ハイパフォーマンス
コンピューティング 84－5
（２０００．１２．８）

地球規模気候モデルNJR-SAGCMの並列計算性能

浅野俊幸　　堀口進
北陸先端科学技術大学院大学 情報科学研究科

近年、先端科学技術分野でのコンピュータ・シミュレーションの果たす役割が注目されている。特に地球温暖化やエルニーニョ等の地球規模の現象のメカニズムを解明する気候変動シミュレーションへの応用が期待されている。このような地球規模のシミュレーションには、大規模なデータを扱い高速計算が必要なため多数のプロセッサを結合した超並列システムでの高速化に期待が寄せられている。しかし、超並列システムを用いた地球規模のシミュレーションでは、プロセッサ間のデータ通信量の増加による並列計算性能の低下という問題が予想される。本稿では地球規模気候モデルNJR-SAGCMを並列計算機(IBM RS/6000 SP, CRAY T3E-600)にインプリメントし、気候シミュレーションの並列計算性能評価を行った。その結果、気候シミュレーションサイズの分解能によって、領域分割による最大並列化可能PE数の約半分でspeed-up比が飽和することが分かった。また、並列シミュレーションの処理時間の30%～50%がCollective通信によるオーバーヘッドであることが分かった。そこで各並列計算機で様々な通信性能を測定し、より詳細に検討を行う。

Parallel Performance Evaluation of a Climate Model NJR-SAGCM

Toshiyuki Asano,Susumu Horiguchi

Graduate School of Information Science,
Japan Advance Institute of Science and Technology

Computer simulations have been an attractive technology in the advance science and technology. Especially, the climate simulation for earth is expected to predict greenhouse effect and climate variability such as El Nino phenomenon. In the large-scale climate simulation, the problem of the increase in the data communication quantity between processors is anticipated. In this paper, parallel climatic model(NJR-SAGCM) is implemented on massively parallel computers ;IBM RS/6000 SP, CRAY T3E-600, and the parallel performances of climate simulation are discussed in detail. By the resolution of the climate simulation, it is proved that speed-up ratio was saturated in about the half of the PE number by the regional division which can parallelize largest. And, it is proved that 30%～50% in the processing time of parallel simulation is the overhead by collective communication. Then, various communication performance of each computer are measured, and the more detailed assesment is carried out.

1 はじめに

近年、計算機の高速化にともなって先端科学技術分野でのコンピュータ・シミュレーションの果たす役割が大きくなっている。特に地球温暖化やエルニーニョ等の地球規模の現象のメカニズムを解明する気候変動シミュレーションが注目されている。科学技術庁と関係機関においては、地球規模の複雑な諸現象をシミュレートするために超高速並列計算機システム「地球シミュレータ」の開発と、その上で利用する気候変動シミュレーションモデルの開発を行っている [1]。地球シミュレータは数百の計算ノード、メモリ 4TByte 以上、ピーク処理速度 40TFLOPS の大規模並列計算機である [2]。

こうした大規模並列計算機上で実行されるシミュレーションでは、多くの部分問題に分割することにより並列化効率は増加するが、各 PE 間でのデータ通信量が増加するために並列化効率が減少する [3]。本稿では、並列計算機上で実行可能な並列大気大循環モデル (NJR-SAGCM) を並列計算機にインプリメントし、大気大循環モデルによるシミュレーションの並列計算性能の評価を行った結果について詳しく検討する。計算機には分散メモリアーキテクチャの並列計算機である IBM 社の RS/6000 SP と共有型分散メモリアーキテクチャの並列計算機である CRAY 社の CRAY T3E-600 を用いて性能予測を行い、2 つの並列計算機の比較を行った。その結果、気候モデルを大規模な並列計算をする場合には、処理時間中の MPI_WAITALL のような待ち時間と Broadcast のような通信の振る舞いが処理時間に大きく影響を与えることが明らかになった。

2 大気大循環モデル (NJR-SAGCM)

大気大循環モデルは、地球上の大気の流れを計算機上で再現するものである。モデルは、様々な物理法則の基礎方程式 (運動方程式、状態方程式 etc) とモデルで直接表現できないサブグリッドスケールに分かれている。各々は、力学過程と物理過程に分類されている。力学過程は大気温度や気圧などのスケールが大きい大気の状態を計算し、物理過程は放射・積雲対流・地表面過程などのスケールが小さい計算をしている。

力学過程の計算にはスペクトル法を用いており、波数空間と格子空間との間での物理量の変換は球面調和関数展開を用いている。

マシン	CRAY T3E-600	IBM RS/6000 SP
プロセッサ	DECchip21164	PPC604e
動作クロック	300MHz	332MHz
トポロジー	3 次元トーラス	SP スイッチ
PE 間通信性能	480MB/s	150MB/s

表 1: 並列計算機の仕様

3 並列計算法

3.1 概要

図 1 に地球規模気候シミュレーションの並列計算における領域分割法と並列計算法を示す。図 1(b) に示すように東西方向にはフーリエ変換を行う。また、図 1(d) に示すように南北方向にはルジャンドル変換を行ってシミュレーションを行う [4]。

並列化手法として図 1(a) に示すように領域分割法を採用している。具体的には全球を緯度線で分割し、各領域をそれぞれノードに割り当て、必要時には図 1(c) に示すようにノード間でデータの交換を行い並列計算を実行する。NJR-SAGCM[5] のスペクトル法による力学過程と物理過程は、東京大学の気候システム研究センターと国立環境研究所と共同で開発した大気大循環モデル CCSR/NIES AGCM を、格子点法による力学過程については気象研究所で開発された大気大循環モデル MPI GCM をベースに用いている。

3.2 並列計算機環境

本稿で使用した並列計算機 CRAY T3E-600 と IBM RS/6000 SP の仕様を表 1 に示す。CRAY T3E-600 は 1 ノード 1PE によるシステム構成で、プロセッサ結合ネットワークは 3D トーラスを用いている。各通信リンクは各次元方向で入力および出力の独立したチャネルを持っている。IBM RS/6000 SP は 1 ノード 4PE によるシステム構成で、プロセッサ結合ネットワークには SP スイッチと呼ばれるクロスバースイッチを用いている。

3.3 入力データ

NJR 用の入力初期データは、地球表面には陸地がなく全て海であるとする全球海面としている。その上にある大気は全球平均気温で風や雲もない気候的に安定で平均的な物理量の分布をしている。従って、実際の大気で起こっている熱帯のスコー

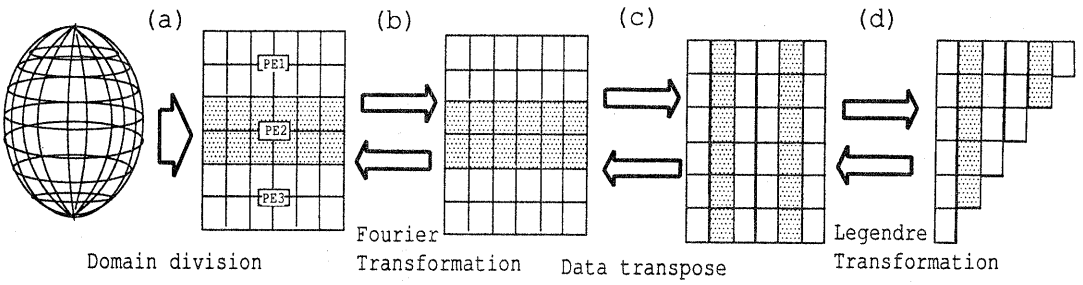

図 1: 地球規模の気候シミュレーションにおける領域分割法と並列計算手法

Truncation	EW x NS(Physical Grid)	Vertical Levels
T21	64 x 32	20
T42	128 x 64	20
T106	320 x 160	20

表 2: 気候シミュレーションサイズの分解能

ルのような激しい雨を降らせるような現象や、季節変化の気候シミュレーションには十分なモデルとは言えない。しかし、基本となる気候現象を表現するには十分なものである。

気候シミュレーションの空間分解能については、表 2 に示すような T21L20、T42L20、T106L20 を用いてシミュレーションを行った。これら入力初期データを用いた場合、領域分割で最大の並列化可能数は T21L20 では 32PE、T42L20 では 64PE、T106L20 では 160PE となる。

4 並列計算性能

4.1 speed-up ratio

異なる空間分解能のシミュレーションでは、解の安定性を保つために 1 日積分をしたときのタイムステップ数は異なる。そこで、使用 PE 数と処理時間の関係を求めるために、各分解能の単位積分時間と使用 PE 数の関係を用いる。また、シミュレーションでは、7 日間の積分計算をさせ、その処理時間の合計から 1 日の処理時間を求める。

図 2、3 に使用した PE 数と積分処理時間との関係を 1 台の PE で実行した処理時間の比、すなわち速度向上比を示す。実行可能 PE 数の約半分を過ぎたところから速度向上比が減少する傾向にあることがわかる。これは計算手法であるスペクトル法と関係があると考える。スペクトル法は、東西方

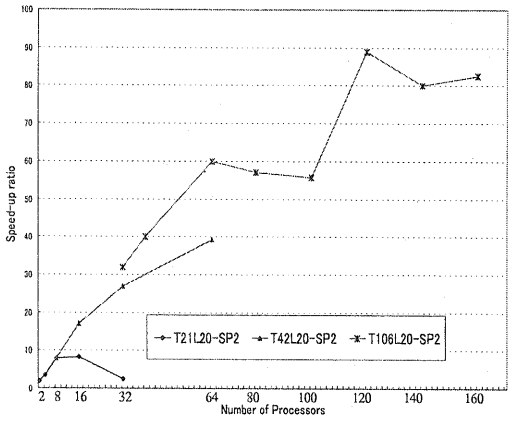

図 2: マルチプロセッサにおける速度向上比 (SP2)

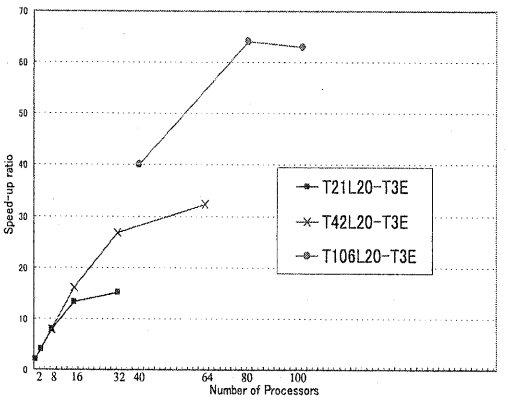

図 3: マルチプロセッサにおける速度向上比 (T3E)

向の計算のあと南北方向の計算をおこなう。この時にデータ転送を行うわけであるが、この通信時間が実計算時間に比べ無視できない時間になっていると考える。図2のT106L20で使用したPE数は、各PEに割り振る領域が実行可能PE数を均等割した数(32、40、80、160)を割り当てた場合と不均等割した数(64、100、120、140)で処理時間を求めた。図2からは均等割り・不均等割が速度向上比に対する影響は少ないと考えられる。これは、データ通信の同期をとる時間が実行PE数が増えたことにより相対的に減少したためと考える。

4.2 処理時間

積分処理時間を固定した場合、使用PE数に対する通信時間の関係を図4、5、6に示す。図にはSP2とT3Eのデータを示している。プロセッサ数の左側がSP2で、右側がT3Eのデータである。但し、計算機の制約によりデータがないものもある。

3つの空間分解能を比べると、実行PE数が少ない場合には1次元方向の分解能の比である1:2:4に近い結果を示す。ところが、実行PE数が増えるにつれてその差は小さくなる傾向がある。

図4、5では使用PE数の増加に比例して処理時間が減少する傾向が確認される。しかし、高分解能である6では実行可能PE数に近づくことにより、より高速化される傾向がある。これは気候モデルの計算方法が高い並列度に向いているためと考える。

4.3 通信時間

図4、5、6共に、使用PE数の増加に従い処理時間が減少する傾向は確認される。しかし、図4、5のSP2の実行時間を見ると、実行可能PE数では実行時間が増加する傾向を示している。一方、図6にはその傾向は顕著ではない。低い分解能の計算では、実行PE数が実行可能PE数の約半分を過ぎたところからは処理時間比べて通信時間が相対的に増加する傾向があると考える。T3EとSP2を比べると、表2からわかるようにT3Eのほうが通信性能が高い。そのため実行PE数が実行可能PE数の約半分を過ぎても処理時間比べて通信時間が増加する傾向は見られないと考える。

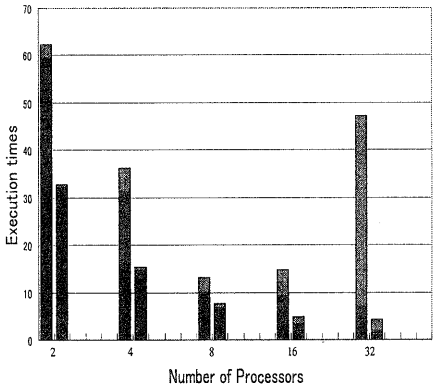

図4: T21L20の処理時間と通信時間

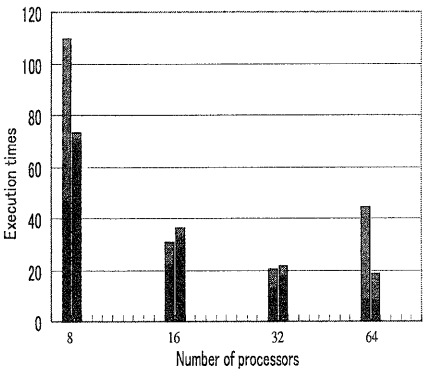

図5: T42L20の処理時間と通信時間

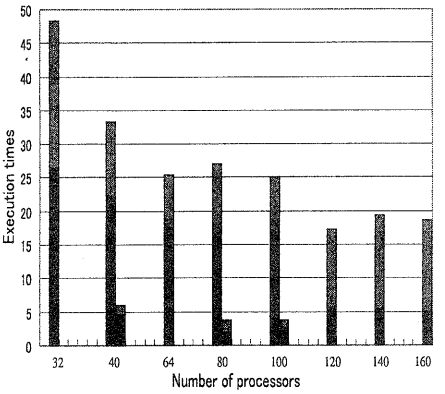

図6: T106L20の処理時間と通信時間

5 並列計算機の通信性能

5.1 シミュレーション結果

第4章で大気大循環モデル NJR-SAGCM の並列シミュレーションによる性能評価をすることにより、並列処理性能評価から以下のことが明らかになった。

- 各シミュレーション対象となる分解能における領域分割による並列化可能 PE 数の約半分以上の PE 数を用いた並列処理の速度向上比は減少する。
- PE 数の増加に従い、通信時間が処理時間に占める割合の増加が顕著である

これらのシミュレーション結果が気候モデルによるものか並列計算機システムに依存したものか議論する必要がある。

5.2 通信ライブラリ

気候モデルで使用する Collective 通信は、メッセージ通信ライブラリの MPI_BCAST、MPI_ISEND、MPI_IRECV 等である。

図6から分かるように T106L20 を SP2 の 140PE を用いて並列実行した場合には、特に MPI_BCAST、MPI_WAITALL は処理時間の約 30%～50%を占めることを確認した。

そこで、各並列計算機システムの Collective 通信の基礎的なデータを測定することにより、より詳細な議論を行う。

5.3 各通信時間の測定

各通信時間の測定は、ベンチマークソフト LLCbench (Low-Level Characterization Benchmarks)[6] を実行して測定した。LLCbench はメッセージ通信ライブラリの MPI の中から 1 対 1 通信、ブロードキャスト、バリア同期などを実行し、実行に要した時間から実行性能を測定するものである。本稿では、遅延時間、通信速度、ブロードキャストを各々100 回繰り返しその平均を測定値とする。

5.3.1 遅延時間

遅延時間は、1 対 1 通信においてメッセージ送信をするのに要した時間を 100 回繰り返し測定した平均時間である。図7に各並列計算機における遅延時間の測定値を示す。T3E、SP2 ともにメッセージサイズが 256Byte 以下のサイズでは通信時間に大きな変化が現れていない。これはメッセージを送る為の手続き時間が通信時間を大きく占めている為と考える。

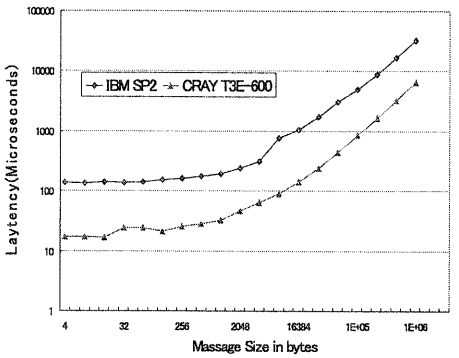

図 7: 各並列計算機における遅延時間の測定結果

5.3.2 平均通信速度

通信速度は、1 対 1 通信において 1 秒間に送受信されたメッセージサイズを 100 秒間繰り返し測定した平均メッセージサイズと定義する。図8に各並列計算機における平均通信時間を示す。T3E、SP2 ともに公称値に比べて 16%,11T3E と SP2 を比較すると通信性能の公称値が 480MB/s:150MB/s に対して約 79MB/s:16MB/s と、SP2 の通信性能が低く感じられる。

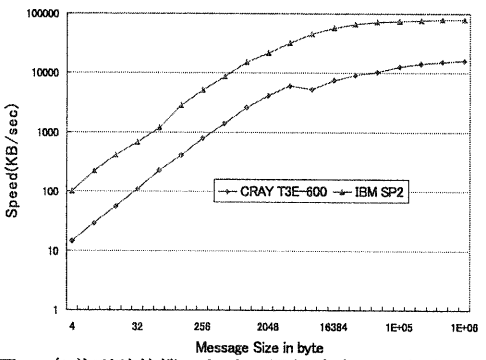

図 8: 各並列計算機における通信速度の測定結果

5.3.3 ブロードキャスト

ブロードキャストは、1 秒間に送受信されたメッセージサイズを測定したものである。各並列計算

機における 16PE に対するブロードキャストの結果を図 9 に示す。T3E の通信性能はほぼ公称値を示した。一方、SP2 は公称値に比べて約 40 ている。ここで送信されたメッセージサイズは 1MByte に満たないものであり、大容量の通信が 多数の PE で行われた時には何らかの傾向が見られることと考える。

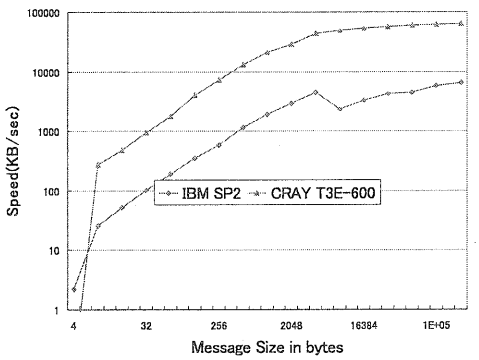

図 9: ブロードキャスト

5.4 通信時間と並列計算性能

1 対 1 ブロッキング通信の場合には SR/6000 SP の通信方式は T3E-600 より通信時間が短い傾向にある。また、Broadcast のような通信の場合には SR/6000 SP の方が通信時間が短い。一方、気候モデルでは 1 対 1 通信よりも Broadcast のような通信を多用している。しかし、図 2、図 3 に示した speed-up 比の測定結果から SP2 と T3E に差が見られない。従って、これは並列気候モデルの特徴と考えられ、実際に送受信しているメッセージサイズの大きさやそのタイミングなどを詳細にに検討する必要がある。

6 おわりに

大気大循環気候モデル NJR-SAGCM を実際に超並列計算機システム (IBM SR/6000 SP、CRAY T3E_600) 上で動作させ、その並列処理性能について議論した。また、各並列計算機における様々な通信時間を測定した結果から、次のことが明らかになった。

- 気候モデルでは Broadcast のような通信の振る舞いが処理時間に大きく影響を与える。
- 均等割り・不均等割が速度向上比に対する影響は少ないと考えられる。
- 気候モデルを大規模な並列計算をする場合には、処理時間中の MPI_WAITALL のような待ち時間が負荷分散に大きく影響を与えることが明らかになった。

今後、MPI_WAITALL のような待ち時間を少なくするような効率的な負荷分散制御を検討する必要があると考える。現在、タイムステップ毎に各 PE の処理時間を記憶したデータを用いて未来の処理時間をトレンド予測し、ある PE の計算負荷が増えると予測されればより負荷の少ない PE へ処理を移動させる方法を検討中である。

謝辞

本研究に当たり、NJR-SAGCM を提供して下さった (財) 高度情報科学技術研究機構に感謝致します。

参考文献

[1] 科学技術庁計算科学技術推進会議, "「地球シミュレータ」計画の推進について", 1996

[2] 横川三津夫, 新宮哲, 荻原孝, 磯辺洋子, 高橋正樹, 河合伸一, 谷啓二, 三好甫, "地球シミュレータ用性能評価システムの開発", 情報処理学会研究報告, 99-HPC-75,pp.55-60

[3] 浅野俊幸, 堀口進, "大気大循環モデル NJR-AGCM の並列計算性能評価", 情報処理学会第 61 回全国大会講演論文集 (1), pp.1-1,2000

[4] Saulo R.M. Barros,Tuomo Kauranne, "On the parallelization of global spectral weather models", Parallel Computing 20,pp.1335-1356,1994

[5] 井上孝洋, 後藤伸寿, 田中幸夫, 山岸米二郎, "並列化気候モデルの特性評価 (第 1 報)", 日本気象学会 1998 年秋季大会論文集, pp.272,1998

[6] LLCbench, http://icl.cs.utk.edu/projects/llcbench/