUE
    
```

図10 Land Mineの主ループ

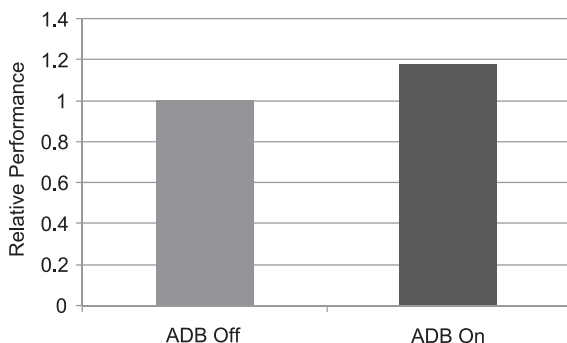


図11 ADBによる性能比 (Land Mine)

データを次の $i+1$ の $H_y(i-1, j, k)$ が再利用することになります。図10の下から5行目の $H_z(i, j, k) - H_z(i-1, j, k)$ も再利用性のある部分で、再利用に必要なADBの容量は24 KBになります。配列 H_y と H_z だけをADBに載せるように指示行で指定した結果が図11です。この場合の性能向上は約1.18倍になり、ADBの利用により1CPUの実効性能が24 GFLOPSになりました。

ADBは256 KBと容量の少ないことが残念ですが、利用の仕方によって、プログラムの性能を大きく向上させる機能であることを示しました。ここに紹介した利用方法以外にリストベクトルのテーブルをADBに載せるなど色々な活用方法が考えられます。本センターでは、SX-9向け高速化としてADBの利用方法と追加されたパイプラインの効率的な利用について今後も研究を行っていく予定です。

5. 大規模シミュレーション

最後にSX-9による大規模シミュレーションの例を紹介しましょう。本センターは、SX-9・16ノードのシステムを導入するに当たり、256個のベクトルプロセッサを効率よく並列動作させるシミュレーションプログラムの開発を、東北大学大学院工学研究科の中橋研究室との共同研究で行いました。

中橋研究室は、スーパーコンピュータを用いて様々な流れの問題を解くための数値流体力学(CFD)の計算手法や、数値流体力学に基づく空力設計手法及び数値最適化アルゴリズムの空力設計への応用について研究を行っています。そして、中橋研究室では、次世代CFDアルゴリズムとして、Building Cube Methodを考案、提唱しています^{3), 4)}。Building Cube Methodは、

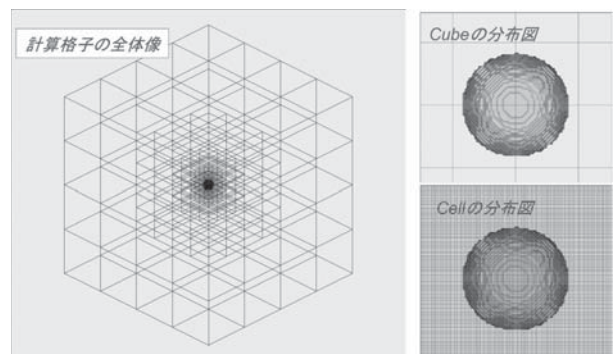


図12 空間格子の例

図12に示したように計算空間に様々なサイズの立方体ブロック（キューブ）を積み上げ、個々のキューブ内には高密度な等間隔直交格子（セル）を用いる計算法で、流れの局所性に適合した格子密度を可能にしています。またキューブ内に等間隔直交格子を用いることで高精度化を可能にするとともに、アルゴリズムの単純さを確保できる利点を有しています。

本共同研究では、中橋研究室の教員・大学院生、本センタースーパーコンピュータ研究部の教員・大学院生、システム管理系の技術職員並びにNECの技術者による、検討会を数回にわたり開催し、非圧縮性流体におけるシミュレーションプログラムの特性解析と並列化の検討を行いました。シミュレーションプログラムは、スタガード格子におけるフラクショナルステップ法をベースに、対流項を3次精度風上差分近似、粘性項を2次精度中心差分近似、時間積分を2次精度陽解法で非圧縮性Navier-Stokes方程式を解き、また、圧力のPoisson方程式については2次精度中心差分を利用し、SOR法を用いて解くものでした。

検討の結果、Building Cube Methodは高い並列性を持つ手法であり、プロセッサごとに複数のキューブを配置することにより、各プロセッサでの並列処理の粒度とロードバランスを調整できることが判明しました。プログラムの並列化では、MPI（Message Passing Interface）ライブラリによるデータ転送の効率化を行うため、SX-9のグローバルメモリ機能⁵⁾を利用し、データ転送の関数としてはMPI_GET、MPI_PUTを採用しました。また、ベクトル化の観点では、非圧縮性流体の計算で最もコストのかかるPoisson方程式の解法に平面法を導入し、SOR法における反復計算のベクトル化を行いました。

性能測定には、2億個のセルを5,930個のキューブを利用して、レーシングカー周辺の風の流れを再現し、表面の圧力分布を求めました。図13に可視化結果を示します。運転席とリアウイングの後方の暗い部分が圧力の低下を示しています。また、圧力の上昇は車体とリアウイングの前面の白い部分に起きています。シミュレーションの精度を上げるため、レーシングカーの近傍は細かいキューブを利用しています。表2にTX7/i9610とSX-9の実効性能比を示します。図13のシミュレーションにおいて、SX-9・256CPUで50分の実行時間を要しました。これらの結果から同じシミュレーションをTX7・1CPUで行った場合には128日かかりますが、SX-9・1CPUでは6日で完了できることが分かります。今回の共同研究により、今まで以上に大規模なシミュレーションを短時間で実行できるように

なりました。

また、図14にSX-9のグローバルメモリ機能を利用した場合と利用しない場合の並列性能を示しました。グローバルメモリ機能を利用した場合は、SX-9・256CPUで130倍の並列性能が得られていますが、利用しない場合には、100CPU以上で性能向上がなくなり、並列性能は40にとどまっています。このことから、SX-9のグローバルメモリ機能はMPIによるデータ転送のオーバーヘッドを抑えることができ、マルチノード

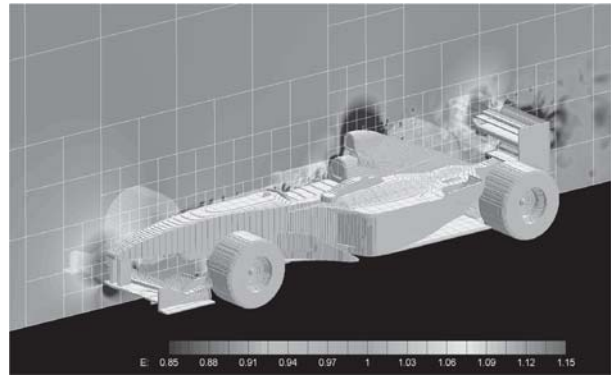


図13 Building Cube Methodによるシミュレーション結果

表2 性能比

システム名	TX7		SX-9		
	1コア	64コア	1CPU	16CPU	256CPU
CPU(コア)数	1コア	64コア	1CPU	16CPU	256CPU
理論性能(FLOPS)	6.4 G	409.6 G	102.4 G	1.6 T	26 T
性能比	1	36	20	285	3700

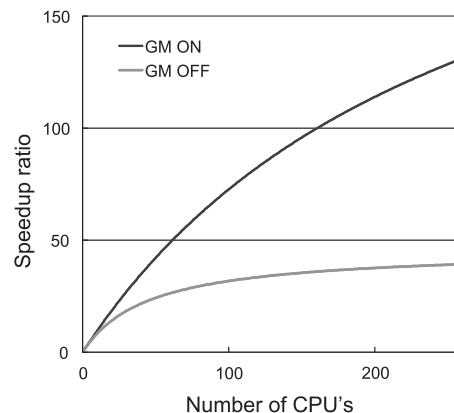


図14 SX-9グローバルメモリ機能の性能

環境の大規模シミュレーションには必要な機能であることが分かります。

6. むすび

本センターは、1986年にスーパーコンピュータSX-1を導入して以来、ベクトル型スーパーコンピュータの運用とその効率的利用、並びにハードウェア・ソフトウェアの機能強化に関する研究を行ってきました。本年、SX-9を本センターに導入し、今まで以上に大規模で効率的なシミュレーションを実行できる環境を整えることができました。本センターでは、国内の学術研究者の方々にSX-9の新しい機能や16ノード26.2 TFLOPSの性能を引き出すための支援活動を継続して行っていく予定です。また、次世代スーパーコンピュータに代表される将来のスーパーコンピュータにおいて直面している、消費電力、発熱量、メモリ性能などの問題について、本センターは、チップ積層化技術、メモリアーキテクチャ及び並列化技術の研究を通して貢献をしていきたいと考えています。

本センターの活動により、日本のコンピュータ技術並びにシミュレーション技術が、よりいっそうの飛躍的な発展を遂げることを心から願っています。

なお、本稿を執筆するにあたり、多くの方々にご協力・ご支援を賜りました。東北大学の長谷川昭名誉教授、澤谷邦男教授、佐藤源之教授、山本悟教授、飯島雅英准教授、東京理科大学の塚原隆裕博士、海洋研究開発機構の有吉慶介博士には評価プログラムを提供していただきました。また、大規模シミュレーションの共同研究においては、東北大学の中橋和博教授、高橋俊様、NECの撫佐昭裕様、NECシステムテクノロジーの曾我隆様、NECソフトウェア東北の下村陽一様から多大なるご協力をいただきました。皆様には心より感謝申し上げます。

参考文献

- 1) 江川隆輔, 「新大規模科学計算システム スーパーコンピュータSX-9の性能評価」, 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.41, No.2, pp.11-18, 2008.
- 2) Kobayashi, H., A. Musa, et al., “The potential of On-chip Memory Systems for Future Vector Architectures”, In M. Resch et al. (Eds.), High Performance Computing on Vector Systems 2007, pp.247-264. Springer-Verlag (2007).
- 3) Nakahashi, K., “Building-Cube Method for Flow Problems with Broadband Characteristic Length,” Computational Fluid Dynamics 2002, edited by S. Armfield, R. Morgan, K. Srinivas, Springer, 2003, pp.77-81.
- 4) Nakahashi, K., “High-Density Mesh Flow Computations with Pre-/Post-Data Compressions,” AIAA 05-4876, 17th AIAA Computational Fluid Dynamics Conference, Toronto, June 2005.
- 5) 日本電気(株), 「SUPER-UX MPI/SX利用の手引き」, 2008

執筆者プロフィール

小林 広明
東北大学
サイバーサイエンスセンター
センター長
教授

江川 隆輔
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助教

岡部 公起
東北大学
サイバーサイエンスセンター
産学官連携研究員

伊藤 英一
東北大学
情報部情報基盤課
技術専門員

大泉 健治
東北大学
情報部情報基盤課
技術専門員