

プロセッサの消費電力測定と低消費電力プロセッサによるクラスタの検討

堀田 義彦[†] 佐藤 三久^{††} 朴 泰祐^{††}
 高橋 大介^{††} 中島 佳宏[†]
 高橋 睦史[†] 中村 宏^{†††}

近年、マイクロプロセッサの消費電力の上昇は著しいものがあり、クラスタなどにおいて高密度な実装が困難になっている。一方、PDA やノート PC に使用されている低消費電力プロセッサの性能が大きく向上している。我々は、低消費電力プロセッサを用いた高密度なクラスタが高性能と低消費電力であることを期待している。本論文では、Pentium4, XScale, Crusoe, Pentium-M の消費電力特性を明らかにし、その結果を基に、低消費電力化手法を示した後、実際に低消費電力クラスタを構築し、評価する。消費電力特性を調べるために我々はホール素子を用いた電力測定環境を構築した。その結果、キャッシュの有効利用が電力量を削減するために非常に有効であることが分かった。Crusoe においては、動作周波数を固定することにより、大きな性能低下なしに電力量を削減できることが分かった。また、Crusoe からなる低消費電力クラスタのプロトタイプで性能と電力効率を調べた。その結果、単一の Pentium4 のような高性能プロセッサを用いるよりも電力的に効率が良くかつ高性能に計算が行えることが分かった。

Measurement of Microprocessor's Power Consumption and Prototyping Low Power Cluster with Low Power Processors

YOSHIHIKO HOTTA,[†] MITSUHISA SATO,^{††} TAISUKE BOKU,^{††}
 DAISUKE TAKAHASHI,^{††} YOSHIHIRO NAKAJIMA,[†]
 CHIKAFUMI TAKAHASHI[†] and HIROSHI NAKAMURA^{†††}

Recently, the power consumption of high performance processors is rapidly increasing, so that it makes high-density packaging difficult when building compact clusters. On the other hand, the performance of low power processors used for PDA or PC is being improved, rapidly. It is expected that high-density clusters will be realized by using the low power processors for power-aware computing. In this paper, we firstly examine the characteristics of power consumption of Pentium4, XScale, Crusoe and Pentium-M. To measure the actual power consumption, we have built environment by Hall device. We found that the total power consumption can be reduce by optimizing memory access. In Crusoe, while it has DVS facility to control voltage dynamically for power reduction, we found that by controlling clock frequency manually, have a possible to the total power consumption can be reduced without performance loss. And we have built prototyping low power clusters with Transmeta Crusoe TM-5800. We have measured the performance and the power efficiency of this low-power cluster using some parallel benchmarks. The result shows that a low power cluster can achieve a good power efficiency than a single high performance processor such as Pentium4.

1. はじめに

近年、マイクロプロセッサの性能は著しく向上して

いるが、高性能化とともに実装面において消費電力を低く抑えるニーズも増している。HPC の分野でプラットフォームとして主流になりつつあるクラスタなどの並列システムにおいても電力の急増による発熱量の増加のためにプロセッサは大きな冷却装置が必要になっており、クラスタを高密度に実装する場合の問題になっている。

一方で、従来 PDA やノート PC 向けであった低消費電力プロセッサの性能が向上してきている¹⁾。低消費電力プロセッサを用いることにより、ファンや冷却に対する制限が少なくなり実装密度の向上が期待でき

[†] 筑波大学大学院システム情報工学研究科

Graduate School of Information and Sciences Engineering, University of Tsukuba

^{††} 筑波大学電子・情報工学系

Institute of Information Science and Electronics, University of Tsukuba

^{†††} 東京大学先端科学技術研究センター

Research Center for Advanced Science and Technology, The University of Tokyo

る．本論文では，このような背景をふまえ，実際に，いくつかのマイクロプロセッサの消費電力を測定した．その結果に基づき，低消費電力プロセッサを用いたクラスタについて，検討する．

計算とそれとともなう消費電力に焦点を当てた研究には，LosAlamos 研究所の Feng らのグループによる低消費電力スーパーコンピュータ「Green Destiny」がある．彼らはこれにより高密度のコンピュータの評価を行い，コストや体積あたりの性能，電力あたりの性能が既存のスーパーコンピュータよりも高くできることが報告されている²⁾．また，BlueGene/L³⁾は低消費電力プロセッサを高密度に実装することによって非常に高い性能と省電力・省スペースを実現している．

マイクロプロセッサの電力計測について，我々はホール素子を用いた電力計測システム環境を使用し，数種の CPU で様々なベンチマークを実行し実際の消費電力を計測した．さらに，実際にそれらのプロセッサを用いたクラスタを試作し，電力・性能について評価を行った．従来，プロセッサの消費電力については，多くの研究がなされているが，それらはプロセッサ単体などに限定され，PC クラスタのようなシステムの全体の消費電力を実際に測定している例は他にない．低消費電力クラスタの検討には，クラスタ全体の消費電力を明らかにする必要があり，プロセッサだけの測定やシステムのシミュレーションでは不可能である．

電力の指標として，瞬間の消費電力と電力量の 2 つの電力に注目する．瞬間の消費電力は，冷却に対して非常に重要である．瞬間の消費電力が高いと大きな冷却装置が必要になり，空間を圧迫し，実装密度を低下させる主要因となる．電力量は瞬間の消費電力とサンプリングを行った時間間隔との積をとったものである．瞬間の消費電力が低くても，電力量が上昇してしまっただけでは適さない場合があり，この値を低くすることで電力あたりの性能を高くすることができる．また，電力性能の基準として様々な指標が存在する^{9),10)}．我々の目的は高密度で高性能かつ低消費電力を実現することであるから，必要な指標としては PDP (Power Delay Product) を用いるのが冷却・実装の面から望ましいが，プロセッサの性能を加味するために EDP (Energy Delay Product) を用いる．この指標は電力と性能のバランスがとれており評価指標として適している．

本論文の目的は，定性的には様々に知られている消費電力について定量的に評価を行うことである．また，消費電力の評価はプロセッサのためのものでなく，システムの評価を行うことを目的としている．シミュレーションにおいてはシミュレータでプロセッサは定量的

に評価されているが，システムの評価をシミュレーションで行うことは難しい．従来，高密度なシステムを構築する際，TDP (Thermal Design Power) の値が目安とされているが，実システムを定量的に測定することによってより現実的な指標が得られる．我々の目指す目標は，高性能な 1U2CPU のシステムよりも電力を維持しながらスペースあたりの性能を上げることである．また，実際の消費電力が TDP と比べてどのようになるかを確認する．

本論文では低消費電力のための新たなアルゴリズムを提案することではなく，既存のアルゴリズムを用いた場合，消費電力がどのように変化するかを定量的に評価するために行う．

2 章では我々の構築した電力測定環境について述べる．3 章ではプロセッサの電力の測定と電力特性について述べる．4 章ではプログラムの最適化による消費電力の削減について述べる．5 章では Crusoe において動作周波数を固定した場合の消費電力の削減について述べる．6 章では低消費電力プロセッサを用いたクラスタの性能と消費電力について述べる．7 章では結果からの考察を述べる．最後に結論と今後の展望について述べる．

2. 電力測定環境

クラスタに用いるプロセッサを調査・評価するにあたって，現在ある各 CPU の電力消費特性を調べるために，(株)シナジェティック社製 CT-30000 を用いた．図 1 に電力測定環境を示す．この装置はホール素子，接続 BOX，A/D コンバータから構成されており，ATX 電源の各電圧の電線に流れる電流をダイナミックに測定が可能である．このようにホール素子の間に電線を通すだけでよく，取扱いが容易であるのもこの装置の特徴である．この装置を用いてプログラム実行時の消費電力を測定した．電力量の単位を [Ws] とする．

ATX 電源には +12V, -12V, +5V, -5V,

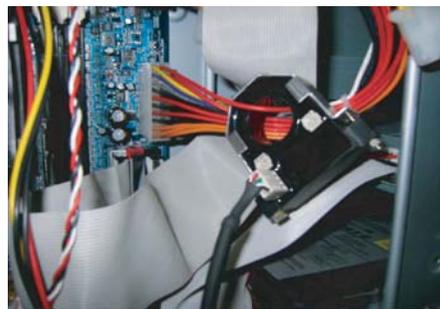


図 1 ホール素子
Fig. 1 Hall device.

表 1 測定 PC の仕様と無負荷時の消費電力
Table 1 Specifications and power consumption of measured systems.

システム名	P4	XScale	Crusoe	PenM
マザーボード	GA-81RX	NPWR	EB5	PFU
CPU	Pentium4	IOP-80200	TM-5800	Pentium-M
Clock	1.80 GHz	733 MHz	933 MHz	1.60 GHz
Cache L1/L2	8 KB/512 KB	64 KB/none	64 KB/512 KB	64 KB/1MB
Memory	512 MB(DDR)	256 MB(SDR)	256 MB(SDR)	512 MB(DDR)
kernel	2.4.18	2.4.18	2.4.22	2.4.18
Compiler	gcc3.2	gcc2.95.2	gcc3.2	gcc3.2
TDP	66.1 W	1.3 W	9 W	25 W
Power (12 V)	15 W (CPU)	0.5 W	1 W	8 W(CPU)
Power (5 V)	1 W	9 W (CPU)	3 W (CPU)	4 W
Power (3.3 V)	3 W	none	0.3 W	2 W

+3.3 V の各電圧があるが、測定の結果 +12 V, +5 V, +3.3 V 以外はほとんど電流が流れていないので測定の対象から外した。また Pentium4 や Pentium-M では CPU への電力供給に +12 V を変圧したものが使われている。他の CPU では +5 V を変圧したものが使用されている。

3. マイクロプロセッサの消費電力の測定

3.1 対象とした CPU

表 1 に測定対象のシステムの仕様と各システムの無負荷時の消費電力を示す。一部のシステムでメモリのサイズが 512 MB であるが、今回のベンチマークではメモリのサイズは 256 MB で十分であり、512 MB であることは性能において問題ない。また、すべてのシステムは DIMM モジュール 1 枚で構成されており、消費電力の面において待機電流に大きな差はなく、実質運用されているのは 256 MB 相当のセル数で問題ない。このことからメモリのサイズの違いは本質的には問題ない。測定を行う CPU は、Pentium4, XScale, Crusoe, Pentium-M の 4 つである。以下 Pentium4 を P4, Pentium-M を PenM と表記する。

XScale⁴⁾ は StrongARM シリーズの上位互換 CPU で Intel が 2000 年に発表した低消費電力 CPU である。主な用途は PDA などの携帯デバイスである。現在は FPU (浮動小数点演算装置) が搭載されていないが、将来的には搭載されることも予定されている。組み込み向け低消費電力プロセッサの代表として XScale を選択した。

Transmeta 社の Crusoe TM-5800⁵⁾ は VLIW アーキテクチャを採用している CPU である。この CPU は CMS (Code Morphing Software) により x86 命令を独自の VLIW 命令 (128 bit) に変換し実行することによって x86 命令を使用する CPU との互換性を備えている。また、LongRun と呼ばれる DVS (Dy-

namic Voltage Scaling)⁶⁾、動的に動作電圧を変化させる機能により、必要な負荷に応じて動作周波数を変更し低消費電力化を実現している。LongRun を用いることによる低消費電力化を試みるために Crusoe を選択した。

P4 は現在の主流な CPU と低消費電力 CPU との比較のために測定を行った。

PenM は現在のノート PC のプロセッサとして主流な低消費電力かつ高性能プロセッサである。このプロセッサが高密度実装において性能と低消費電力をどのくらい満たすかどうか調べるために測定を行った。

3.2 ベンチマークプログラム

前述のシステムにおいていくつかのベンチマークを実行し、性能と電力消費の振舞いを定量的に動的に確認した。測定に以下のベンチマークを用いた。

- datascan
- dhrystone
- NPB (NAS Parallel Benchmark) version 2.3
- matrix multiply ブロッキング版
- LU 分解ブロッキング版

3.2.1 datascan

datascan は連続したメモリ領域を順にアクセスするプログラムで、その領域のサイズを順に大きくすることにより、キャッシュヒット/ミス時の消費電力の変化を見ることができる。ループではデータを使って整数の saxpy の演算を行っている。このプログラムを各システム上で実行し、キャッシュヒット時とミスヒット時での性能の違いと電力差について確認した。図 2, 図 3, 図 4 にそれぞれのプロセッサでの消費電力の変化を示す。

P4 の場合 (図 2), プログラムを実行すると急激に消費電力が上昇する。一方、後半部分において消費電力が減少している。減量する間隔が徐々に長くなるのは、アクセス領域の増加にともない、キャッシュから

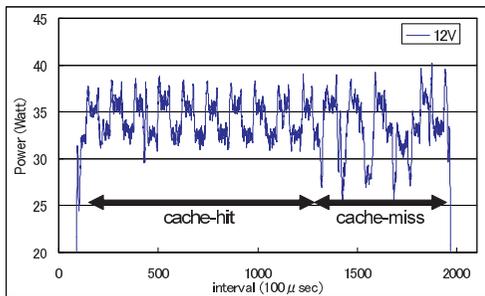


図 2 P4 12V datascan 実行時電力変化
Fig. 2 P4 power consumption.

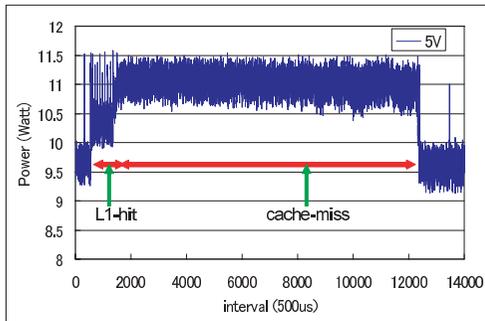


図 3 XScale 5V datascan 実行時電力変化
Fig. 3 XScale power consumption.

外れている時間が增大するためである。また、キャッシュに当たるときと当たらないときの電力差は数 W と少ない。キャッシュヒットしない場合は、性能が著しく低下することから電力量はキャッシュにヒットし続ける方が少なくて済む。

XScale の場合 (図 3), P4 とは異なり、プログラムの開始直後のキャッシュにヒットしている部分での消費電力の方が、メモリアクセスしている部分よりも少ない。これは、CPU の消費電力が非常に小さく、メモリの消費電力の方が大きいためである。キャッシュにヒットしない場合は、消費電力が高く、実行時間も急増するため電力量は増大する。

Crusoe の場合 (図 4), プログラム実行時に消費電力は段階的に無負荷時と比べて数 W 上昇する。これは動作周波数が段階的に上昇しているためである。他のプロセッサと異なり、消費電力が階段状に減少する。これはまず、キャッシュから外れたために電力が減少する。そして待機時間が長くなるために、プロセッサは動作周波数を落とすことによってさらに消費電力が減少するためである。キャッシュヒット時とミスヒット時の消費電力の差は約 2W で、実行時間も急増するため電力量はキャッシュヒットし続けるほうが少なくて済む。

PenM の場合 (図 5), 消費電力の振舞いは P4 に似ている。実行時に上昇し、キャッシュミス時の方が消

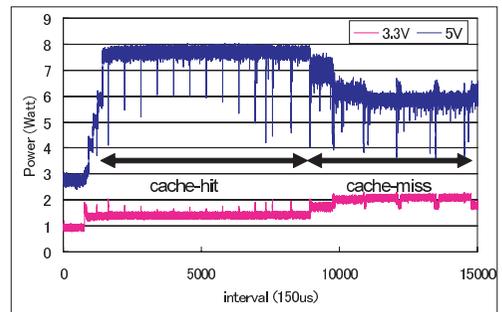


図 4 Crusoe datascan 実行時電力変化
Fig. 4 Crusoe power consumption.

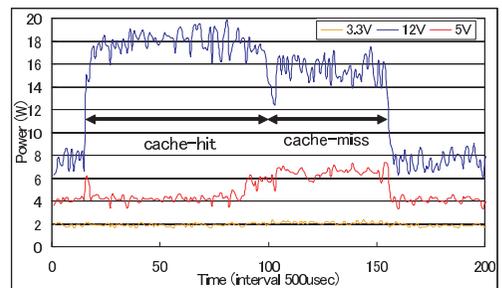


図 5 PenM datascan 実行時電力変化
Fig. 5 PenM power consumption.

費電力は低い。キャッシュヒットとミスヒットの消費電力の差は約 2W と少なく、P4 と同様に電力量はキャッシュヒットし続けるほうが少なくなる。

3.2.2 dhrystone

dhrystone とは整数演算や文字列コピーを繰り返し実行するベンチマークである。プログラム全体がキャッシュに収まる場合の各システムでのパフォーマンスと電力量を比較した。図 6 に実行時間と電力量のグラフを示す。P4 は高速に実行が可能である一方で電力量が最も多くなっている。XScale はこのプログラムにおいて P4 との性能差は大きい。しかし、このようにキャッシュにヒットし続けるプログラムの場合、XScale P4 に比べて少ない電力量で実行できる。Crusoe では実行時間は P4 に及ばないが、電力量は約 4 分の 1 となっており電力あたりの性能では P4 よりも良い。PenM は最も高い性能を示し、電力量も少なくなっている。電力あたりの性能では最も良い結果となる。

3.2.3 NPB IS

実用的なアプリケーションで性能と消費電力を比較するために逐次版の NPB の IS で評価を行った。IS はソートの実行部分のみで電力量の評価を行った。表 2 に IS の各クラスでの性能を示す。図 7 に IS の CLASS W でのソート部分の実行時間と電力量を示す。P4 は性能が非常に高いが XScale では非常に性能

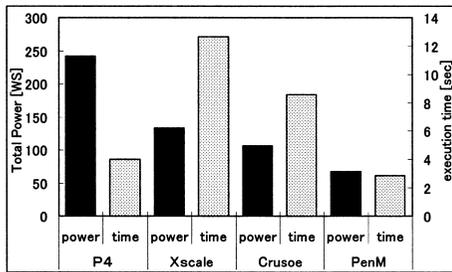


図 6 dhrystone 電力量と実行時間

Fig. 6 Execution time and the total power consumption in dhrystone.

表 2 IS の性能 [Mop/s]
Table 2 Result of IS [Mop/s].

CLASS	S	W	A
P4	73.22	50.69	7.82
XScale	5.12	1.40	1.04
Crusoe	30.25	17.57	3.09
PenM	204.93	60.15	22.53

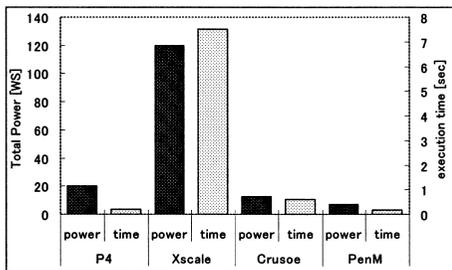


図 7 IS (CLASS W) の電力量と実行時間

Fig. 7 Execution time and the total power consumption in IS (CLASS W).

が低くなっている．特に CLASS W 以降の性能低下が大きく、キャッシュ容量の少なさが効いていると思われる．Crusoe は性能も良く、電力量も少ない．PenM は最も高性能で、電力量も少ない．電力あたりの性能も PenM が他よりもはるかに良い結果となる．

3.3 各 CPU の電力消費特性

以下に各 CPU の電力特性をまとめる．

—P4 は高性能であるが瞬間的な電力は非常に高いものであった．また、プログラムがキャッシュヒットしているときとしていないときで、数 W の違いが見られた．P4 の電力特性としては、より高速にプログラムを実行することで全体の電力量が少なくなる傾向がある．他の CPU と比べて、瞬間の電力は急激に上昇するが、電力量は少なくなる．

—XScale では、キャッシュヒット時の消費電力の方がミスヒット時より少なくなっている．これは、CPU の消費電力の上下よりもメモリの消費電力の上下の方がはるかに大きくなるのが原因である．また、ミスヒッ

ト時にヒット時と比べて急激に性能が低下することも電力量が増大する原因となっている．これにより電力あたりの性能は非常に低くなっている．キャッシュが有効利用され、整数演算の多いプログラムの実行においてのみ、低消費電力でプログラムを実行できる．

—Crusoe では実行時に消費電力が数 W しか上昇しない．他のプロセッサとは異なり、消費電力が実行時に段階的に変化する．また、キャッシュヒット時とミスヒット時の消費電力の差は約 2 W であり、キャッシュミス時には実行時間が急増することから、総合的に電力量は大きくなる．

他の CPU と比べて消費電力は最大で 9W 程度と少なく、電力あたりの性能も高い．電力あたりの性能が非常に高く、冷却の必要がないほど瞬間の消費電力が低い．

—PenM では、実行時に消費電力が上昇する．しかし、P4 ほど上昇は大きくない．また、性能は今回用いた P4 よりも高く、低消費電力かつ高性能を実現している．しかしながら、熱設計は消費電力の最大値を考慮して行わなければならない、Crusoe に比べて消費電力の最大値が高いため大きな冷却装置による冷却を考えなければならない．そのために高密度実装には不向きである．

4. プログラム最適化による消費電力の低減

これまで述べたように、キャッシュミス時には消費電力は減少するが、減少の度合いは性能の低下の度合いに比べると少なく、電力量を増加させることから、低消費電力 CPU ではなるべくキャッシュを有効利用することが、電力量の削減につながるということが分かった．そこでキャッシュブロッキングを行うことによってキャッシュを有効に利用するときの電力量の変化を確認した．ブロッキングは定性的にはよく知られているアルゴリズムであり、これを行列積に施すことによって電力量と性能に定量的にはどのような変化があるのかを評価するために行った．対象としたプログラムは、単純な行列積と LU 分解のプログラムである．ここでは P4、と Crusoe に限定する．

4.1 行列積

ここまでの測定の中で、キャッシュミスヒットによるメモリアクセスは消費電力が上昇するだけでなく、減少することが分かった．このことから実際にオフチップアクセスが起こる場合の電力量の変化はどの程度であ

一例として Transmeta 社によると冷却ファンが必要かどうかの境界線は CPU の消費電力が 7W 以下であるかどうかである⁷⁾．

表 3 ブロッキングサイズごとの実行時間 [sec] (括弧内は電力量 [Ws])
Table 3 Execution time in each block size [sec] (the total power consumption in parenthesis [Ws]).

bsize	none	32	64	128	256	512
P4	16.5 (802.40)	8.8 (460.21)	8.3 (432.04)	7.7(391.54)	11.7 (582.63)	14.3 (703.14)
Crusoe	93.7 (1071.40)	13.7 (198.18)	12.4 (182.65)	12.3 (172.59)	13.0 (189.42)	21.3 (285.08)

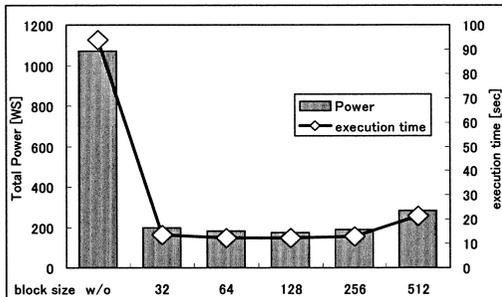


図 8 Crusoe での行列積の電力量と実行時間

Fig. 8 Matrix multiply result by blocking in Crusoe.

るのかを確認するためにプログラムにキャッシュブロッキングを施し、同一作業量におけるオフチップの回数が多いかどうかによる電力量の変化を確認する。このために通常の行列積とキャッシュブロッキングを施したものを実行し、実行時間、電力量を比較、検証を行う。

キャッシュブロッキングは、空間的局所性を向上させることによりキャッシュを利用する割合を多くし、性能を向上させる手法である。このブロッキングのサイズを変化させ、実行時間と電力量の変化を確認した。行列のサイズは 1024 である。

図 8 に Crusoe における各ブロッキングサイズでの実行時間と電力量を示す。表 3 に各 CPU でのブロッキングを行わない場合と行った場合の実行時間と電力量を示す。すべての CPU で効果があるが、特に Crusoe で効果が大きい。これは、ブロッキングを行う場合、計算を実行している間、ほとんどキャッシュにヒットしているのが原因である。一方で P4 では実行時間が減少はするが約半分と減少の割合が少ない。これは、Crusoe ではメモリアクセスの速度の差や、分岐予測の精度の差にとまなう、キャッシュミスによる性能低下が大きいことを示している。また、ブロックサイズが 512 のときに性能が下がるのは表 1 に示してあるように、キャッシュのサイズが 512 KB であるためである。この結果から、低消費電力 CPU にとっては、低消費電力化には、実行時間短縮が効果的である。キャッシュの有効利用は、低消費電力 CPU にとっては実行時間短縮に効果的である。

4.2 LU 分解

ブロッキングされた LU 分解のプログラム⁸⁾を用いて、電力量と性能を測定した。表 4 に性能と電力量

表 4 LU 分解における電力量 (括弧内は性能 [MFlops])
Table 4 The total power consumption in LU calculation ([MFlops] in parenthesis).

	w/o blocking	with blocking
P4	270.3 (138.4)	92.2 (442.8)
Crusoe	119.7 (56.2)	37.6 (174.5)

の結果を示す。ブロッキングを行うことにより約 3 倍性能が向上する。性能に反比例するように電力量はブロッキングを行うと 3 分の 1 に削減できている。前述の行列積と同じように、キャッシュのヒット率を上げることにより実行時間を減らすことが低消費電力化への有効なアプローチであることが確認できた。

5. DVS と消費電力

5.1 DVS の機能

Crusoe には LongRun とよばれる、DVS の機能が備わっている。この機能はプロセッサが動作する必要のないときは、電圧を下げることによって動作周波数を下げ消費電力を抑える働きを持っている。また、高動作周波数で連続動作をしている際に、発熱量が増大するときには、自動的に電圧を下げる。瞬間の消費電力を制限することができ、また発熱量が多いときはそれを抑えるように動作するため、ファンなどの装置が必要なく。これにより低消費電力システムにとって、高密度な実装が可能になっている。PenM にも同様に“Speed Step”と呼ばれる機能が備わっているが、今回は Crusoe のみの評価を行った。

5.2 DVS による動作周波数変化と消費電力

DVS をコントロールし、Crusoe の動作周波数を固定することによって、性能・電力量にどのような変化が見られるのか、また実行するプログラム内で動的に周波数をコントロールすることによって、電力量を削減する可能性があるのかを定量的に調べるために性能と電力量の評価を行った。

表 5 に動作周波数を固定した場合の動作周波数と FSB の動作周波数、電圧を示す。これ以降、図・表において DVS と表記されているものは動作周波数を固定していない場合を指す。また、933 MHz に固定した場合、熱暴走によりシステムがダウンするという症状がおきるため、測定の対象から除外した。

各動作周波数においてキャッシュヒットがどのよう

表 5 Crusoe の動作周波数と電圧

Table 5 Clock frequency and voltage in Crusoe.

Clock	Voltage	FSB
933 MHz	1.35 V	133 MHz
800 MHz	1.25 V	133 MHz
667 MHz	1.20 V	133 MHz
533 MHz	1.10 V	133 MHz
300 MHz	0.90 V	100 MHz

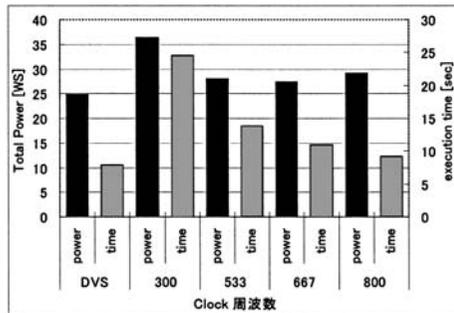


図 9 dhrystone 実行時の各動作周波数ごとの性能と電力量
Fig. 9 Execution time and total power consumption of dhrystone in each clock frequency.

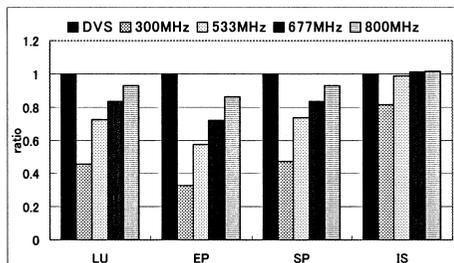


図 10 NPB CLASS A 実行時の各動作周波数ごとの性能
Fig. 10 Performance ratio of NPB A in each clock frequency.

な影響を及ぼすかを確認するために dhrystone を用いて評価を行った．図 9 に dhrystone における実行時間と電力量を示す．この結果から，キャッシュヒットを頻繁にする場合，動作周波数は高いほうが性能・電力量ともに良い結果になることが分かる．次に，実用的なアプリケーションにおける電力量を調べるために，この各動作周波数で NPB の CLASS A の LU，IS，EP，SP を実行した．また，Crusoe において動作周波数を固定することによって省電力化につながるかどうかを調べた．図 10 に NPB における動作周波数を固定しない場合を基準とした各動作周波数での性能の比を示す．動作周波数を 300 MHz にしたときに大きな性能低下が見られる．これは，メモリに対する FSB の動作周波数が他の動作周波数のときと比べて下がるのが原因であると思われる．また，IS では性能低下の割合が低く，逆に LU などでは性能低下が大き

表 6 動作周波数ごとの NPB A の電力量 [Ws]

Table 6 The total power consumption of NPB A in each clock frequency.

	IS	LU	SP	EP
DVS	295.9	9844.1	12568.9	1941.9
300 MHz	271.6	10895.3	13343.0	2742.2
533 MHz	230.5	10272.7	11626.9	2115.4
667 MHz	259.8	13152.7	10806.1	2199.7
800 MHz	252.5	12135.9	11979.0	1910.9

い．動作周波数を下げた場合には，CPU が計算を行うようなアプリケーションでは性能低下が大きく，メモリアクセスが頻繁に起きるアプリケーションでは性能低下が低いということが分かる．

表 6 に動作周波数ごとの各アプリケーションの電力量を示す．この結果から性能低下の低いアプリケーションのほうが，電力量が低い傾向にあることが分かる．メモリアクセスが頻繁なアプリケーションにおいては，動作周波数を固定することによって省電力化が可能であるが，LongRun の動作はおおむね，消費電力と性能について有効に動作していることが分かる．

6. 低消費電力 CPU を用いたクラスタの構築

低消費電力クラスタを構築するにあたって，ここまでの測定の結果より XScale はクラスタとするには性能が非常に低く，またキャッシュの容量が少ないことから大きなプログラムになると電力性能比が著しく低下するために低消費電力クラスタの CPU として採用するには不向きであると考え，Crusoe をクラスタに用いる CPU の候補とした．そこで実際に Crusoe クラスタを構築し，評価をした．また，PenM に関しては，確かに高性能かつ，低消費電力であるが我々が目的とできるだけ高密度な実装による低消費電力，高性能クラスタのアプローチには大きな冷却装置が必要な PenM は高密度実装が困難になるため採用しなかった．このクラスタは表 1 にある Crusoe のシステムを 4 台，Fast Ethernet で接続したクラスタである．クラスタの各ノードの NIC はオンボードのものをを用いており，クラスタの消費電力の測定は NIC の消費電力も含まれている．スイッチの消費電力に関しては製品によって差があり，固定的な電力であるため評価には加えなかった．このことに関しては 7.3 節で考察する．

また，このクラスタは Diskless boot で動作するようにし，Disk 装置などの余分な消費電力を省いている．我々の目的は低消費電力・高密度クラスタを構築することであるが，このことは実行速度とのトレードオフの関係となる．できるだけ性能を落とすことなく，低消費電力・高密度実装を実現する必要がある．そこで，

表 7 クラスタにおける IS の電力量 (括弧内は性能 [Mop/s])
Table 7 The total power consumption of IS in cluster ([Mop/s] in parenthesis).

	Power[W]s ([Mop/s])	電力性能 [Mop/s/Ws]
2node	257.5 (4.28)	0.0166
4node	570.3 (3.98)	0.0069
P4	1032.5 (7.82)	0.0075

既存の高性能プロセッサを用いたシステムよりも電力性能 (電力あたりの性能) で上回ることを目標とする。

6.1 並列ベンチマーク NPB での消費電力評価

低消費電力クラスタとして既存の高性能なシステムとの比較のために新たに電力性能という指標を用いる。これは計算の効率を示す指標であり、以下のように定める。

電力性能 = 性能 [Mop/s など] / 電力量 [Ws]

これは ED 積である。電力性能の指標には様々なものがある^{9),10)}。まず PDP (Power Delay Product) は、低消費電力システムにとって非常に重要な指標ではある。PDP の値は冷却システムの必要性に直結し、この値を小さくすることによって大きな冷却装置を除くことができる。しかし、性能が指標に含まれにくく高性能と高密度を実現しようという目的に反するためこれのみではシステムの評価はできない。ED 積は指標としてはトレードオフになるが、システムの性能を加味するために用いることにした。以降、いくつかの評価ではこの評価指標を用いることにする。

NPB の IS, LU で性能を計測した。表 7 に IS の CLASS A での 2node, 4node, P4 における性能と電力量, 電力性能を示す。台数が増えても性能に変化がなく, 結果として電力量も大きくなっている。IS は並列化がうまくいっておらず台数が増えても効果のないプログラムである。電力性能で比較を行うと, 2node の場合が最も良く, P4 の約 2 倍の結果となる。

表 8 に LU CLASS A での Crusoe クラスタの性能と電力量, 電力性能を P4 の場合と比較したものを示す。台数が増えると性能が大きく向上している。また P4 よりも高性能でかつ電力量が小さくなる。このような負荷分散がうまくいき, かつ計算の多いプログラムにおいては低消費電力クラスタは非常に良い結果を示す。電力性能を比較すると, P4 に比べて Crusoe クラスタは 58% の向上が見られる。

この結果から, LU のような並列性の高いプログラムでは, Crusoe クラスタにおいて非常に効率の良く計算が実行できるのが分かる。

6.2 ブロッキングによるクラスタの低消費電力化
キャッシュの有効利用が低消費電力クラスタにおい

表 8 クラスタにおける LU CLASS A の電力量 (括弧内は性能 [Mop/s]) と電力性能

Table 8 The total power consumption of LU in cluster ([Mop/s] in parenthesis).

	Power[W]s ([Mop/s])	電力性能 [Mop/s/Ws]
2node	15634.2 (192.3)	0.0122
4node	15249.3 (351.7)	0.0231
P4	19532.4 (286.4)	0.0146

表 9 並列版ブロッキング行列積実行時間 [sec]

Table 9 Execution time of parallel matrix multiply by blocking [sec].

bsize	w/o	32	64	128	256	512
single	93.7	13.7	12.4	12.3	13.0	21.3
2node	108.6	6.1	5.4	5.4	6.0	9.8
3node	109.1	6.1	5.4	5.4	6.4	9.9
4node	54.6	3.4	3.0	3.0	3.6	5.2

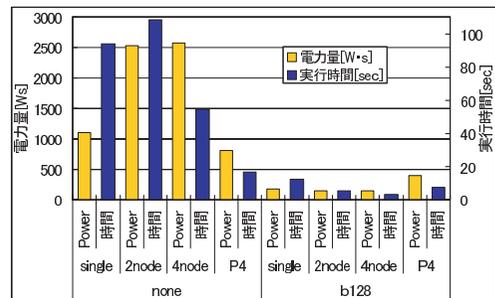


図 11 blocking 有無による Crusoe クラスタと P4 の電力量と実行時間の比較

Fig. 11 The total power consumption and execution time with or w/o blocking.

てどれくらい性能・電力量に影響があるのかを確認するために行列積を並列化し, サイズが 1024 の行列にブロッキングを行い, キャッシュの有効利用が低消費電力クラスタにおいてどれくらい影響があるのかを調べた。

表 9 に各ノード数でのブロッキングサイズごとの実行時間を示す。single の結果は MPI を使用していない完全な逐次版での結果である。このプログラムは 2 次元分割を行っており, 3 ノードではうまく並列化ができていない。そのために 3 ノードにおいて性能が向上していない。図 11 に Crusoe クラスタと P4 における, ブロッキングサイズ 128 のときとブロッキングを施さなかったときの実行時間と電力量の比較を示す。これらの結果から, 並列化によってわずかな電力量の上昇で, 大幅な性能向上が実現できていることが確認できる。これらのことからクラスタにおいてもブロッキングは, 大幅な低消費電力化には有効な手法であることが分かる。電力性能においても非常に良い結果となる。

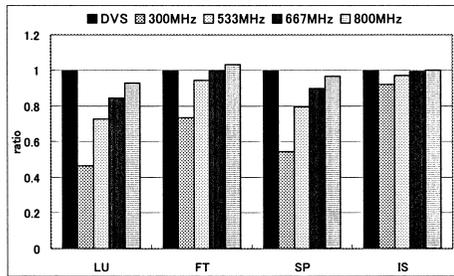


図 12 動作周波数ごとの NPB における性能の変化

Fig. 12 Performance ratio of NPB in each clock frequency.

表 10 動作周波数ごとの電力量の比較 [Ws] (括弧内は電力性能 [Mop/s/Ws])

Table 10 The total power consumption in each clock frequency ([Mop/s/Ws] in parenthesis).

	IS	FT
DVS	570.30 (0.00698)	1630.58 (0.07121)
300 MHz	488.38 (0.00752)	1500.22 (0.05701)
533 MHz	481.80 (0.00803)	1413.82 (0.07745)
667 MHz	509.14 (0.00776)	1476.20 (0.07836)
800 MHz	542.18 (0.00736)	1571.80 (0.07618)

6.3 クラスタでの DVS による省電力化

動作周波数を固定した場合、クラスタにおいて性能と電力量にどれくらいの影響があるかを確認するために NPB で性能と電力量を測定した。図 12 に動作周波数を固定しない場合と、各動作周波数での性能を DVS の性能を 1 とした場合の比率を示す。逐次のときと同様に、FT や IS などメモリアクセスが多いアプリケーションにおいてクロック周波数を下げても性能低下が抑えられているのが分かる。

電力量を下げることもできて、性能が大幅に低下しては、電力あたりの性能が低くなってしまふ。そこで動作周波数を低く固定しても性能がそれほど変化しないならば、瞬間的な消費電力の低いほうが電力量の削減が期待できる。そこで性能低下の少ないアプリケーションとして IS と FT に注目し、動作周波数ごとに電力量を求め、動作周波数を固定しない場合と比較を行った。表 10 に IS と FT における動作周波数ごとの電力量と電力性能を示している。動作周波数を固定することによって十数%の省電力化を実現している。また電力性能を見ると約十数%の効率化ができています。図 13 に IS と FT における動作周波数を固定しないときを基準とした電力性能比を示す。

これらの結果から、動作周波数を固定することによってメモリアクセスが頻繁に起きるアプリケーションでは、以下のことが推測される。

- 性能低下が少ない。
- 電力量を約十数%削減可能。

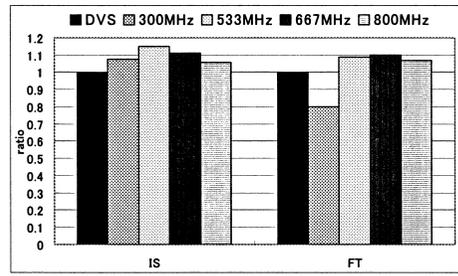


図 13 動作周波数ごとの IS, FT 実行時の電力性能比

Fig. 13 Power performance ratio of IS and FT in each clock frequency.

表 11 動作周波数ごとの行列積行列サイズ 2048 の電力量 (括弧内は性能 [MFLOPS])

Table 11 The total power consumption of matrix multiply ([MFLOPS] in parenthesis).

	電力量	電力性能 [MFLOP/s/Ws]
DVS	49699.5 (9.78)	0.000196
300 MHz	41758.6 (6.84)	0.000164
533 MHz	38892.1 (8.96)	0.000230
667 MHz	40952.3 (9.45)	0.000231
800 MHz	41119.8 (9.41)	0.000229

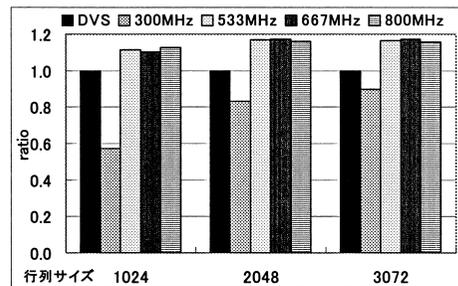


図 14 動作周波数ごとの行列積の電力性能比

Fig. 14 Power of matrix multiply performance ratio in each block frequency.

- 電力性能比を向上させることができる。

NPB の結果より、メモリアクセスが多いならばクロック周波数を低く固定しても性能低下が少ないことが分かった。そこでメモリアクセスとクロック周波数によるバランスを定量的に調べるために、メモリアクセスが頻繁に起きる例として最適化を施していない行列積のプログラムで実証を行った。表 11 に動作周波数ごとの電力量と性能を示す。図 14 に、動作周波数を固定しない場合の電力性能を 1 とした場合の動作周波数ごとの電力性能の比を示す。これらの結果から以下のことが分かった。

- 動作周波数を固定しても性能は大きく低下しない。
- 電力量を最大 22%削減。
- 電力性能比を約 1 割向上。

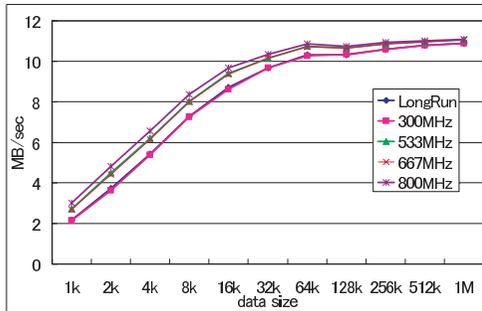


図 15 動作周波数ごとの通信バンド幅性能

Fig. 15 Performance of bandwidth in each clock frequency.

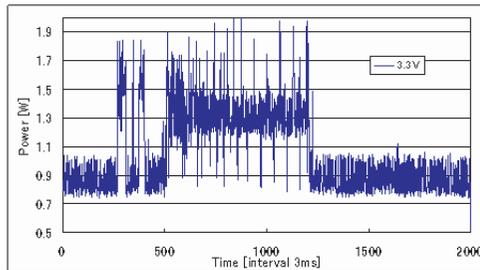


図 16 DVS 使用時の通信時消費電力

Fig. 16 Power consumption of communication using DVS.

これによって、メモリアクセスが頻繁に起きるアプリケーションにおいて動作周波数を低く固定することが、性能を維持しながら電力量を削減でき、その結果電力性能比が向上することが定量的に確認された。

6.4 通信の消費電力と性能

動作周波数を固定することによって、クラスタの通信性能と消費電力がどのように変化するか評価を行った。図 15 に動作周波数ごとの通信性能を、図 16 に通信時の消費電力を示す。図 15 の結果から、DVS を用いるよりも動作周波数を固定することによって通信バンド幅が約 1 割向上する。また図 16 の結果から、通信時の消費電力の上昇は約 1W と非常に小さいことが分かる。

7. 考 察

7.1 クラスタの消費電力

本論文では低消費電力 CPU である Crusoe を使用してクラスタを構築し評価を行った。

Crusoe クラスタは、性能も大きく向上し電力量も小さくすることができた。

科学技術計算の分野であれば、プログラムが並列化してある場合が多く、性能が低くても低消費電力な CPU を多数接続したクラスタは性能・電力両方の面で

高性能で高消費電力な Pentium4 のようなプロセッサを用いるより性能が良い場合がある。また、動作周波数周波数を固定することにより、瞬間の消費電力を下げると同時に、性能が高く、電力量が少ない場合がある。

7.2 冷却・実装

Pentium4 と Pentium-M は高速に実行が可能であるが、Crusoe に比べ瞬間の消費電力が非常に大きくなり、発生する瞬間の熱量が多くなる。よって、定常的に冷却をしなければならず、大きな冷却装置が必要になる。これが高性能 CPU を高密度で実装する場合の最大の問題である。一方、Crusoe や XScale はどれだけプログラムを実行しても、CPU 自体にはファンなどの冷却装置が必要ないため、高密度な実装が可能になる。また、これらの低消費電力 CPU は使用されないときは通常時より電力を下げることで、定常的な電力を大幅に削減できる。これらは、高密度で低消費電力なクラスタを構築するときに重要なことである。今回の結果から高密度に実装することによって、現在の主流なプロセッサよりもスペースあたりの性能や電力あたりの性能で上回る可能性があることを示した¹¹⁾。

7.3 ネットワークインタフェースの消費電力

通信に関しては、ネットワークカードの消費電力が CPU の消費電力に比べても小さいので全体の消費電力に影響を与えることはほとんどない。また、スイッチに関しても通常的に電力を消費するので通信を行うときに特に変化は見られない。このクラスタの NIC はオンボードのものであり、ネットワークの消費電力は電力量に含まれている。今回の評価にはスイッチの消費電力は含まれていない。その消費電力を無視するわけにはいかないので既成品のデータをもとに議論を行う。図 17 に既存のスイッチ（レイヤ 2、スイッチングハブに限定）の消費電力をポート数ごとに比較したものを示す。スイッチの消費電力は製品によって様々であり、性能と消費電力のトレードオフとなるが我々の目的とするクラスタには高消費電力なスイッチを用いることは適さず、ポートあたり 1W のスイッチを選択することによって低消費電力を維持できる。しかしながら、100BASE に関してはほぼポート数に比例した形で消費電力が増しており、おおむねポートあたり 1W で推移している。我々の目的とするのは高密度・低消費電力なクラスタであり消費電力とネットワークとのトレードオフの関係から高性能なスイッチを選択することは難しくなる。スイッチの消費電力に関してはポートあたり 1W を選択することによってスイッチの消費電力はノードあたり 1W となり全体の約 1 割

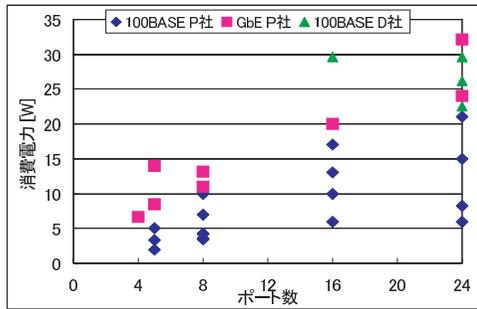


図 17 既存のスイッチの消費電力の比較

Fig. 17 Power consumption of existing switch.

となるため、十分に低消費電力を維持できる。

8. おわりに

本研究では、Pentium4 や Pentium-M, XScale, Crusoe の消費電力をダイナミックに測定することにより各 CPU の電力特性を評価した。各 CPU の電力特性として、Pentium4 は非常に性能が高いが、瞬間の消費電力も非常に高くなるため Pentium4 には大きな冷却装置が必要になる。低消費電力 CPU は瞬間の消費電力は低いために冷却装置が必要ない。低消費電力 CPU では、ソフトウェアからのアプローチとして、キャッシュブロッキングを行い電力量を大幅に削減できる。

低消費電力なプロセッサをクラスタに用いることによって、Pentium4 よりも性能が高く、電力量も低くすることが実現できる可能性があることを示した。ソフトウェアの最適化はクラスタにおいても重要であり、最適化を行うことによって電力量を最大で 10 分の 1 にできることを示した。

また、動作周波数周波数をコントロールすることによって、性能を大幅に下げることなく省電力化の可能性を示した。しかし、ほとんどのアプリケーションでは動的に周波数を変化させる DVS は、おおむね、電力・性能面で有効に動作していることも分かった。今後の課題として、MMX や SSE などの命令を使用する場合の消費電力についても測定を行い、特性を調べることによって、より低消費電力でかつ高性能を維持することを検討する¹³⁾。清水ら¹²⁾が述べているようにメディア専用命令などを用いることにより低消費電力化と電力性能の向上が期待できる。さらに、DVS の利用を含めた低消費電力化について検討する必要があると考えている。

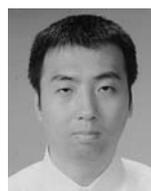
謝辞 様々なご助言をいただいた CREST チームの方々に感謝します。本研究は、科学技術振興機構・戦略的創造研究「低消費電力化とモデリング技術によるメガスケールコンピューティング」による。

参考文献

- 1) Segars, S.: Low Power Design Techniques for Microprocessors, *ISSC* (Feb. 2001).
- 2) Warren, M., Weigle, E. and Feng, W.: High-Density Computing: A 240-Node Beowulf in One Cubic Meter, *SC2002* (Nov. 2002).
- 3) IBM and Lawrence Livermore National Laboratory: An Overview of the BlueGene/L Supercomputer, *SC2002* (Nov. 2002).
- 4) Intel: *Intel XScale™ Microarchitecture Technical Summary*, Intel Corporation (2000).
- 5) *Crusoe Processor Model TM5800 Product Brief*, Transmeta Corp (Feb. 2003).
- 6) Powelse, J., Langendoen, K. and Sips, H.: Dynamic Voltage Scaling on a Low-Power Microprocessor, UbiCom-TechnicalReport (2000).
- 7) da Silva, J.A.F.: A new Processor for Power Efficient Computing, *ICCA04* (Jan. 2004).
- 8) 寒川 光: RISC 超高速化プログラミング技法, pp.113-148, 共立出版 (1995).
- 9) Brooks, D.M., Bose, P., Schuster, S.E., et al.: POWER-AWARE MICROARCHITECTURE: Design and Modeling Challenges for Next-Generation Microprocessor, *IEEE MICRO*, pp.26-44 (Nov. 2000).
- 10) Gonzalez, R. and Horowitz, M.: Energy Dissipation In General Purpose Microprocessors, *IEEE Journal of Solid-State Circuits*, Vol.31 (Sep. 1996).
- 11) Hotta, Y., Sato, M., Boku, T., Nakashima, H., Nakamura, H., Matsuoka, S., et al.: MegaProto: A Prototype of Ultra Low-Power Mega-Scale System, *CoolChips VII* (Apr. 2004).
- 12) 清水雄歩, 津邑公暁, 中島康彦, 五島正裕, 森真一郎, 北村俊明, 富田真治: 距離画像生成処理におけるメディアプロセッサの評価, 情報処理学会論文誌: コンピューティングシステム (ACS 3), pp.257-267 (Aug. 2003).
- 13) Intel: *Notebook Computing: Power and Performance Analysis*, Intel Corporation (June 1997).

(平成 16 年 1 月 31 日受付)

(平成 16 年 5 月 9 日採録)



堀田 義彦

昭和 54 年生。平成 15 年筑波大学第三学群情報学類卒業。現在、同大学大学院システム情報工学研究科在学中。低消費電力クラスタ等の研究に従事。



佐藤 三久 (正会員)

昭和 34 年生。昭和 57 年東京大学理学部情報科学科卒業。昭和 61 年同大学大学院理学系研究科博士課程中退。同年新技術事業団後藤磁束量子情報プロジェクトに参加。平成 3

年通産省電子技術総合研究所入所。平成 8 年新情報処理開発機構並列分散システムパフォーマンス研究室室長。平成 13 年より、筑波大学電子・情報工学系教授。同大学計算物理学研究センター勤務。理学博士。並列処理アーキテクチャ、言語およびコンパイラ、計算機性能評価技術、グリッドコンピューティング等の研究に従事。日本応用数理学会、IEEE 各会員。



朴 泰祐 (正会員)

昭和 59 年慶應義塾大学工学部電気工学科卒業。平成 2 年同大学大学院理工学研究科電気工学専攻後期博士課程修了。工学博士。昭和 63 年慶應義塾大学理工学部物理学科助手。

平成 4 年筑波大学電子・情報工学系講師、平成 7 年同助教授、平成 16 年同大学システム情報工学研究科および計算科学研究センター助教授、現在に至る。超並列処理ネットワーク、超並列計算機アーキテクチャ、クラスタコンピューティング、並列処理システム性能評価等の、ハイパフォーマンスコンピューティングシステムの研究に従事。日本応用数理学会、IEEE 各会員。



高橋 大介 (正会員)

昭和 45 年生。平成 3 年呉工業高等専門学校電気工学科卒業。平成 5 年豊橋技術科学大学工学部情報工学課程卒業。平成 7 年同大学大学院工学研究科情報工学専攻修士課程修了。

平成 9 年東京大学大学院理学系研究科情報科学専攻博士課程中退。同年同大学大型計算機センター助手。平成 11 年同大学情報基盤センター助手。平成 12 年埼玉大学大学院理工学研究科助手。平成 13 年筑波大学電子・情報工学系講師。平成 16 年筑波大学大学院システム情報工学研究科講師。博士(理学)。並列数値計算アルゴリズムに関する研究に従事。平成 10 年度情報処理学会山下記念研究賞、平成 10 年度、平成 15 年度情報処理学会論文賞各受賞。日本応用数理学会、ACM、IEEE、SIAM 各会員。



中島 佳宏 (学生会員)

昭和 55 年生。平成 15 年筑波大学第三学群情報学類卒業。現在、同大学大学院システム情報工学研究科在学中。グリッドコンピューティング等に関する研究に従事。



高橋 陸史 (学生会員)

昭和 55 年生。平成 14 年筑波大学第三学群情報学類卒業。現在、同大学大学院博士課程システム情報工学研究科在学中。ハイパフォーマンスコンピューティング向けプロセッサ

に関する研究に従事。



中村 宏 (正会員)

昭和 60 年東京大学工学部電子工学科卒業。平成 2 年同大学大学院工学系研究科電気工学専攻博士課程修了。工学博士。同年筑波大学電子・情報工学系助手。同講師、同助教授

を経て、平成 8 年より東京大学先端科学技術研究センター助教授。この間、平成 8 年~9 年カリフォルニア大学アーバイン校客員助教授。高性能・低消費電力プロセッサのアーキテクチャ、ハイパフォーマンスコンピューティング、ディペンダブルコンピューティング、デジタルシステムの設計支援の研究に従事。情報処理学会より論文賞(平成 5 年度)、山下記念研究賞(平成 6 年度)、坂井記念特別賞(平成 13 年度)各受賞。IEICE、IEEE、ACM 各会員。