

## Heterogeneous Multi-Computer System: 連続体・多粒子系融合型超並列計算機システム

朴 泰 祐<sup>†1</sup> 牧野 淳一郎<sup>†2</sup> 須 佐 元<sup>†3</sup>  
梅 村 雅 之<sup>†3</sup> 福 重 俊 幸<sup>†4</sup> 宇 川 彰<sup>†3</sup>

Heterogeneous Multi-Computer System (HMCS) は、マルチスケールな計算物理シミュレーションを行なうために開発された、連続体向け超並列計算機と多粒子系向け超並列計算機を融合した新しい計算機システムである。我々は HMCS のプラットフォームとして、科学技術計算用超並列計算機 CP-PACS と、重力計算専用計算機 GRAPE-6 を、並列コモディティネットワークによって結合したシステムを開発中である。HMCS では、GRAPE-6 で行なわれるミクロスコピックな重力多体計算と、CP-PACS で行なわれるマクロスコピックな輻射流体計算を並行して行ない、非常に詳細な計算宇宙物理シミュレーションを実現する。両システムは PIO (Parallel I/O System) と呼ばれる、コモディティネットワークを用いた並列入出力システムで結合され、各計算タイムステップ毎に、計算対象となる全粒子データを交換し合う。本稿では、HMCS の全体構想・性能予測・実装・予備性能評価及びそれによってもたらされる新しい計算物理学に関して報告する。

## Heterogeneous Multi-Computer System: A massively parallel processing system combining continuum and particle simulators

TAISUKE BOKU,<sup>†1</sup> JUNICHIRO MAKINO,<sup>†2</sup> HAJIME SUSA,<sup>†3</sup>  
MASAYUKI UMEMURA,<sup>†3</sup> TOSHIYUKI FUKUSHIGE<sup>†4</sup>  
and AKIRA UKAWA<sup>†3</sup>

HMCS (Heterogeneous Multi-Computer System) is a new parallel processing platform combining massively parallel processors for continuum simulation and particle simulation to realize multi-scale computational physics simulations. We are constructing a prototype system of HMCS with a general purpose scientific parallel processor CP-PACS and a gravity calculation parallel processor GRAPE-6 connecting them via commodity-base parallel network.

On the prototype of HMCS, a micro-scopic gravity calculation on GRAPE-6 and a macro-scopic radiation hydrodynamic calculation on CP-PACS are performed simultaneously to realize very detailed simulation on computational astrophysics. Both systems are connected via parallel network controlling system named PIO (Parallel I/O System). On each time step, all data of particles are exchanged between two systems hiding communication latency with a special algorithm and buffering effect by PIO.

In this paper, we report the overall concept of HMCS, predicted performance, implementation and preliminary performance evaluation of the system as well as new physics provided by this system.

†1 筑波大学 電子・情報工学系

Institute of Information Sciences and Electronics, University of Tsukuba

†2 東京大学 理学系研究科 天文学専攻

Department of Astronomy, Graduate School of Science, The University of Tokyo

†3 筑波大学 物理学系

Institute of Physics, University of Tsukuba

†4 東京大学 総合文化研究科 広域科学専攻

Multi-Disciplinary Sciences, Graduate School of Arts and Sciences, The University of Tokyo

### 1. はじめに

計算物理学における大規模シミュレーションには、並列ベクトル計算機・超並列型スカラー計算機・専用並列計算機等、様々な HPC 向け並列計算プラットフォームが用いられる。従来、これらのプラットフォームは単体で用いられることがほとんどであり、最近になって、Grid 環境等を用いていわゆる「スパコン」間の接続を行おうという機運も生じているが、実用研究と

してはまだ時間がかかる。このような環境下では、当然、対象となる計算もそのプラットフォームに合わせざるを得なくなり、より広汎な物理計算を行なう際に実効性能・問題規模等の点で制約条件とならざるを得ない。

このような背景の下、我々は次世代の大規模シミュレーションプラットフォームとしてHMCS (Heterogeneous Multi-Computer System) を提案する。HMCS は異なるアーキテクチャを持つ (並列) 計算環境の融合体の総称であり、例えば超並列計算機と専用並列計算機を高速・大容量ネットワークによって有機的に結合し、互いを相補的に用いることにより、従来実現不可能であった新しいパラダイムによる物理シミュレーションを可能とするものである。このようなシステムを実現するに当たり、ボトルネックとなるのは、言うまでもなく複数のシステム間を結合するネットワークである。現在、コモディティネットワークの性能は日々増大しつつあり、Gbps クラスのネットワークが PC にも導入できる時代になっている。しかし、実際の超並列システム等にこれを用いる場合、入出力プロセッサにおける処理の集中・絶対的バンド幅の不足等の問題が生じる。これに加え、一世代前のプラットフォームにおいてこれらの最先端ネットワークインタフェースが必ずしも利用できない等の問題もある。

我々は、これらの問題を解決する手法として、筑波大学において研究開発中である、コモディティネットワークに基づく並列入出力システム PIO (Parallel I/O System) を用い、超並列計算機と重力多体問題専用並列計算機を結合した HMCS プラットフォームを開発中である。

## 2. HMCS プロトタイプの基本設計

我々が開発中の HMCS プラットフォーム (以下、便宜上、単に HMCS と呼ぶ) では、汎用超並列計算機 CP-PACS<sup>1)</sup> と、重力計算専用並列処理システム GRAPE-6<sup>2)</sup> を用いる。CP-PACS は 2048 台の計算専用プロセッサ (PU: Processing Unit) と、128 台の入出力プロセッサ (IOU: I/O Unit) を持つ。また、GRAPE-6 システムは 32 チップの重力計算専用プロセッサを 1 つのボードに実装したものであり、これをさらに複数用いることにより、プロセッサ数に比例した超高性能な重力計算エンジンが提供される。

CP-PACS には 128 台の IOU のうち 16 台に対し、100base-TX イーサネットのインタフェースが実装されている。この他に汎用インタフェースとしての 10base-T イーサネットと、フロントエンドコンピュータとの間の大容量データ通信のための HIPPI (ピーク性能 100MB/s) も別途用意されているが、今回の HMCS 設計では後述の GRAPE-6 ホストシステムとの並列接続によりネットワークバンド幅を確保をするため、

16 本の 100base-TX イーサネットを用いる。2048 台の全ての PU は、どの IOU を経由しても通信を行なえる。

GRAPE-6 ボードそのものはスタンドアロンタイプの計算機システムではなく、インタフェースを通じてホスト計算機と接続し、制御される。現在のインタフェースは 32bit PCI であり、Alpha 系及び Pentium 系のプロセッサを持つ PC と接続可能である。GRAPE-6 の PCI インタフェースには制御用プロセッサが実装されており、ホストからは PCI バスを通じて命令及びデータのやりとりを行なう。GRAPE-6 では  $N$  個の粒子に対する  $N^2$  の総当たりの重力計算を行なう。このため、まず  $N$  個の全粒子データ (質量・位置・速度・前ステップにおける加速度等) を GRAPE-6 上の高速メモリに格納しておき (これを  $j$  粒子と呼ぶ)、その後で  $N$  個の粒子データを改めて内部のパイプライン演算器のレジスタにセットして一気に処理する。GRAPE-6 チップでは、1 回のパイプライン演算でセットできる粒子 (これを  $i$  粒子と呼ぶ) 数は 48 個である。従って、 $i$  粒子に関しては  $N/48$  回の演算を行なうことになる。

GRAPE-6 ボードを多数用いれば、 $i$  粒子を分散させることにより、処理性能がボード数に比例して向上する。GRAPE-6 では、複数のボードの入出力を束ねるツリー型通信ボードも別途用意されているが、HMCS では以下の 2 つの理由によりこれを用いない。まず第一に、ホスト計算機にあまり多数の粒子を集中させると、GRAPE-6 の演算性能に対し、ホスト上での汎用処理及び通信処理を行なう際にこちらがボトルネックとなってしまう。第二に、CP-PACS との通信において、並列ネットワークを用いたバンド幅向上を狙う都合上、ホスト計算機も並列構成になっていた方が都合が良い。以上の理由により、HMCS では GRAPE-6 のホスト計算機を、複数の PC によるクラスタ構成とし、ノード PC が GRAPE-6 ボード 1 枚ずつを担当し、必要に応じてノード間通信も行ないながら、重力計算部を構成する。

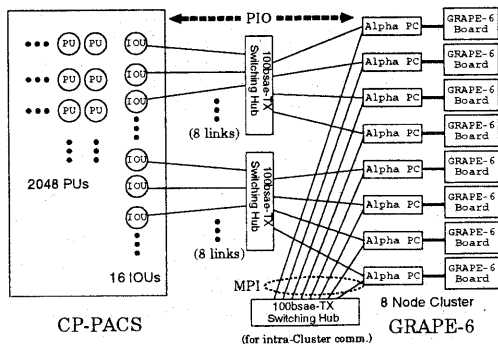


図1 HMCSの全体構成図

図1にHMCSの全体構成を示す。現在の構成では、8枚のGRAPE-6ボードを用いるため、クラスタのノードも8台になっている。CP-PACSからの16本の100base-TXイーサネットリンクと、クラスタの8本のそれは、2台のSwitching Hubを介して接続される。本システムで用いるPIO (Parallel I/O System)<sup>4)</sup>は、ネットワークが複数のHubを用いたサブネット構成になっている場合でも、静的及び動的負荷分散機能を用い、これに対応するように設計されている。本システムの場合、総イーサネットリンク数は24本となるため、1台のHubでも収容可能ではあるが、これ以外に周辺機器として、ファイルサーバであるSGI Origin-2000 (8プロセッサ、100base-TXイーサネット8本)及びグラフィックサーバであるSGI Onyx2 (4プロセッサ、100base-TXイーサネット4本)も収容しなければならないため、実際のリンク数は30本以上になる。よって、PIOのサブネット機構を積極的に利用し、総バンド幅を広くするためにこのような構成を取った。

### 3. 基本アルゴリズムと性能予測

GRAPE-6を用いたHMCSでは、重力多体計算をベースにした様々な物理シミュレーションを扱える。現在、我々は重力効果を取り入れた、SPH (Smooth Particle Hydrodynamics) に基づく宇宙輻射流体計算<sup>3)</sup>を最初のアプリケーションとして考えている。従来の宇宙輻射流体計算では、重力効果による物体の移動が存在しないという前提で、固定的な粒子配位に対する処理を行ってきた。しかし、実際の銀河生成のシミュレーションをより正確に行うためには、力学的現象、輻射輸送、化学反応の全てを加味した精密なシミュレーションが必要である。これをHMCSによって実現するのが目的である。

計算アルゴリズムは以下の通りである。

- (1) 全粒子の位置・速度、場の温度・密度・化学種の量を初期条件として与える。
- (2) CP-PACSからGRAPE-6に粒子データを送り、重力効果を評価する。
- (3) SPHにより密度・圧力・圧力勾配を計算する。
- (4) 速度・力から時間刻み幅を決定する。
- (5) SPH粒子の大きさを変更する。
- (6) 輻射輸送計算・化学反応計算・SPHエネルギー方程式の計算を、温度・化学種が収束するまで繰り返す。必要によってタイムステップをさらに短く調整する。
- (7) SPH運動方程式により粒子の運動を積分する。
- (8) ステップ(2)~(7)を必要タイムステップだけ繰り返す。

計算量のオーダーは以下の通りである。以下、 $N$ は総粒子数、 $N_{neib}$ は、ある粒子の近傍の粒子数、 $N_{src}$ は

輻射の源となる粒子数、 $f_i$ は計算にかかるファクタを現わす。

- 重力 (GRAPE-6):  
 $f_1 N^2, f_1 \sim O(10^2)$   
 $\Rightarrow 1Tflop(N/10^5)^2/step$
- 流体 (CP-PACS):  
 $f_2 N N_{neib}, f_2 \sim O(10^3), N_{neib} \sim O(10^2)$   
 $\Rightarrow 10Gflop(N/10^5)/step$
- 輻射輸送+化学反応 (CP-PACS):  
 原始組成のネットワークで調べる限り、化学反応に要する時間は輻射輸送に比べ無視できる。  
 $f_3 N^{4/3} N_{neib} N_{src}, f_3 \sim O(10^4)$   
 $\Rightarrow 100Gflop N_{src} (N/10^5)/step$

従って、演算量だけを評価すると、10万程度程度の計算を行う場合、CP-PACSとGRAPE-6での計算がバランスするには、

$$N_{src} \sim 10 \left( \frac{N}{10^5} \right) \frac{[CP-PACS flops]}{[GRAPE flops]}$$

程度の $N_{src}$ を扱うことになり、その下で $10^4 \sim 5$ ステップの計算を行なう。

一方、通信性能に関する評価は以下の通りである。GRAPE-6は重力多体計算の副産物として、各粒子に対する近傍粒子リストを生成することができる。しかし、現在のHMCSでは両システムの計算パワーは非常に強力であるため、通信コストを極力抑える必要がある。このため、近傍粒子リストは転送せず、CP-PACS側で直接生成するものとする。この場合のCP-PACSからGRAPE-6へのデータ転送では、各粒子に対して質量・3次元座標・1ステップ前の加速度のファクタ・1ステップ前のポテンシャルの合計6つの値を送る。倍精度浮動小数点を用いるため、総転送量は

$$N \times 6 \times 8/step = 48Mbyte \left( \frac{N}{10^5} \right) /step$$

となる。逆に、GRAPE-6からCP-PACSに戻されるデータは、各粒子に対して3次元加速度・ポテンシャルの合計4つの値である。よって、総転送量は

$$N \times 4 \times 8/step = 32Mbyte \left( \frac{N}{10^5} \right) /step$$

となる。

これらのパラメータを、予備実験のCP-PACS flops値とGRAPE-6 flops値、及びPIOの性能<sup>4)</sup>と組み合わせた結果、10万程度を扱う場合、1ステップ当たりCP-PACS側で約16秒、GRAPE-6側で1秒未満、通信時間も1秒未満となり、CP-PACS側でのSPH計算が支配的となることがわかった。しかし、実際に重力効果をCP-PACSで処理した場合、現実的な時間で計算することは不可能に近いので、GRAPE-6をHMCSに導入する意義は十分に高い。

## 4. プロトタイプの実装

### 4.1 CP-PACS

本論文はHMCS自体に関する説明を主とするため、CP-PACS上におけるSPH計算の詳細については省略し、並列化の要点と、GRAPE-6とのインタフェースに関して説明する。

CP-PACS上で動作するSPH計算プログラムでは、1000PU規模の超並列アルゴリズムを実装している。ここでは、全粒子を座標に従って空間的に分割し、それらを並列処理する。ただし、粒子は全PUにおいてその数が均等になるように分割し、必ずしも直行空間のセルに従って厳密に分割するわけではない。すなわち、粒子処理における負荷分散が主目的であり、なるべく近い粒子を1つのPUに収めるようにすることは、近傍粒子計算を効率的に行なうためのものである。

1つの計算ステップでは、GRAPE-6から返された現在の粒子配置における加速度より、粒子の位置を決定し、上述のPU間分割を行なう。その後、全粒子に対し輻射輸送+化学反応の計算を行なう。前節で述べた通り、途中の計算では化学過程の収束のための反復計算を伴う。計算終了後、全粒子の質量・座標・加速度のファクタ・力のポテンシャルをGRAPE-6に送り、次の計算ステップに移行する。なお、今回の実装では各粒子の質量は変化しないが、後述のGRAPE-6クラスタAPIに合わせるため、毎回転送している。また、1000以上のPUから一斉に、個別にデータ転送を行なうのは通信回数と通信粒度の点であまりに非効率的である。よって、適当な数の入出力取りまとめPUを設け、これらがエージェントのような形でGRAPE-6クラスタと通信する。エージェントの数は必ずしもGRAPE-6クラスタのノード数と一致はしていない。

### 4.2 GRAPE-6クラスタ

HMCSプロトタイプでは、8枚のGRAPE-6ボードを用いている。これらを制御するため、8ノードのPCクラスタを用意し、各ノードのPCIスロットにGRAPE-6インタフェースボードを実装している。各ノードの主な仕事は、CP-PACSから並列ネットワーク経由で送られてくる粒子データをGRAPE-6ボードに送り、結果を再びCP-PACSに返すことであるが、これに付随して、クラスタ内での全粒子データの交換作業も行なう。

CP-PACSから送られてくる $N$ 粒子のデータは、ネットワークにおける通信量を最小化するため、8台のノードに $N/8$ 粒子ずつ送られる。しかし、GRAPE-6ボードは、全ボードに対し、全 $j$ 粒子をセットし、その上で $i$ 粒子を順次セットしながら演算を進めていく。従って、各ノードはGRAPE-6ボードをアクセスするに当たり、全粒子のデータを持っている必要がある。

このため、ノード間で全対全のデータ交換を行なう。GRAPE-6クラスタのこれらの並列制御はMPIプログラミングによって行なわれている。クラスタ間の通信には、CP-PACSとの通信において用いられているネットワークとは別のネットワークが用いられる。これは、CP-PACSとの通信用並列ネットワークは複数のサブネットに分割されているため、クラスタ全体を束ねるフラットなネットワークが別途必要なためと、MPIによる通信がCP-PACSとの間の通信を邪魔しないように考慮してのことである。

クラスタのノードにはSamsung UP1100マザーボードを用いた。ここにCPU周波数600MHzのAlpha 21264Aを載せ、768MBのSDRAMメモリで運用する。OSはRedHat 6.2J for Alphaで、Linuxカーネルは2.2.16である。UP1100にはon-board EthernetとしてDEC Tulipベースのものが載っているが、上述の用途のためIntel EthernetPro100を別途実装している。

クラスタの各ノードでは、通信及びデータ交換を行なうためのg6pioservと呼ばれるサーバプログラムが1つずつ走る。1つのg6pioservはCP-PACS上の複数の入出力エージェントと通信するようプログラムされている。GRAPE-6インタフェースの制御には、公開されているGRAPE-6 API<sup>5)</sup>を用いる。これはGRAP-6ボードに対する一連のアクセスルーティンのライブラリである。また、PIOを用いた通信を行なうため、PIO-APIを用いる。

### 4.3 PIOネットワーク

両システムを結合する100base-TXベースのネットワークは、並列入出力システムPIO<sup>4)</sup>によって制御される。PIOは並列に張られたコモディティネットワークにおいて、バンド幅の増大・大容量バッファリングによる通信時間隠蔽・並列ネットワーク間の自動負荷分散制御等を目的とした、ユーザレベルの超並列計算機向け入出力システムである。今回の実装では、CP-PACS側のエージェントと、GRAPE-6クラスタ上のg6pioservとの間でPIOによる通信が行なわれる。エージェント当たりの送受信粒子データ数が均一化され、またデータ送受信タイミングもアプリケーションによりほぼ同期しているため、HMCSではPIOが提供する動的負荷分散機能は用いていない。

エージェントでのプログラミングを容易にするため、CP-PACS側の任意のノードから、GRAPE-6クラスタをアクセスするためのg6cplibと呼ばれるAPIを新たに設けることにした。このAPIライブラリは、g6pioservとの間のプロトコルに基づき、PIO-APIを用いた通信を行なう。このライブラリを用いることにより、今後、CP-PACSから自由にGRAPE-6クラスタをアクセスすることが可能となる。表1にAPIの概略を示す。

表1 GRAPE-6 クラスタアクセス API

ルーティン名	機能
g6cpp_start	通信モードをセットし処理を開始する
g6cpp_unit	粒子数をセットし座標と時刻の値をセットする
g6cpp_calc	担当粒子データをセットし計算を開始する
g6cpp_wait	GRAPE-6 による演算結果を待ち加速度等のデータを返す

## 5. 予備性能評価

HMCS は現在実装中であり、残念ながら現段階では実装に基づく性能評価及び物理計算結果は得られていない。ここでは、HMCS 実装のための、4 ノードの GRAPE-6 クラスタからなる実験システムにおける性能を示す。

CP-PACS を用いたフルサイズのシステム実装の前に、PIO ネットワークの性能と GRAPE-6 クラスタの性能を測定した。測定においては、CP-PACS の代わりに、8 ポートの 100base-TX イーサネットインタフェースを持つ Origin-2000 (8 プロセッサ) を使い、実際には 4 ポートを PIO 用に設定し、4 ノードの GRAPE-6 クラスタとの間で、g6cpplib を用いた通信と GRAPE-6 ボードでの重力計算時間を求めた。Origin-2000 上のテストプログラムは、ダミーデータを g6cpplib を用いて GRAPE-6 クラスタに転送する。GRAPE-6 クラスタでは g6pioserv を走らせ、各ノードに設置された GRAPE-6 ボードで計算を行なう。Origin-2000 上のプログラムがダミーである以外は、PIO 通信に関しては実際の HMCS と基本的に同じである。

仮想的に、8 つのエージェントが各々 16,384 粒子ずつ、すなわち全体で約 13 万粒子の処理を GRAPE-6 クラスタに要求した場合の結果を表 5 に示す。ここに示したのは、g6pioserv 内の各部の処理内訳と全処理時間を、それぞれ 1 計算ステップ当りに換算した値である。実際には 4 ノードでの並列処理になるため、内訳に関しては平均値を、総処理時間に関しては mpirun で g6pioserv を起動してから終了するまでの wall clock に基づく時間を示している。

表2 実験システムの 1 ステップ当たり処理時間 (13 万粒子) 単位 (秒)

内訳	時間
g6cpp_unit	0.843
g6cpp_calc	0.334
クラスタ内の MPI 処理	0.746
全 $j$ 粒子のセット	0.510
全 $i$ 粒子の処理	0.435
g6cpp_wait	0.085
総処理時間	3.243

表では g6cpp\_wait の処理が非常に短く見えるが、こ

れは PIO が提供するバッファに加速度データを格納しただけで処理が終わったように見えるため、実際にはこの裏で PIO ネットワークでのデータ転送が行なわれる。その結果、次の g6cpp\_unit のリクエストが来るまでは逆に時間がかかるため、表中の g6cpp\_wait、g6cpp\_unit はセットで考える必要がある。さらに、クライアント側では g6cpp\_unit の結果を待たずに g6cpp\_calc を実行しているため、これら全てを含めた時間で粒子データの相互通信が行なわれていると考えるのが妥当である。

実験の結果、総処理時間は 1 ステップ当たり 3.24 秒となっている。これは、ノード数が 4 しかないため、各ノードにおけるデータ受信・GRAPE-6 への粒子データセット ( $j$  粒子を除く)、実際の重力計算時間等が全て 8 ノードの場合より長くなっているためで、8 ノードを用いた場合、総処理時間は 2 秒前後になるものと期待される。13 万粒子の処理をこの時間で行なえる効果は大きい。また、CP-PACS 側の SPH プログラムでは、GRAPE-6 ノードへの通信処理を前倒しして行ない、通信時間の隠蔽を行なっているため、HMCS の実質的な計算時間は、ほぼ CP-PACS での SPH 計算処理にかかっているとと言える。

## 6. HMCS によってもたらされる新しい計算宇宙物理学

宇宙には、恒星・ブラックホール・星団・銀河。銀河団・宇宙大規模構造など様々な階層の構造が存在する。しかし、それらを支配する基礎物理過程の観点から見ると、およそ 3 つに集約される。それは、流体過程、重力過程、輻射過程である。宇宙に登場する流体過程は主に圧縮性流体であり、航空工学で使われるものと基本的に同じである。また、宇宙の現象では重力が本質的な役割を果たしている。地上の重力場と違う点は、地上重力は一定であり外場として扱うことができるが、宇宙の現象では各部分に働く引力の総和によって運動が決まるという点である。このような重力の働き方を、外場重力と区別して“自己重力”と言う。自己重力系を扱うためには粒子数  $N^2$  の計算が必要になる。もう一つ重要な基礎過程に輻射 (電磁波) がある。輻射は宇宙に普遍的に存在するものであり、あらゆる宇宙現象において、エネルギー収支を司っている。輻射は、3 次元空間の各点において方向 2 次元、振動数 1 次元の自由度を持つため、輻射を正確に取り扱うためには、6 次元の自由度で光子の分布関数を求めなければならない。光子の分布関数の変化を記述する式は、輻射輸送方程式と呼ばれる。従って、宇宙現象の本質的な解明には、流体方程式、自己重力場、輻射輸送方程式を組み合わせる計算を行うことが必要である。宇宙現象は、基本的に非線型であるため、これらの式は解析的に扱うことは難しく、超高速計算機による数値シミュ

レーションが極めて重要になる。これまでは自己重力と輻射輸送方程式の計算量の膨大さから、これらをすべて採り入れた重力輻射流体力学の実現は困難であった。HMCSでは、 $N^2$ のコストのかかる自己重力計算を重力計算専用機 GRAPE-6 によって高速に処理し、6次元問題である輻射輸送を汎用スーパーコンピュータ (CP-PACS) によって解き上げ、数値流体と結合すること、重力輻射流体力学の実現を可能にする。

現在、重力輻射流体力学の重要なターゲットとして、銀河形成問題への適用を進めている。銀河は、宇宙を構成する基本的単位となっている天体であるが、その形成史は多くの謎に包まれている。中でも、楕円型銀河と円盤型銀河という2種類の銀河が存在するという事実は、長年の未解決問題であり、それは銀河形成問題を解く鍵を握ると考えられてきた。我々は、銀河形成期の宇宙には紫外線光が充満しており、宇宙は電離状態にあるということに注目した。このような紫外線光が銀河形成に及ぼす効果を正確に調べるためには、重力輻射流体力学計算が不可欠である。ワークステーションを用いた数千粒子程度の予備的な計算で、比較的初期に出来る銀河では背景紫外線光に対する自己遮蔽という効果が強く効いて、星形成が活発に起こるのに対し、後期にできる銀河では背景紫外線光に阻害され星形成がゆっくりしか進まない傾向が示された。前者はやがて力学的緩和を起こし、楕円型銀河となる可能性がある。また、後者は角運動量が効くまでガスのまま重力収縮し最終的に円盤型銀河へと進化する可能性がある。目下、HMCSを用い数十万体のガス粒子と星の系を重力輻射流体力学によって解き上げること、銀河形成問題の本質的解明に迫ることを計画している。この計算によって、実際に楕円銀河や円盤銀河が形成される過程が明らかになるであろう。さらに今後、HMCSの計算能力が大きく向上すれば、宇宙における様々な階層構造の形成史の解明に大きな進展をもたらすであろう。

## 7. おわりに

本稿では、マルチスケールな複合物理シミュレーションを行なうための新しい計算プラットフォームとして、HMCS (Heterogeneous Multi-Computer System) を提案し、そのプロトタイプとして、汎用超並列計算機 CP-PACS と重力計算専用並列システム GRAPE-6 を、並列コモディティネットワーク制御システム PIO によって結合した実装例を示した。

HMCS は現在最終段階の実装を行っており、予備実験では GRAPE-6 と Alpha 21264A を用いた4ノードのクラスターで、13万粒子に対する処理を、ネットワーク遅延を含め1ステップ当たり約3.2秒で処理できることが示された。これに CP-PACS 上での SPH 計算を組み合わせて、銀河形成の詳細なシミュレーショ

ンを行なうことが当面の目標である。このシミュレーションでは、CP-PACS の1024台のPUと16台のIOU、そして8ノードからなるGRAPE-6クラスターを用いることが計画されている。

今後の課題として、HMCSの構成要素である、g6pioserv サーバプログラムとg6cpplib API ライブラリのチューニングを行なうことが必要である。特にPIOを用いたプロトコルでは、ユーザデータの配列の制約上、10万粒子程度では通信粒度が比較的小さく、より効率の良い実装を考える必要がある。また、現在のg6cpplibではGRAPE-6が提供する機能のうち一部のみを用いており、より広汎なシミュレーションを行なうため、APIとしての充実化を図る必要がある。

今後の大きな課題としては、HMCSそのものの枠組の見直しがある。現在のHMCSは基本的に2つの高性能並列処理システムを結合した形を取っているが、より柔軟で高い性能を求めるならば、両者の結合ネットワークがボトルネックになる。例えば、並列システムのノード内において、汎用プロセッサの横にコプロセッサのような形でGRAPEチップを設け、並列処理システムが持つ高いネットワークバンド幅をそのまま活かした通信を行なうような、「汎用・専用ノード一体型」のような形でHMCSを実現するのが理想であろう。

**謝辞** 本研究を行うにあたり、御意見・御協力を頂いた筑波大学計算物理学研究センター未来開拓プロジェクト関係各位に感謝致します。なお、本研究の一部は日本学術振興会未来開拓学術研究推進事業「計算科学」のプロジェクトによるものである。

## 参考文献

- 1) K. Nakazawa, et.al., "CP-PACS: A massively parallel processor at the University of Tsukuba", *Parallel Computing*, Vol.25, pp.1635-1661, 1999.
- 2) J. Makino, et.al., "A 1.349 Tflops simulation of black hole in a galactic center on GRAPE-6", in *Proc. of SC2000 (CD-ROM)*, IEEE, Los Alamitos, 2000.
- 3) M. Umemura, "Three-dimensional hydrodynamical calculations on the fragmentation of pancakes and Galaxy formation", *The Astrophysical Journal*, 406, pp.361-382, 1993.
- 4) 松原 正純 他, "分散メモリ型超並列計算機における並列入出力", *情報処理学会論文誌ハイパフォーマンスコンピューティングシステム*, Vol.41, No. SIG 5 (HPS1), pp.58-69, 2000.
- 5) GRAPE-6 documents (on WWW), <http://grape.astron.s.u-tokyo.ac.jp/pub/people/makino/software/GRAPE6/>