

東大・国立天文台グループのスーパーコンピュータ、
1448 Mflops/W の超低消費電力スーパーコンピューティングを実現

Green 500 リストで僅差の 2 位

1. 発表者：

平木 敬（東京大学大学院情報理工学系研究科 創造情報学専攻 教授）

牧野淳一郎（国立天文台 理論研究部 教授）

稲葉真理（東京大学大学院情報理工学系研究科 創造情報学専攻 准教授）

2. 発表内容：

2010 年 11 月に発表された Green500 リスト¹において、東京大学・国立天文台が共同開発したスーパーコンピュータシステム GRAPE-DR システム²が HPL ベンチマーク³で 1 ワットあたり **1448.03 Mflops/W**⁴（1 ワットあたり 14 億 4 千 8 百万回の演算）の処理性能を実現し、第 2 位にランクされました。この省電力性能は、今年 6 月に GRAPE-DR システムが Little Green 500⁵で達成した 815Mflops/W の省電力性能を 1.8 倍と大幅に更新したものです。なお、Green500 リストは 12 月 22 日に更新され、同時に発表された Little Green 500 リストにおいても、GRAPE-DR は第 2 位にランクされています。

Green 500 リストの第 1 位、第 3 位のシステムは、冷却方式として水冷技術⁶あるいはラック単位水冷技術⁷を用いている一方、GRAPE-DR は空冷技術⁸を採用していることで、不利な条件下にあるにもかかわらず、第 3 位に大きく差をつけた優れた省電力性能を示しました。GRAPE-DR システムは、空冷スーパーコンピュータとして世界一の省電力性能を持っています。

GRAPE-DR のプロセッサが日本で設計・開発されたものであることから、世界中がしのぎを削っている「スーパーコンピュータの超低消費電力化」において、日本の技術が世界のトップレベルを保持していることを示しています。また、開発した超低消費電力技術はスーパーコンピュータに限定されず、将来の超低消費電力情報システム全般を構築する基礎技術となるものと期待されます。

3. 発表詳細：

[本研究成果の意義]

- 1) 省電力スーパーコンピュータ技術を実現し、スーパーコンピュータで問題となる環境問題を解決し、運転費用を大きく削減することを可能にしました。
- 2) 2010 年 6 月に Little Green 500 で達成した省電力性能から約 **1.8 倍**の演算速度／消費電力比を実現しました。米国の BlueGene/Q システムから僅かに下回ったものの、国産コンピュータとして画期的な省電力性能を示しました（付録 1，2）
- 3) GRAPE-DR は 90 ナノメートル半導体を使い、半導体テクノロジー的には 2 世代前で古くなっているにもかかわらず、最新の半導体テクノロジー(32, 40 もしくは 45 ナノメートル)で

製造したスーパーコンピュータに並ぶ省電力性能を示し、省電力スーパーコンピューティングへの GRAPE-DR システム基本設計の優位性を示しました。

- 4) 消費電力低減が第一の技術的課題である次々世代スーパーコンピュータの基礎技術として、省電力型アクセラレータが優れていることを実証し、基礎技術として確立しました。

[競合システムとの性能比較]

今回の第1位は、米国 IBM 社 T.J. ワトソン研究センタに設置されたブルーグリーン/Q システム⁹で、水冷技術を駆使して効率を上げ、1684 Mflops/W の省電力性能をあげました（GRAPE-DR より 16%演算性能／消費電力比が優れています）。第3位は東工大の TSUBAME 2.0 システムで、ラック単位水冷技術により効率を上げ、958 Mflops/W の省電力性能でした。第4位は米国 NACA の GPU クラスタで、933Mflops/W の省電力性能、第5位は日本の次世代スーパーコンピュータ「京」で、829 Mflops/W の省電力性能でした。GRAPE-DR は、これらに対して、各々51%（対3位）、55%（対4位）、75%（対5位）もの優れた省電力性能を達成しています。

[詳細説明]

Green 500 プロジェクトは、バージニア工科大学の Feng 教授が 2006 年に IEEE¹⁰で提案したことから始まっており、Linpack 性能を測定する HPL ベンチマークでの電力あたり性能を順位付けすることで、電力あたり性能が高い計算機の開発・普及を促進しています。

今回の速度・消費電力の測定は、国立天文台三鷹キャンパスに設置された GRAPE-DR システム(写真1)で行いました。測定に使ったシステムは、GRAPE-DR システム全体のうち 81 ノードであり、1 ノードは GRAPE-DR ボード1枚、インテル製 Core i7-920 CPU、ASUS 製マザーボード、18GB DDR3 メモリからなるもので、東大・天文台で開発した GRAPE-DR ボード(写真2)以外は、一般市販部品を使用しています。

GRAPE-DR ボードを利用せず、インテル製 CPU だけを使った場合、今回測定対象にした HPL ベンチマークでの消費電力あたりの性能は 250Mflops/W 程度となります。GRAPE-DR ボードを使うことで、電力あたり性能を約 6 倍の **1448.03 Mflops/W** に引き上げることができました。

今回、電力あたり性能を 6 月から更に約 1.8 倍改善したベース技術は以下の通りです。

1. 超低消費電力を実現する 1 チップ超並列プロセッサ「GRAPE-DR プロセッサチップ」の開発。チップ単体で、200Gflops の性能を 50W の消費電力で実現。(SC2007 で論文発表)
2. ソフトウェアを改良し、GRAPE-DR チップの実効性能を大幅に上げました。HPL ベンチマークのコアである行列演算で 94%という高い効率を実現しました。
3. メモリおよび電源を低消費電力・高効率のものに交換。メモリの消費電力を半減、電源では効率が大きく上がり消費電力を減らしました。
4. 空冷システムを採用していることから、6 月時と比較して気温が低下したことで、サーバー室の室温が下がり、消費電力が減少しました。
5. 部品レベルでの細かい無駄電力の削減、ソフトウェアの最適化を更に進めました。

[今後の発展]

GRAPE-DR には、実装方法、冷却方法など更に省電力性能を向上させる余地が多くあります。今後、ソフトウェアの最適化、実装方法、冷却方法を改良し、数十%の更なる性能向上を実現する予定です。

また、2017 年前後に利用可能な半導体テクノロジー、冷却テクノロジーを仮定すると、今回用いた GRAPE-DR アーキテクチャ技術を用い、100Gflops/W レベルの省電力性能を実現することが可能と考えています。

なお本発表は、文部科学省科学技術振興調整費「分散共有型研究データ利用基盤の整備」で得られた成果に基づいています。

4. 参考URL：

Green 500 リスト：<http://www.green500.org/lists/2010/11/top/list.php>

Little Green 500 リスト：<http://www.green500.org/lists/2010/11/little/list.php>

5. 注意事項：

特になし

6. 問い合わせ先：

○東京大学大学院情報理工学系研究科 教授 平木敬

○国立天文台理論研究部 教授 牧野淳一郎

○東京大学大学院情報理工学系研究科 准教授 稲葉真理

7. 用語解説：

1) **Green 500 リスト**：これからのスーパーコンピュータでは、エネルギー効率が最重要である、という見地から、2005 年に始まったプロジェクト。バージニア工科大学の Feng 教授を中心とするグループが 2007 年 11 月から年 2 回発表している。Green 500 リストの対象は HPL ベンチマークで Top 500 にランク入り可能な性能のシステムです。HPL ベンチマークの性能を消費電力で除した演算性能／消費電力比でリストの順を決定しています。

付録 3 に Green500 の規則の概要を示します。（<http://www.green500.org/>）

2) **GRAPE-DR**：東京大学・国立天文台が共同開発したスーパーコンピュータシステム。通常の PC クラスタに独自開発した GRAPE-DR プロセッサを搭載したアクセラレータボードを付加することで、多様なアプリケーションに対して高い性能を実現しています。（<http://grape-dr.adm.s.u-tokyo.ac.jp/>）

3) **HPL ベンチマーク (High Performance Linpack)**：多様なスーパーコンピュータの性能評価法としてもっとも普及し、上位 500 システムの順位付けを出す Top 500 で用いられているベンチマーク。巨大な連立一次方程式を直接法で解くことを要求します。

4) **Mflops/W**：プログラム実行において、1W の電力で実行できる浮動小数点演算回数（単位は 100 万演算）。最新の家庭用 PC を用いると約 250 Mflops/W、携帯用ノート PC を用いると約 500Mflops/W の省電力性能です。

5) **Little Green 500 リスト** : 2009 年 11 月から Little Green 500 リスト として対象となるシステムの最低性能を小型システム側に広げ、過去 19 ヶ月の間に Top 500 にランク入りできる性能があるシステムについて演算性能／消費電力比のリストを作成したもの。2010 年 6 月の Little Green 500 リストでは、GRAPE-DR システムが第 1 位。

6) **水冷技術** : プロセッサチップなどに直接冷水が通る熱交換器を付け、プロセッサが出す熱を冷却する技術。古くはメインフレームなど大消費電力システムの冷却技術として多用されたが、コストやシステム設置が大変なこと、システムの消費電力の低下から使われなくなりました。

最近ではチップ温度が同じ場合にはより少ない電力で冷却を実現し、更にチップ温度を下げれば半導体の微細化に伴うリーク電流が激減することから、省電力化の方法として再び使われるようになってきています。日本の次世代スパコン「京」、IBM の省エネルギー・スパコン「Aquasar」や、開発中の 20Pflops スパコン「BlueGene/Q」、「BlueWater」では水冷技術が使われています。

7) **ラック単位水冷技術** : 前記水冷技術が、水により直接チップを冷却することに対し、ラック単位に液冷熱交換器を入れ、閉鎖されたラック内の空気循環によりプロセッサを冷却する技術。閉じたラック内で熱交換器を含め空気循環させるため、ラック内温度を低く保ち、省電力化するとともに、プロセッサ密度を上げることが可能になります。ラック単位液冷技術は、2004 年に CERN で開発された LHC 加速器用クラスタシステムに導入され、大きな成果を得た後、Cray XT-6、中国の天河-1A、Lenovo の DeepComp7000、TSUBAME 2.0 などで広く使われています。

ラック単位水冷技術には多くの利点がありますが、ラックへ往復する冷媒の配管が必要であり、空冷スーパーコンピュータより設置先に制限が付くことが欠点です。

8) **空冷技術** : プロセッサチップなどコンピュータの構成要素を全てファン等による空気の流れて冷却する放熱技術。コンピュータが誕生して以来、常に最も使われる放熱技術でしたが、乱流を用いた高レイノルズ数領域での衝突冷却の進展により、後記水冷に匹敵する冷却能力が実現し、パソコンからスーパーコンピュータまでの全ての領域で空冷システムが可能となっています。

空冷システムは騒音が出るという問題点を除けばどこにでも設置できる便利さを持ち、現在でも主流の冷却技術となっています。

9) **BlueGene/Q** : 米国 IBM 社が 2012 年に 20Pflops の性能を実現することを目標として開発しているスーパーコンピュータ。1 チップ 16 コアのプロセッサを水冷技術により高度に集積し、96 ラックで 20Pflops を実現する予定です。

10) **IEEE** (The Institute of Electrical and Electronics Engineers, Inc) : アメリカに本部を持つ世界的な電気・電子技術の学会で、活発な標準化活動も行っています。

(写真は全て三浦健司撮影)

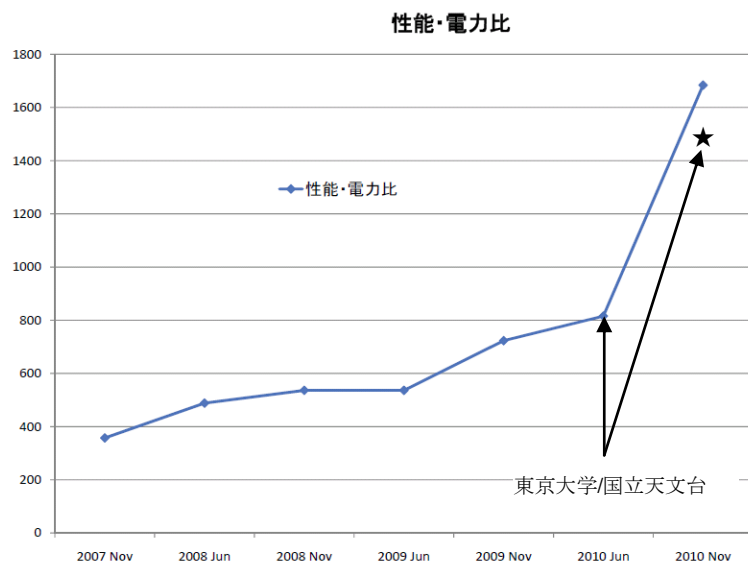


写真1. GRAPE-DR システム



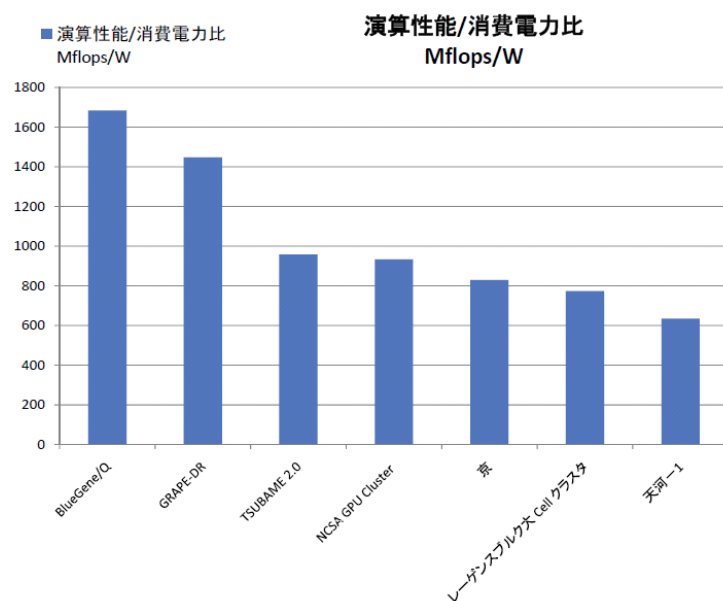
写真2. 今回使用した GRAPE-DR 4 プロセッサボード

(付録 1)
演算性能／消費電力比の推移



年	性能・電力比	HPL 性能 (Mflops)	研究機関名	コンピュータ名
2007/11	357	13.9	Council Daresbury Lab	BlueGene/P
2008/6	488	9.3	Fraunhofer IT WM	BladeCenterQS22 PowerXCell 8 3.2GHz
2008/11	536	18.6	University of Warsaw	BladeCenterQS22 PowerXCell 8 4GHz
2009/6	536	18.6	University of Warsaw	BladeCenterQS22 PowerXCell 8 4GHz
2009/11	722	55.7	Forschungszentrum Juelich	QPACE Cluster PowerXCell 8i 3.2GHz
2010/6	815	23.4	東京大学/国立天文台	GRAPE-DR (Little Green500)
2010/11	1684	65.4	IBM T.J. ワトソン研究センタ	BlueGene/Q

(付録 2) Green 500 リストでの主要システム省電力性能の比較



(付録 3) Green500 の規則。バージョン 0.9 (日本語訳 平木 敬)

- [ルール 1] Green500 リストの対象は、現行 Top500 リストの 500 位システム以上の性能を持つスーパーコンピュータである。
- [ルール 2] Green500 リストの性能は、HPL ベンチマークを実行している時の実行性能 (Flops) を、平均消費電力(W)で除した数値である。
- [ルール 3] Green500 リストに申請する時の HPL 実行性能は、TOP500 リストの 500 位の性能より高い限り、TOP500 リストに掲載される Rmax 性能と一致していなくてよいが、Top500 リスト用性能測定で用いた全てのコアを使うことが必要である。DVFS(可変電圧・周波数による省電力技術)や他の省電力方法を用いて良い。Rmax を実行する時の演算性能/消費電力比と当該スーパーコンピュータ全体での最良の演算性能/消費電力比が異なる場合は、双方を報告することが望ましい。
- [ルール 4] スーパーコンピュータのサブコンポーネントとは、スーパーコンピュータの一部であり、電力を、その部分だけに関して測定できる単位である。電力測定は、1つのサブコンポーネントに対して行うこと。
- [ルール 5] サブコンポーネントの電力は、AC コンセントまたはスーパーコンピュータ用の電力源において測定すること。
- [ルール 6] 測定に用いる AC 電源は、定格の $\pm 1\%$ 以内の電圧、周波数であり、AC 電源電圧の総歪率が 2% THD 以下であること。周囲温度が摂氏 15 度から 25 度の間であり、周囲の相対湿度が 30-80%であること。
- [ルール 7] 測定時の HPL 実行には Top500 リスト申請時のルールが全て適用されていること。Rmax 電力を測定する場合 (それが報告される場合)、電力測定は、Top500 リストへ申請するための HPL 実行と同時にされること。
- [ルール 8] 真の RMS 値 (実効値) が測定でき crest factor (平均電力値に対して測定可能なピーク電力の比率) が 5 以上の電力測定器を使用すること。システム中からランダムに選択され、[4]で規定されるサブコンポーネントの消費電力を測定すること。測定時に使用する電力測定器のフルスケール値を crest factor で乗じた値は、オシロスコープ (またはデジタルマルチメータ) で観測したピーク値より大きいこと。
- [ルール 9] 瞬間的な消費電力を、少なくとも 1 秒間に一回記録して電力測定すること。電力計には別の PC かノート PC が接続され、電子的に値を記録すること。同じ実験を、同じサブコンポーネントに対してもう一回繰り返すこと。もし、同一サブコンポーネントの平均消費電力が 1%以上変化するならば、別のサブコンポーネントを用いて、測定を繰り返すこと。
- [ルール 10] 測定するスーパーコンピュータは、測定前に少なくとも 15 分間電源が入っている状態でウォームアップを行うこと。
- [ルール 11] HPL ベンチマーク実行開始から、HPL ベンチマーク全実行時間の 10%以上の時間が経過後に、電力の測定を開始すること。
- [ルール 12] HPL ベンチマーク全実行時間の 20%以上の期間について電力測定を行うこと。
- [ルール 13] スーパーコンピュータ全体がタスクを実行する際の消費電力を、サブコンポーネントの消費電力にシステム全体におけるサブコンポーネント数を乗じて求めること。