

# PowerEdge11Gサーバでの HPC ワークロードに最適な BIOS 設定

デル・テクニカル・ホワイトペーパー

Jacob Liberman、Garima Kochhar 著

デル・ハイパフォーマンス・コンピューティング・  
エンジニアリング製品グループ

2009 年 7 月 13 日



本書は情報の提供のみを目的としたものであり、誤植や技術的に誤った情報が含まれる場合があります。本書の内容は明示もしくは黙示の別を問わず、一切の保証なく現状のまま提供されるものです。

© 2009 Dell Inc. All rights reserved. Dell Inc. の明確な書面による許可なく本書を複製することは、いかなる形であれ一切禁じられています。詳細は Dell までお問い合わせ下さい。

Dell、DELL のロゴ、EqualLogic、PowerEdge、PowerConnect、OpenManage、および DELL のマークは、それぞれ Dell Inc. の商標であり、Microsoft、Outlook、および Windows は、米国および各国の Microsoft Corporation の商標もしくは登録商標です。Intel、Core、Xeon、および Pentium は米国および各国の Intel Corporation の登録商標もしくは商標です。Red Hat および Enterprise Linux は、米国および各国の Red Hat, Inc の登録商標です。

## 目次

はじめに .....	5
Intel Nehalem アーキテクチャ .....	5
テスト方法 .....	7
11 G BIOS オプションの概要 .....	8
Node Interleaving .....	8
論理プロセッサ .....	9
Power Management プロファイル .....	10
C-state .....	10
Turbo Mode .....	11
結果 .....	11
消費電力 .....	11
性能 .....	13
Node Interleaving .....	13
同時マルチスレッディング .....	14
Turbo Mode と C-state .....	14
消費電力効率 .....	16
Power Management プロファイル .....	17
BIOS 設定 .....	17
高消費電力効率プロセッサ .....	18
結論 .....	20
参考文献 .....	21
付録 A- 調査結果のまとめ .....	22
Power Management プロファイル .....	22
性能 .....	22
消費電力効率 .....	22
付録 B - Linux コマンドラインから BIOS オプションを変更する DTK .....	23
付録 C - 各種ベンチマーク .....	24

## 概要

デルの第 11 世代目となるデュアルソケット PowerEdge<sup>®</sup> サーバは、インテルの最新マイクロアーキテクチャ（コード名：Nehalem）をベースにした Intel<sup>®</sup> Xeon<sup>®</sup> 5500 番台のプロセッサを搭載しています。このマイクロアーキテクチャが提供する機能は同サーバの BIOS で制御することが可能で、これらは幅広いサーバワークロードに応じて性能と消費電力効率を向上させる設計となっています。さらに、この 11G サーバは同マイクロアーキテクチャの仕様以外にも demand-based power management（DBPM）の各種プロファイルを BIOS レベルで導入しました。

本書は、BIOS 機能と DBPM プロファイルを検証してハイ・パフォーマンス・コンピューティング（HPC）環境に最適な設定を規定しています。複数のオープンソース／商用アプリケーションを HPC クラスタ上で実行し、各種 BIOS 設定や DBPM プロファイルごとに比較を行いました。これら機能の性能への影響や消費電力の計測結果を元にして、ベストプラクティスな構成を推奨しています。また、まとめとして、最後にシステム性能と消費電力効率を最大限に引き上げるためのアドバイスも提供しています。

## はじめに

デルの第 11 世代(11G)デュアルソケット PowerEdge サーバ(註 1) は、インテルの最新マイクロアーキテクチャ(コード名: Nehalem) をベースにした Intel® Xeon® 5500 番台のプロセッサを搭載しています。前世代の Intel x86 プラットフォームと異なり、Intel Xeon 5500 番台ではメモリコントローラをプロセッサ上に直接搭載しています。隣接したプロセッサに接続されたメモリには、QuickPath Interconnect (QPI) と呼ばれるプロセッサソケット間高速バスを経由してアクセスします。QPI は従来のフロントサイドバス (FSB) アーキテクチャで問題であったメモリ帯域競合を排除する技術です。Nehalem は non-uniform memory access (NUMA) アーキテクチャを採用したために、各々のプロセッサは高速にローカルメモリにアクセスできます。5500 番台プロセッサはアーキテクチャの強化以外にも、性能と消費電力効率を向上する新機能を搭載しています。本書では、これらの機能と、ハイ・パフォーマンス・コンピューティング・クラスタ (HPCC) におけるそれらの利点を解説します。

HPCC は、ハイ・パフォーマンス・コンピューティング (HPC) を実現するアプローチの 1 つです。低価格なサーバを、高速インターコネクトで接続し、スーパーコンピュータと同等の性能を実現するものです。桁外れのコストパフォーマンスを実現することから、計算機クラスタはこの 10 年で HPC アプリケーションのデファクト・スタンダードとして台頭してきました(註 2)。11G サーバは、Intel Xeon 5500 番台プロセッサ、第 2 世代 PCI Express のサポート、そして消費電力効率の高い筐体を組み合わせることで HPCC に適した基盤となっています。

本書は Intel Xeon 5500 プロセッサを紹介し、11G サーバと一緒に提供される BIOS 機能と DBPM プロファイルについて解説します。また、テスト方法も詳しく解説し、複数の典型的な HPC ワークロードにおける各種 BIOS 設定の性能への影響も明らかにします。そして最後に、性能や消費電力効率を最大限に引き出す、さらには消費電力の制約範囲内で性能を最大限に引き出すための BIOS 設定に関するアドバイスを提供します。

HPC ワークロードにはバランスの取れたアーキテクチャが必要とされるため、一つのサブシステムだけが実行時間に著しい影響を与えるということはありません。ただ、ここで紹介する指針は、一般的に I/O の負荷が高いデータベースやメールサーバといった業務アプリケーションのワークロードには適さない場合もあります。

## Intel Nehalem アーキテクチャ

インテルの「Penryn」マイクロアーキテクチャの後継が「Nehalem」マイクロアーキテクチャです。Nehalem - EP は、Xeon 5500 番台プロセッサをサポートするプロセッサソケットを 2 つ装備しています。この 5500 番台プロセッサには、以下のような 5400 番台との共通点があります。

- 共通の 45 nm 製造プロセスを採用しています。
- いずれもソケットあたり 4 コアを搭載し、スカラーおよびパック浮動小数点命令実行用にストリーミング SIMD 拡張 (SSE) のバージョン 4 をサポートします。
- 5500 番台のクロック周波数は前世代からほぼ変更がありません。本書執筆時点(2009年6月)における 11 G サーバの最高速プロセッサは 2.93 GHz となっています。第 10 世代 (10 G) Dell PowerEdge サーバ用で最も高い周波数の 5400 番台のプロセッサは 3.16 GHz でした。

5500 番台のプロセッサはあらゆる面で 5400 番台と似ていますが、基本的な違いもいくつかあります。5500 は新しいキャッシュ構造を持っており、5400 番台は 1 つのコアに共有 2 次キャッシュを最大 6MB 割り当ててことができますが、5500 番台ではコアごとに 256KB の専用 2 次キャッシュと、1 つのソケット内の全コアで共有できる 8MB の内蔵 3 次キャッシュを搭載しています。

Nehalem とこれまでのアーキテクチャの最大の違いがメモリサブシステムです。Xeon 5400 番台のプロセッサファ

ミラーは、1333 もしくは 1600 MHz のフロントサイドバス (FSB) による共有メモリコントローラへのアクセスをサポートしています。インテルは新しいアーキテクチャでレガシーな FSB アーキテクチャを断念し、DDR - 3 メモリコントローラをプロセッサに直接内蔵しました。内蔵メモリコントローラの方がローカルメモリに高速にアクセスできるほか、FSB アーキテクチャ固有の競合を排除して、共有バスでマルチコアプロセッサをサポートしています。図 1 は Intel Nehalem - EP プロセッサのブロック図です。

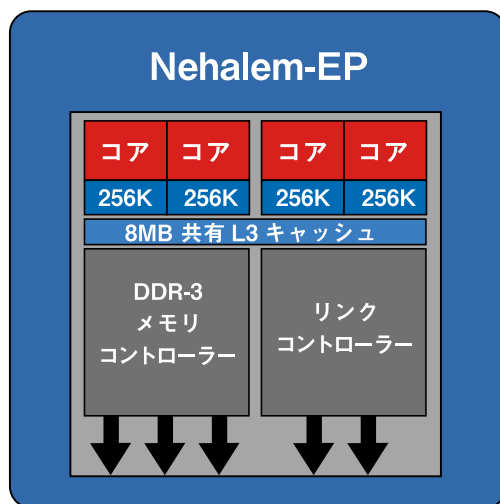


図1: Xeon 5500プロセッサブロック図

それぞれの Nehalem - EP メモリコントローラは、3つの DDR - 3 チャンネルを用意しています。Dell 11G デュアルソケット PowerEdge サーバは、モデルによりチャンネルあたり最大 2 ～ 3 枚の DIMM (DPC) をサポートしています。PowerEdge R710 と M710 の両サーバは最大で 3DPC をサポートするため、プロセッサソケットあたりでは DIMM が 9 枚、あるいはサーバあたりでは DIMM が 18 枚となります。R610 および M610 は最大 2DPC をサポートし、CPU ソケットあたりでは DIMM が 6 枚、あるいはサーバあたりでは DIMM 12 枚をサポートします。

プロセッサコアは、内蔵メモリコントローラ経由で直接ローカルメモリにアクセスします。Nehalem は、リモートメモリアクセスと共有 I/O コントローラへの接続をサポートするプロセッサソケット間高速バスの QPI を搭載しています。図 2 はデュアルソケット Nehalem - EP アーキテクチャのブロック図です。

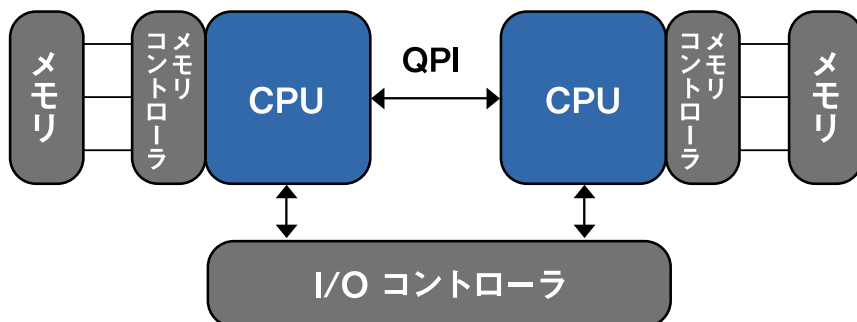


図2: Nehalem - EPアーキテクチャブロック図

内蔵メモリコントローラ経由でのローカルメモリアクセスは、Nehalem アーキテクチャの QPI リンクを使ったりリモートメモリアクセスより高速です。QPI リンクの速度は表 3 にあるようにプロセッサの周波数ビンに応じて変化します。

## テスト方法

前項で解説したアーキテクチャの強化以外に、Nehalem では消費電力効率と性能向上も目指した BIOS 機能を新たに搭載しており、11G サーバには新たな電力制御機能も搭載されています。本書では、HPC ワークロードにおける性能と消費電力効率を最大限に高めるための指針を導き出すために、これらの機能の影響を数値化します。

単一サーバやクラスタは、一連の典型的な HPC アプリケーションやマイクロベンチマークを使ってベンチマークを計測します。マイクロベンチマークは独立サブシステムの性能を計測するもので、特定のサブシステムが搭載した機能による最大の影響を特定するのに便利な理想的ワークロードとなっています。現実的な BIOS 設定やメモリプロファイルの影響評価にはクラスタレベルのアプリケーションが使用されました。また、この調査にはオープンソースと商用アプリケーションの両方が選択されています。これらのベンチマークやアプリケーションは表 1 に一覧されており、ベンチマークの詳細は付録 C の「各種ベンチマーク」にあります。

ベンチマーク	内容	タイプ
STREAM	スレッドメモリ帯域幅テスト	メモリマイクロベンチマーク
lat_mem_rd	メモリレイテンシテスト、アイドルアレイの追跡	LMBench のメモリマイクロベンチマーク
DGEMM	スレッドマトリックスの乗算ルーチン	CPU マイクロベンチマーク
HPL	分散浮動小数点ベンチマーク	CPU と通信のベンチマーク
Fluent	計算流体力学	商用クラスタ対応アプリケーション
Ansys	構造力学	商用クラスタ対応アプリケーション
ECLIPSE	油層シミュレーション	商用クラスタ対応アプリケーション
WRF	気候モデリング	オープンソースクラスタ対応アプリケーション
LU	LU 分解、物理システム	オープンソースクラスタ対応人工カーネル

表1: 使用したベンチマークとアプリケーション

Nehalem の BIOS 機能は、性能向上だけでなく、消費電力の削減も目指しています。このような理由から、ベンチマークの性能結果は消費電力の計測で補足しており、HPC アプリケーションの性能向上には消費電力の増加が伴うケースが多くなります。

そこで、さまざまな BIOS 設定に関連した消費電力とパフォーマンスのトレードオフを数値化するために、アプリケーションのパフォーマンスデータと一緒に消費電力のデータも使用します。消費電力効率は「パフォーマンス」／「消費電力」もしくはワットあたりの性能で計算しています。「レート」(1 日の最大アプリケーション実行回数)は、ベンチマーク全体で共通の性能測定単位になっています。アプリケーションのレートは、1 日の秒数をアプリケーション実行時間(秒)で割った数です。すべての結果は Dell の HPCC エンジニアリングラボで生成された性能データから導き出されています。テストクラスタの構成については表 2 に示されており、各ベンチマークの具体的な構成の詳細は必要に応じて説明されています。

コンポーネント	説明
サーバ	Dell PowerEdge R610、Dell PowerEdge M610 (16 台)、PowerEdge M1000e シャシー採用
サーバBIOS	1.1.4
プロセッサ	Intel Xeon X5550、Intel Xeon X5570、Intel Xeon E5540
メモリ	4GB の 1333 MHz RDIMM X 6 枚、4GB の 1066 MHz RDIMM X 6 枚
ストレージ	Dell SAS 6iR コントローラ、73GB の 1 万回転 SAS ハードディスク X 2、M610 で RAID 1 Dell Perc6i コントローラ、73GB の 1 万 5000 回転 SAS ハードディスク X 2、R610 で RAID 0
インターコネクト	InfiniBand - Mellanox MTH MT26428 [ConnectX IB QDR、第 2 世代 Gen2 PCI Express]
IBスイッチ	Mellanox 3601Q QDR ブレードシャシー内蔵 I/O スイッチモジュール
GbEネットワーク	Broadcom BCM5709
GbEスイッチ	PowerConnectTM M6220 シャシー内蔵 I/O スイッチモジュール、PowerConnect 6248 ラックスイッチ
ソフトウェア	ClusterCorp Rocks+ 5.1 for Dell (註 3)
OS	Red Hat Enterprise Linux 5.3 x86_64 (2.6.18-128.el5 kernel)
IBスタック	Mellanox OFED 1.4

表2: テストクラスタの構成

註3 本製品にはカリフォルニア大学サンディエゴ校San Diego Supercomputer CenterのRocks Cluster Groupとその貢献者らの開発によるソフトウェアが含まれています。

## 11G BIOSオプションの概要

本項は、今回の調査で実験した 11G の各種 BIOS オプションの解説を行います。11G サーバは新たな BIOS 設定もサポートしていますが、今回の調査では Node Interleaving、論理プロセッサ、C-state、Turbo Mode、および Power Management プロファイルなど、HPC に関係する設定とプロファイルに重点を置いています。本項では、各オプションを有効にするための説明や手順に言及するほか、調査結果の理解に必要なほかの予備知識や性能測定についても述べます。

### Node Interleaving

Nehalem が採用する NUMA アーキテクチャでは、プロセッサによるローカル／リモートメモリへの非対称アクセスが可能になっています。表 3 にローカル／リモートメモリアクセスの理論帯域幅を示します (4)、(5)。5500 番台では、プロセッサのクロック周波数が内蔵メモリコントローラの最大帯域幅を決定します。

CPU 周波数 (GHz)	QPI リンク (GT/ 秒)	メモリコントローラ (GB/ 秒)	QPI リンク速度 (GB/ 秒)
<b>2.66~3.20</b>	6.40	31.99	25.60
<b>2.26~2.53</b>	5.86	25.58	23.44
<b>1.86~2.13</b>	4.80	19.20	19.20

表3: 理論メモリ帯域幅

表 3 の理論値はローカル／リモートメモリ操作の境界を示していますが、実効性能は理論上の性能で定められた最大値よりも常に低くなります。図 3 は、STREAM と lat\_mem\_rd の両マイクロベンチマークで計測したローカル／リモートメモリ操作の帯域幅とレイテンシの違いを示しています。



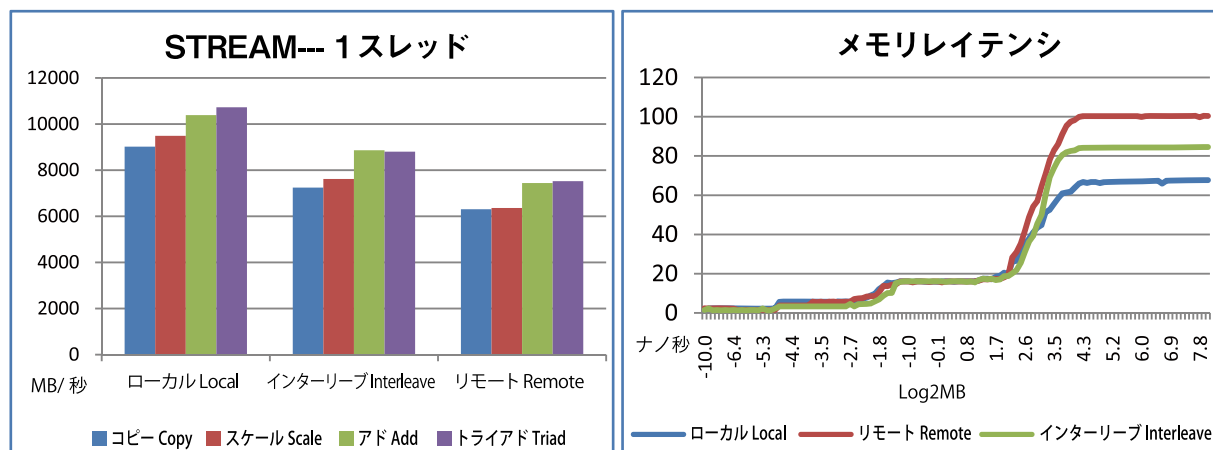


図3: PowerEdge R610 (Dual Intel Xeon X5550、4GBの1333 MHz RDIMM X 6枚を搭載) のローカル／リモートメモリアクセス時におけるメモリレイテンシと帯域幅

図3は、ローカルメモリアクセスの方がリモートメモリアクセスよりも帯域幅が約40%高く、レイテンシが短いことを示しています。リモートメモリアクセスはQPIリンクを通してメモリにアクセスするために、性能ペナルティを招きます。11Gサーバは、Node Interleavingを3つめのメモリアクセス方法として提供しており、ローカルメモリコントローラとリモートメモリコントローラ間でデータアクセスをストライピングさせることで、性能ペナルティを部分的に相殺しています。図3は、Node Interleaving利用時のメモリ帯域幅とレイテンシ性能がローカルとリモートメモリアクセス性能の中間に位置することを示しています。

サーバのBIOSでNode Interleavingを有効にするには、サーバのブート画面で「F2」を選択してから次のオプションを選びます。

Memory Settings → Node Interleaving, Values = Disabled (default), Enabled

Node InterleavingとIntel Xeon 5500メモリスバシステムの詳細な説明は、HPC用メモリの選択基準を説明したDellの技術白書を参照して下さい。

## 論理プロセッサ

論理プロセッサ機能はインテルの同時マルチスレッディング（SMT）テクノロジーをベースにしています。SMTが有効になったシステムでは、各物理コアに2つの「論理」コアがあるとみなすことで、OSからはプロセッサのコア数が実際の2倍に見えます。SMTは、各論理コアにスレッドを割り当てることで性能の向上を可能にします。論理コアは、物理コアのリソースを共有することで自身のスレッドを実行します。

サーバのBIOSで論理プロセッサ機能を有効にするには、サーバのブート画面で「F2」を選択してから次のオプションを選びます。

Processor Settings → Logical Processor, Values=Enabled (default), Disabled

## Power Management プロファイル

周波数スケーリングとは、性能向上もしくは消費電力削減のためにプロセッサのクロック周波数を変更する手法です。Nehalem アーキテクチャでは、プロセッサソケットで周波数スケーリングが可能になっています。この電力管理機能は、Demand Based Switching (DBS) と呼ばれる業界標準となっています。一般的に、DBS はプロセッサの利用率に応じてプロセッサ周波数をスケーリングする OS インターフェース経由でインプリメントされます。Linux で DBS は、cpuspeed サービスやプロセッサ専用のカーネルドライバ経由で導入できます。

11G サーバは、BIOS Power Management メニューで複数の DBS スキーマが設定されており、DBS 初期設定は「OS Control」になっています。この設定では、OS が cpuspeed サービスを使ってプロセッサ周波数のスケーリングをコントロールできるようになっています。11G サーバは、このほかに「Active Power Controller (APC)」と呼ばれる BIOS レベルの Power Management オプションも導入されています。OS Control と異なり、APC は OS とは独立して動作します。APC は、ハードウェアの消費電力／温度センサーから直接読み込んだ利用量データに応じてプロセッサ周波数スケーリングを開始することでワットあたりの性能を向上させる設計となっています。

HPC ユーザの多くにとっては、電力管理の問題よりも最高の性能を達成することの方が重要です。11G サーバには「Maximum Performance」オプションも用意されています。Maximum Performance はアイドル時でもメモリやファンなどの内部コンポーネントに最大電圧をかけ続け、高負荷と低負荷のフェーズ移行時における性能の低下を解消します。11G の各種 demand based power management 設定を表 4 にまとめます。

Power Management プロファイル	説明
Active Power Controller (アクティブパワーコントローラ)	Dell の設計による BIOS コントロール式のプロセッサ周波数スケーリング
OS Control	OS コントロール式のプロセッサ周波数スケーリング
Maximum Performance (最大性能)	すべてのプロセッサに最高電圧をかけて OS サービスをオーバーライド
Custom (カスタム)	ファン、メモリ、およびプロセッサ電圧のユーザ定義設定

表4: 11GのDemand Based Power Management設定

サーバ BIOS の Power Management メニューにある管理プロファイルを選択するには、サーバのブート画面で「F2」を選択してから次のオプションを選びます。

Power Management → Values = OS Control (default), Active Power Controller, Custom, Static Max Performance

本書の「結果」の項には、Power Management プロファイルごとの消費電力削減についての解説があります。

## C-state

C-state は 5500 プロセッサに導入されている消費電力削減機能であり、ソケットレベルで機能する DBS とは異なり、個々の CPU コアで周波数スケーリングを実現します。コアあたりの電圧が低下すると、それにしたがってコア周波数が低下しますが、プロセッサソケット内のアクティブコアはすべて同じ周波数で動作します。

コアがアクティブかどうかという情報を提供する C-state もあります。表 5 で C-state をいくつか紹介します。

C-state	CPU 状態	説明
C0	アクティブ	アクティブ状態 - 命令がコアによって実行されていない。
C1	アクティブ	アクティブ状態 - コアはアクティブだが命令は実行されていない。
C3	非アクティブ	アイドル状態 - コアは非アクティブで、コアキャッシュはフラッシュされる。
C6	非アクティブ	パワーゲートが消費電力を 0 近くに削減、キャッシュはフラッシュされる。

表5: C-stateの例

5500 番台のプロセッサは多くの C-state をサポートしており、その利用モデルはハードウェアベンダーごとに異なります。

## Turbo Mode

Turbo Mode は「Turbo Boost」(註 6)とも呼ばれ、一定の条件下でプロセッサコアが基本動作周波数より高速に動作できるようにします。プロセッサが定格電力より低かったり制限温度以下で動作している場合は、Turbo Mode が CPU のクロックレートを引き上げて性能を向上させることができます。

プロセッサがサポートしているときにサーバの BIOS で Turbo Mode を有効にするには、サーバのブート画面で「F2」を選択してから次のオプションを選びます。

Processor Settings → Turbo Mode, Values=Enabled (default), Disabled

プロセッサが到達できる最高周波数はシステムのアクティブコアの数に依存し、プロセッサモデル番号によっても変わります。(註 7) たとえば、Intel Xeon X5550 プロセッサの基本周波数は 2.66 GHz です。このプロセッサの「Turbo」ブーストレートは 2/2/3/3 です。このレートは、コアが 4 基、3 基、2 基、もしくは 1 基アクティブな場合にプロセッサが引き上げ可能な周波数ステップ数（133 MHz 単位）を表しています。たとえば、Xeon 5500 は 4 コアすべてがアクティブな場合に周波数を 2 ステップ上げて、周波数を 2.66 から 2.93 GHz へと引き上げることができます。プロセッサ内のアクティブなコアはすべて同じ周波数で動作するようになります。

## 結果

本項は、複数クラスタや単独サーバ上で実施された消費電力、性能、消費電力効率の各テスト結果について解説します。まず、Power Management プロファイル全体のアイドル時の消費電力を比較し、次にさまざまな BIOS 設定の性能への影響と、性能を最大限に引き上げるベストプラクティスについて解説します。最後は、性能向上と消費電力削減の関係を数値化する消費電力効率調査で結果の項を締めくくっています。

## 消費電力

クラスタが増え続けるなか、消費電力はクラスタ設計の主要検討事項の 1 つになっています。電源の供給や冷却に関連したインフラコストは、クラスタの良好な経済効果をたちどころに損なってしまいます。さらに、多くのクラスタは常にフル活用されているわけではなく、(ユーザがスケジューラを使わずにジョブをサブミットする) インタラクティブクラスタにはアイドル時間があります。これらのクラスタでは、アイドル時に消費電力を最小限に抑えることが重要な検討事項となっています。

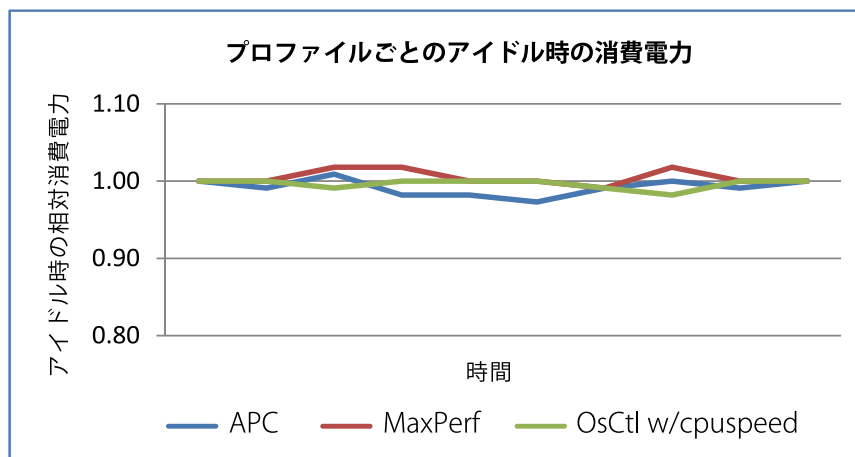


図4: 16ノードPowerEdge M610クラスタ (それぞれにDual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) におけるPower Managementプロファイルごとの相対アイドル時消費電力

図4は、11G サーバがサポートするすべての Power Management プロファイルを比較しています。Turbo Mode、C-state、Node Interleaving、そして SMT BIOS オプションは無効になっています。このグラフは、システムのアイドル時には3つのプロファイルすべての消費電力がほぼ同じであることを示しています。アイドル状態のクラスタにはどの Power Management プロファイルでも適しています。したがって、消費電力プロファイルの選択は、負荷のかかった状態における消費電力で行う必要があります。これらのデータは次項で解説します。

図5は、C-state を有効にするとシステムのアイドル時消費電力がさらに大幅に削減されることを示しています。このグラフは、16 ノードクラスタのアイドル時消費電力を各消費電力プロファイルごとに比較しています。棒の高さは C-state を有効にしたときと、C-state を無効にしたときの削減消費電力量を消費電力プロファイルごとに示しています。本テスト中は Turbo Mode、Node Interleaving、論理プロセッサを無効にしています。

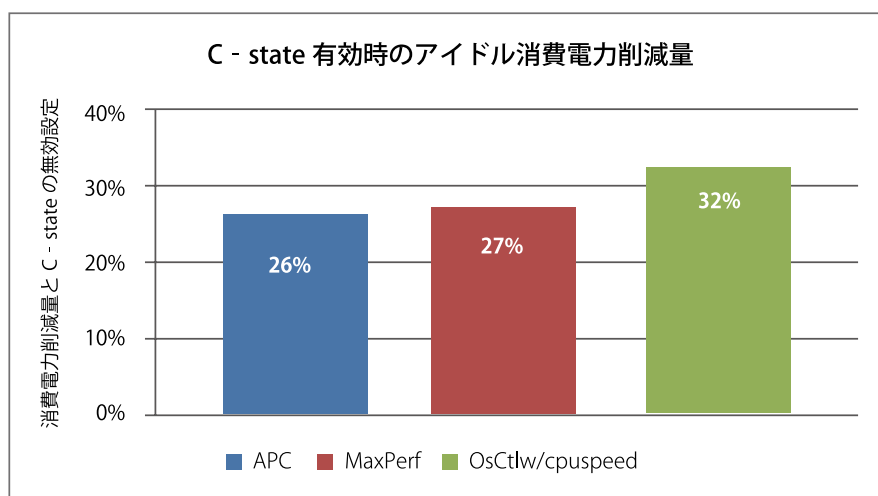


図5: 16ノードPowerEdge M610クラスタ (それぞれにDual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) におけるC-state有効時のアイドル時消費電力削減量

この調査では、C-state を有効にするとアイドル時にクラスタ全体で消費電力が最大 32%削減されるという結果が分かります。したがって、アイドル時もしくはインタラクティブクラスタの消費電力を最小限に抑えるには C-state を有効にする必要があります。

## 性能

前項では、HPC クラスタでアイドル時消費電力を最小限に抑えるためのベストプラクティスを解説しています。本項では、性能を最大限に高める BIOS 設定を解説します。そして、最終結果を示す「消費電力効率」の項では、ワットあたり性能を向上させるためのベストプラクティスを紹介します。

### Node Interleaving

この調査では、Node Interleaving が HPC アプリケーションの性能に与える影響を理解するため、9 種のベンチマークを使用して、Node Interleaving を有効にした場合と無効にした場合の両方でシングルノードのアプリケーション性能を測定しました。今回の調査結果は図 6 に示されるとおりです。1.0 より長い棒は、Node Interleaving が有効のときに無効のときより性能が向上したことを示しています。

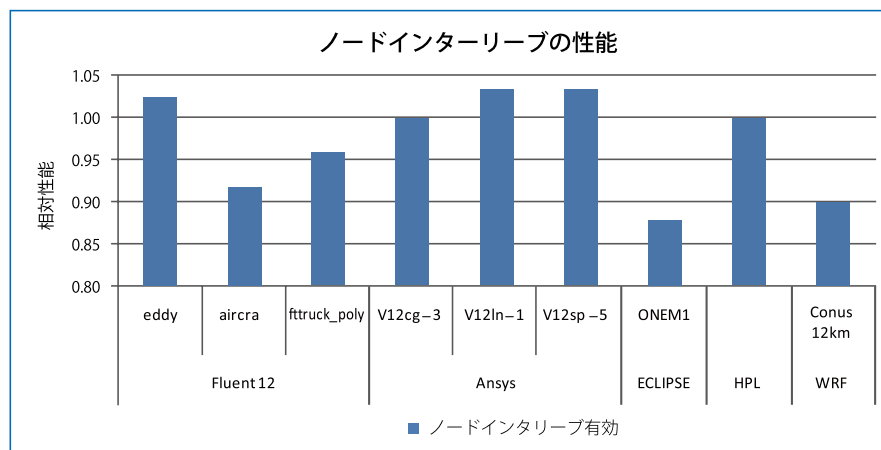


図6: PowerEdge R610 (Dual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) におけるノードインターリーブの相対性能

図 6 を見ると、Node Interleaving は 9 種のうち 3 種のベンチマークで 2 ~ 4% の性能向上に役立ち、9 種のうち 4 種のベンチマークで 4 ~ 13% 性能を低下させています。まとめると、ノードインターリーブが役立つことはめったになく、役に立ってもほんのわずかであるため、典型的な HPC ワークロードにおいてはこれを無効にしておくべきだということになります。

ただ、Node Interleaving を活用した方が良いケースが 3 つあります。1 つ目のケースは、実行中のプロセスのメモリ要求がシングルソケットの使用可能なメモリ容量を越えたときです。このような状況においては、リモートメモリアクセスの性能の低下を最小限に抑えたり、メモリアドレス空間がフルになるとプロセスが動かないときに備えて Node interleaving を有効にする必要があります。

2 つ目のケースは、スレッドの配置が悪い場合マルチスレッドアプリケーションのメモリ局所性が悪いときです。ソケット内で利用できるコアの数より多くのスレッドをプロセスが作り出すことがあります。リモートソケットに配置されたスレッドは、ほかのメモリコントローラに接続されたメモリへのアクセスが必要になる場合があります。この場合、Node Interleaving はリモートメモリアクセスに関連した性能低下を抑えることができます。この問題への対応としては、

1つのプロセスが作り出すスレッドの数をローカルプロセッサコアの数に制限する戦略の方が優れています。そして、Node Interleaving を有効にする必要がある3つ目のケースは、NUMA 非対応のレガシー OS を運用しているときです。

### 同時マルチスレッディング (SMT)

インテルは、Pentium® 4 プロセッサでハイパースレッド技術を導入してきました。登場時の同 OS スケジューラは論理プロセッサにプロセスを最適にアロケートしておらず、性能の低下を頻繁に招きました。このような理由から、ハイパースレッド技術は HPC では無効にされることが多くありました。(註 8) Xeon 5500 と Red Hat® Enterprise Linux® (RHEL) 5.3 の世代になると、SMT がこれまでのバージョンより効率的に動作するようになり、多くのケースで性能を向上させました。図 7 は、7 種の HPC アプリケーションベンチマークにおける SMT の性能への影響を示しています。このテストでは、BIOS で SMT を有効にし、アプリケーションはシステム内の論理コア数と同じプロセス数で実行されました。そして、SMT 有効時の性能と SMT 無効時のそれとを比較しました。1.0 以上の数字は SMT の性能向上値を示しています。

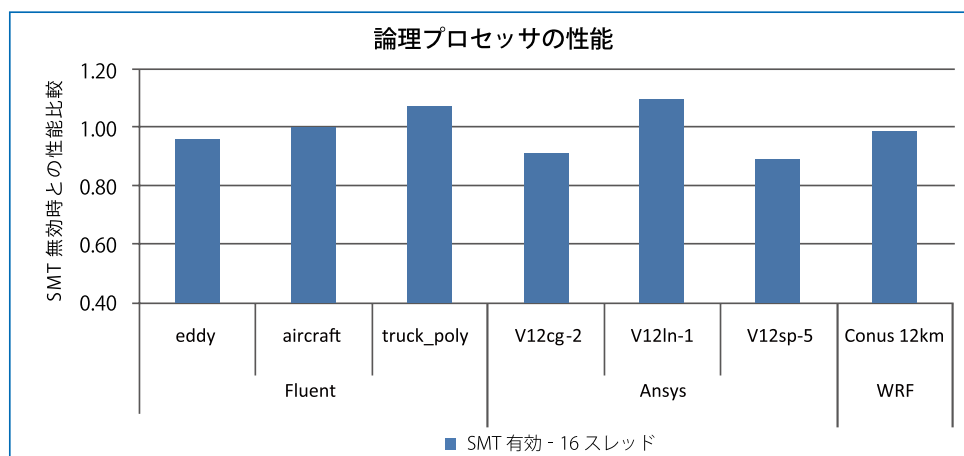


図7: PowerEdge R610 (Dual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) で論理プロセッサを有効にしたときと無効にしたときの論理プロセッサの相対性能

このグラフは、SMT 有効時の結果に良し悪しがあることを示しています。SMT は2つのテストケース (Fluent truck\_poly と Ansys V12ln-1) で性能を7~10%向上させました。一方、性能を11%落とした Ansys V12sp-5 など、SMT が性能を低下させるケースもいくつかありました。ほかのケースについては結論が出ませんでした。また、SMT は5つのワークロードにおいて性能向上につながりませんでした。したがって、HPC ワークロードで性能を最大限に引き上げるには、SMT の性能への影響をアプリケーション単位で必要に応じて有効もしくは無効にしてテストすることが推奨されます。

### Turbo Mode と C-state

Turbo Mode では、プロセッサコアがその基本動作周波数より高速に動作するようになります。本項では、Turbo Mode を有効にする場合の性能上の利点について解説します。Turbo Mode で C-state を有効にしておく必要はありませんが、Turbo Mode はその時々システム環境に応じて動作する機能であり、C-state が有効になっているとシステムが Turbo Mode で動作する可能性が高まります。これは、C-state が有効だと処理時のシステム熱出力カーブが拡大するためだと思われます。図 8 は、Turbo Mode を有効および無効にして CPU 負荷の高いベンチマークを実行したときの Turbo 対応プロセッサの CPU 周波数を示しています。

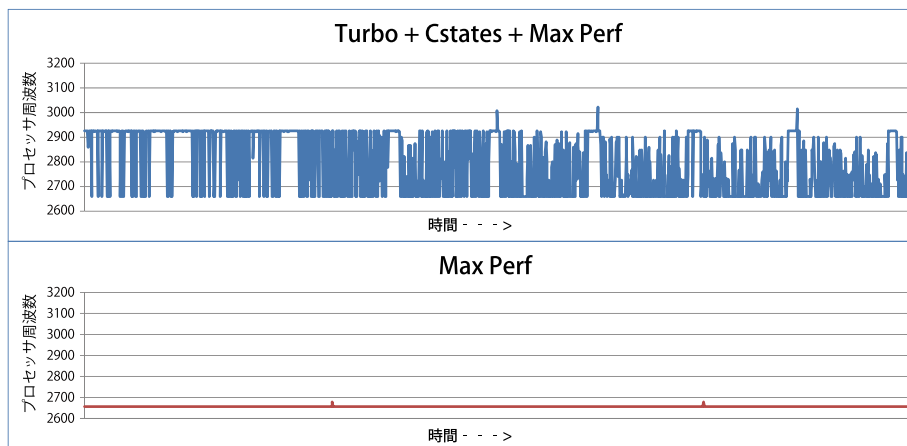


図8: PowerEdge R610 (Dual Intel Xeon X5550と4GBの1333 MHz RDIMMを6枚搭載) における Turbo Mode有効時のCPU周波数

Turbo Mode で C-state を両方とも有効にした場合、コアの周波数は 2.66 GHz と 2.93 GHz の間で変化します。Turbo Mode にしない場合はプロセッサ周波数が変化しません。図を見ると、Turbo Mode を有効にしてこのベンチマークテストを行ったときの性能向上は、Turbo が実現する周波数向上が実際の要因となっていることは明らかです。したがって、Turbo Mode を活用するためには、BIOS で Turbo Mode を有効にしてください。

Turbo Mode に入る確率を最大限に高めるには、C-state も有効にする必要があります。Power Management プロファイルは利用可能なオプションにも設定することができます。「Power Management」の「OS Control」プロファイルでは、C-state と Turbo Mode を活用するには Linux の cpuspeed サービスを有効にする必要があります。Turbo Mode が動作するにはアイドル状態のコアから生じる消費電力の余裕が必要だというのは一般的な誤解です。Turbo Mode はすべてのコアがフルに利用されていても性能は向上できると覚えておくことが重要です。

図 9 は、16 ノードクラスタで、2つのインターコネクトごとに、4 種の HPC ベンチマークを各々実行した場合の性能比較を示しています。このグラフは、インターコネクトごとに Turbo Mode を有効にしたときと無効にしたときの両方の結果を示しています。1.0 以上の数字は、Turbo Mode が特定のインターコネクトに対応する分、性能を向上させたことを示しています。これらのテストでは、C-state は有効、消費電力プロファイルは OS Control に設定、Node Interleaving と SMT は無効になっています。

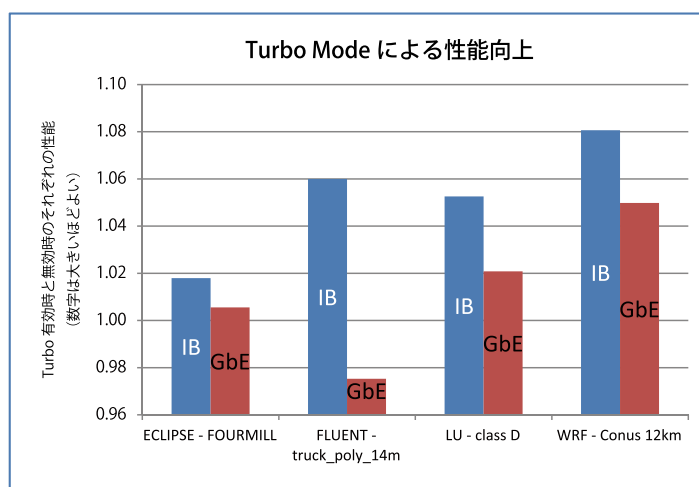


図9: 16ノードのPowerEdge M610クラスタ (それぞれDual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) においてTurbo Modeを有効にしたときのクラスタワークロードとインターコネクトの性能向上



図 9 は、InfiniBand インターコネクトにおいて Turbo Mode と C-state を有効にすると性能が 2 ～ 8% 向上することを示しています。ギガビットイーサネットでは、Turbo Mode が ECLIPSE、LU、および WRF を 1 ～ 5% 向上させていますが、Fluent には変化がありません。ギガビットイーサネットにおいて Turbo が Fluent の性能を 2% 低下させているのは、このベンチマークの誤差の範囲内であり、ギガビットイーサネットを用いた場合、Turbo Mode は Fluent には有効でないという結論になります。

ギガビットイーサネットにおいて Turbo が実現する向上分は、InfiniBand で見られるものよりも小さくなっています。ギガビットイーサネットの通信に関するオーバーヘッドの方が Turbo Mode で得られる性能向上分より重要である可能性が高いようです。Turbo は、高速インターコネクト搭載クラスタの方により大きいメリットがあります。HPC ワークロードで性能を最大限に引き出すには、Turbo Mode と C-state を有効にする必要があります。

C-state 機能は主に消費電力の削減を目指したのですが、図 10 は C-state と性能の関係を示しています。C-state は Turbo Mode と併用すると性能の向上に役立ちますが、C-state を有効にすると性能が低下する状況もあります。C-state 有効時は、ステートの移行を開始するために CPU の消費電力率と放熱が継続的にポーリングされます。このポーリングと、それに関連する C-state の移行は、メモリとインターコネクトのレイテンシを増やし、待ち時間レイテンシの制約を受けるアプリケーションの性能を低下させる場合があります。図 10 は、OSU MPI レイテンシ マイクロベンチマーク実行時の C-state 移行に関連した MPI レイテンシの増加を示しています。

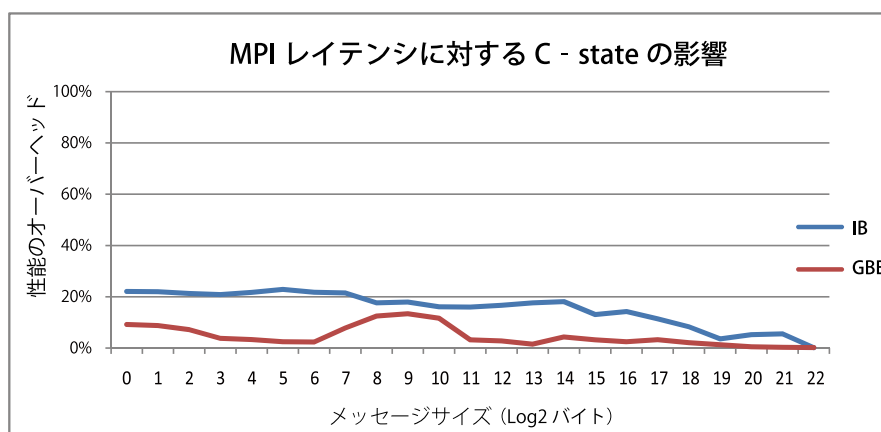


図10: 2台のPowerEdge R610 (Dual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載) におけるC-state移行時の待ち時間オーバーヘッド

図 10 は、C-state を有効にすると MPI レイテンシが InfiniBand で最大 20%、ギガビットイーサネットでは 12% 延びることを示しています。このような理由から、金融系で利用されるリアルタイムトレーディングアプリケーションなど、レイテンシに依存する HPC ワークロードでは、C-state を有効にすることは適切ではないかもしれません。

## 消費電力効率

これまでの項では、システムのアイドル時に消費電力を削減したり、アプリケーションの性能を最大限に引き上げたりするための BIOS 設定や Power Management プロファイルに関するアドバイスを示してきました。本項では、消費電力効率を最大限に引き上げるためのベストプラクティスについて解説します。

消費電力効率を最適化するということは、性能と消費電力の最適なバランスを見つけ出すことを意味します。一般的に、消費電力効率は性能を消費電力で割ることにより算出します。今回の調査目的では、1 日あたりアプリケーション実行可能回数を示す「レート」にアプリケーション性能の結果をすべて換算します。実行時間が短ければレートは高くなります。



## Power Management プロファイル

本書の消費電力の項では、Power Management プロファイル全体で C-state を有効にするとアイドル時の消費電力が低下することを紹介しました。この調査では、4 つの HPC ワークロードを実行してフル活用されている 16 ノードクラスタ全体の Power Management プロファイルの消費電力効率を比較しています。図 11 は、「Max Performance」プロファイルと比較した各 Power Management プロファイルの消費電力効率をグラフ化しています。1.0 以上の数字は、当該プロファイルの消費電力効率が「Max Performance」プロファイルのそれを上回る分の値を意味します。これらのテストでは、Turbo と C-state が有効にされ、Node Interleaving と SMT は無効にされています。

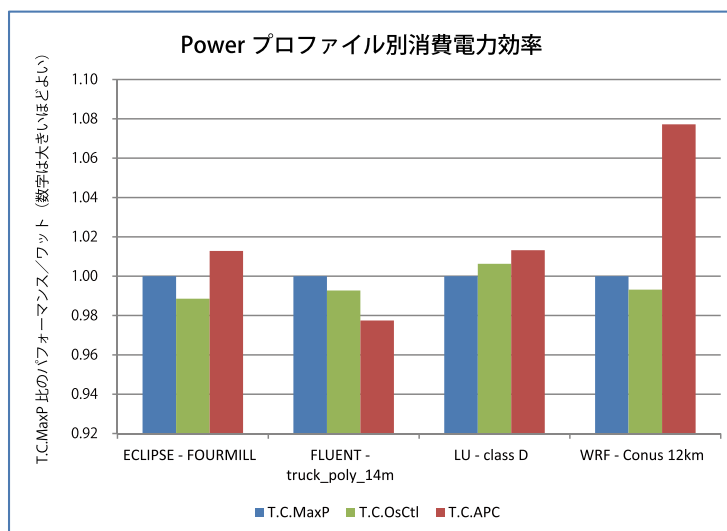


図11: 16台のPowerEdge M610 (それぞれDual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚搭載)とMellanox QDR HCAにおけるPower Managementプロファイルごとのパフォーマンス/ワット

図 11 は、フル利用されているクラスタの消費電力効率には Power Management プロファイルの影響がほとんどないことを示しています。すべてのプロファイルにおいて、ECLIPSE、Fluent、そして LU 実行時の消費電力効率はお互いの値の 2 ～ 3% 以内に収まっています。

注目に値する例外は WRF です。WRF では、「Max Performance」と「OS Control」の両プロファイルが同様の消費電力効率を出す一方で、APC が最高のパフォーマンス/ワットを出して、ほかのプロファイルを 8% 近く上回っています。したがって、APC は WRF のように通信の多いアプリケーションの実行にフル活用されているクラスタに推奨される消費電力プロファイルです。消費電力削減量は、クラスタの数が増えれば増えるほど増加することが期待されます。

## BIOS 設定

HPC ワークロードでは、併用することで最高の性能が出る Turbo と C-state は常時有効にしておくべきだと思います。しかし、Turbo Mode と C-state を無効にすべき状況が 2 つあります。

第 1 に、C-state はネットワークやメモリのレイテンシを最小限に抑える必要があるアプリケーションの性能向上には役立ちません。図 10 を参照して下さい。このカテゴリーには、リアルタイム信号処理や高頻度トレードアプリケーションなどがあります。これらのアプリケーションでは、C-state を無効にすることで最高の性能を達成することができます。

第 2 に、Turbo Mode による性能向上には消費電力の増加という犠牲が伴う点に注意することが重要です。性能と消

費電力効率はトレードオフの関係にあり、これは図 12 に示されています。このグラフには、4 ノードクラスタ対応ベンチマークにおけるワットあたりの性能データが示されています。それぞれのベンチマークで Turbo Mode 無効時の結果を Turbo Mode を有効にしたときの結果に対して正規化しています。

1.0 以上の値は、Turbo Mode を無効にすることでクラスタの消費電力効率が上がったことを意味します。消費電力効率は性能と消費電力の 2 つのコンポーネントで構成されます。消費電力効率の向上は、消費電力が低下するか、性能が向上するか、あるいはその両方によって実現します。図 12 にある棒グラフ中の記述は、Turbo 無効時の性能低下と消費電力削減量を示しています。たとえば、ECLIPSE で Turbo を無効にすると、Turbo 有効時と比較して性能が 2% 低下しますが消費電力は 21% 削減されます。つまり、Turbo 無効時は Turbo 有効時と比較して消費電力効率が 18% 向上することになります。この実験では、C-state が有効にされ、消費電力プロファイルは OS Control に設定、Node Interleaving と SMT は無効にされ、InfiniBand インターコネクトが使われました。

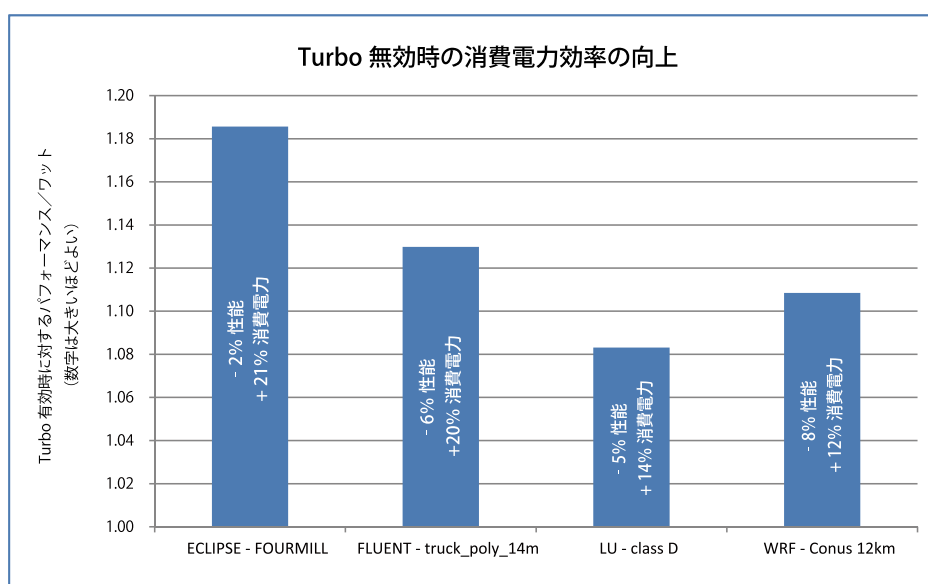
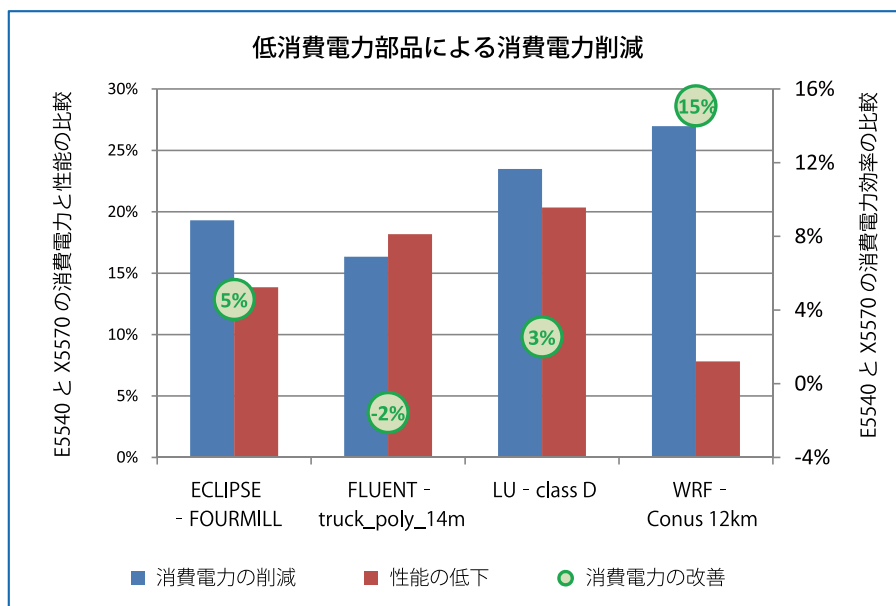


図12: 16台のPowerEdge M610 (それぞれDual Intel Xeon X5570と4GBの1333 MHz RDIMMを6枚とMellanox QDR HCAを搭載) で4ノードクラスタ対応アプリケーションを実行したときの消費電力効率比較

図 9 は、Turbo Mode を有効にすると、4 ノードクラスタ対応アプリケーションを InfiniBand インターコネクト上で実行したときに全体の性能が 2～8% 向上することを示しています。しかし、図 12 を見るとこれらの性能向上の結果、同じアプリケーションで消費電力が Turbo Mode 無効時よりも 12～21% 増加したことも分かります。したがって、Turbo Mode 有効時の運用は無効時よりも消費電力効率が低いことを意味します。バランスの取れたパフォーマンス/ワットが問題になる HPC 環境において、Turbo Mode の利用は最高の性能が絶対に重要であるとき、もしくは性能ベンチマークを計測するときに限定して使用すべきです。

#### 高消費電力効率プロセッサ

Power Management の項で解説した結果によると、BIOS の設定では、最高の性能が必須でフル活用されたクラスタの消費電力効率が向上する可能性はほとんどないことになります。Fluent や LU のように CPU に対する負荷の高いアプリケーションでは特にそうです。これらのクラスタで消費電力効率を向上させる最良の手段は、低消費電力のコンポーネントと最速のインターコネクトを搭載することです。このアプローチは、性能や利用率をさほど犠牲にすることなく消費電力効率の向上につながるケースが増えています。図 13 は、95 ワットの Xeon X5570 と 80 ワットの Xeon E5540 の消費電力効率を比較しています。



**図13: 16台のPowerEdge M610 (それぞれDual Intel Xeon X5570、4GBの1333 MHz RDIMMを6枚、Mellanox QDR HCAを搭載) で低消費電力CPUとメモリを採用したときの消費電力効率向上**

この実験では、4種のクラスタ対応アプリケーションを128コアのInfiniBandクラスタ上で実行しました。BIOSオプションはTurbo ModeとC-stateが有効、Max Performanceプロファイル、そしてNode InterleavingとSMTが無効です。最高速向けの構成では最速プロセッサのX5570と1333 MHz DIMMの組み合わせになっています。低消費電力構成では、消費電力効率の高いE5540を使っています。E5540メモリコントローラは1066 MHz以上の速度が出ないため、この構成では1066 MHzのDIMMを使用しています。

低消費電力の低速コンポーネントを使うことで消費電力と性能が低下する結果となりました。これらの数値をX5570クラスタと比較したものが図13に示されています。たとえば、ECLIPSEでは低消費電力構成により消費電力は19%削減されますが、性能も14%低下しました。E5540クラスタの消費電力効率はX5570クラスタ比で5%向上しています。

LUやFluentのようにCPU負荷の高いアプリケーションの消費電力効率はいずれのクラスタもかなり近づいています。E5540クラスタの性能低下分は消費電力低下分に比例しており、消費電力効率はX5570クラスタと同レベルのままになっています。

WRFでは、低電圧プロセッサが消費電力を大幅に削減しています。WRFは通信がかなり多いため、通信中のCPUの周波数低下が消費電力効率の大幅な向上につながりました。ただ、性能の低下も同様に低くなっています。その結果、最高速構成より消費電力効率が15%の向上が見られました。大規模クラスタやインターコネクト速度が遅いクラスタも、通信のオーバーヘッドが高まることから、消費電力効率の高いコンポーネントによって同様の消費電力効率向上が得られる可能性が高いと思われます。

要約すれば、低速の低電圧プロセッサと低速メモリは、適切な組み合わせを適用することにより、性能を損なうことなく大幅な消費電力削減に寄与することができます。

## 結論

Intel Nehalem - EP アーキテクチャベースの 11G サーバは、性能と消費電力効率の向上を目指して BIOS 機能と Power Management プロファイルを新たに用意しました。HPC で性能と消費電力効率を最適化するには、クラスタの特性と利用パターンを深く理解した上での慎重な機能選択が必要です。1 つのアプリケーションに特化した小規模インタラクティブクラスタと、各種アプリケーションを運用する大規模な計画運用クラスタではおそらくニーズが異なると思われます。

多くのクラスタ設計者は、性能の向上と消費電力効率の向上を相反する目標だと考えることが多々あります。しかし、常にそうとも限りません。この調査では、通信負荷の高いアプリケーションでは最適な BIOS 設定によって性能に影響を与えずに消費電力を抑えられるケースが多いことが示されています。また、クラスタは可能な限り多くの場所に低消費電力コンポーネントを採用することで設計段階でも消費電力効率を向上させることができます。

今回の調査結果に基づき、複数の利用率モデルに対応した BIOS 設定に関するアドバイスを表 6 にまとめます。

設定	消費電力効率	最大性能	バランス重視	最低レイテンシ
Power Management	APC	Max Perf	OS Control	Max Perf
C-state	有効	有効	有効	無効
Node Interleaving	無効	無効	無効	無効
SMT	無効	無効	無効	無効
Turbo Mode	無効	有効	有効	有効
cpuspeed サービス	無効	無効	有効	無効

表6: クラスタの推奨 BIOS 設定

今回のような調査の結果を元に、Dell の HPC エンジニアリングは工場出荷時に表 6 のようなバランスの取れた BIOS 設定を行うカスタム HPCC BIOS を開発しました。これらの設定は、さまざまな特性を持った幅広いクラスタで性能と消費電力効率を最大限に引き上げるはずです。

## 参考文献

**1.Data Center Workhorses: New Dell PowerEdge Rack and Blade サーバ**

<http://www.dell.com/downloads/global/power/ps2q09-20090246-Nguyen.pdf>

**2.Top 500 Supercomputers Over Time**

<http://www.top500.org/overtime/list/33/archtype>

**3.ClusterCorp Rocks+5.1 for Dell**

<http://www.dell.com/downloads/global/power/ps2q09-20090231-Clustercorp.pdf>

**4.Intel Xeon 5500 series Memory controller bandwidth**

<http://www.intel.com/cd/channel/reseller/asmo-na/eng/products/server/410125.htm>

**5.Intel QuickPath Interconnect**

<http://www.intel.com/technology/quickpath/introduction.pdf>

**6.Intel Turbo Boost Technology**

<http://www.intel.com/technology/turboboost/>

**7.Turbo Rating**

<http://www.intel.com/Assets/PDF/specupdate/321324.pdf> ,Table1

**8.Using Intel Hyper - threading Technology to Achieve Computational Efficiency**

<http://www.dell.com/downloads/global/power/4q03-cel.pdf>

## 付録 A- 調査結果のまとめ

### Power Management プロファイル

- 1) アイドル時における OS Control、APC、および Max Performance プロファイルの消費電力はどれも同様の結果でした。「図4 - Power Management プロファイルごとの相対アイドル時消費電力」を参照して下さい。
- 2) CPU への負荷の高い HPC ワークロード実行中における OS Control、APC、および Max Performance プロファイルの消費電力はどれも同様の結果でした。「図11 - Power Management プロファイルごとの性能／ワット」を参照して下さい。
- 3) 通信の多いクラスタ対応アプリケーション実行時は、APCの方が OS Control や Max Performance 消費電力プロファイルよりも消費電力が低下しました。「図11 - Power Management プロファイルごとの性能／ワット」を参照して下さい。

### 性能

- 4) 典型的な HPC ワークロードにおいては Node interleaving は無効にしておく必要があります。「図6 - ローカルに対するノードインターリーブの相対性能」を参照して下さい。
- 5) SMT はアプリケーション単位でテストし、必要に応じて無効もしくは有効にします。「図7 - 論理プロセッサを有効にしたときと無効にしたときの論理プロセッサの性能」を参照して下さい。
- 6) Turbo Mode はすべてのコアがフル活用されている場合でも性能を引き上げることができます。「図8 - Turbo Mode 有効時の CPU 周波数」を参照して下さい。
- 7) Turbo Mode を有効にすると大半の HPC ワークロードで性能が向上します。さらに、C-states を有効にすると Turbo Mode の動作する可能性が最大限に高まります。「図8 - Turbo Mode 有効時の CPU 周波数」および「図9 - Turbo Mode を有効にしたときのクラスタワークロードとインターコネクトの性能向上」を参照して下さい。
- 8) C-state を有効にするとレイテンシの影響が大きいアプリケーションの性能が低下する可能性があります。「図10 - C-state 移行時におけるレイテンシのオーバーヘッド」を参照して下さい。

### 消費電力効率

- 9) C-state を有効にすると、すべての Power Management プロファイルでアイドル時消費電力が最小限に抑えられます。「図5 - C-state 有効時のアイドル時消費電力削減量」を参照して下さい。
- 10) 一般的に、消費電力増加分より性能向上分の方が低くなる傾向があるため、Turbo Mode を有効にすると消費電力効率は下がります。「図12 - 4種のクラスタ対応アプリケーションを実行したときの消費電力効率比較」を参照して下さい。
- 11) 一部のアプリケーションでは、性能を最大限に引き上げるために消費電力を節約する BIOS 機能を無効にする必要があります。消費電力の低いコンポーネントは、それを搭載するクラスタの消費電力効率を高めます。「図13 - 低消費電力 CPU とメモリを採用したときの消費電力効率向上」を参照して下さい。

## 付録 B - Linux コマンドラインから BIOS オプションを変更する DTK

「ClusterCorp Rocks+ 5.1 for Dell」には、「Dell OpenManage™ Deployment Toolkit」(DTK) をクラスターの各ノードにインストールする Dell Roll が含まれています。Dell では、PowerEdge サーバ用に「Dell OpenManage™ Deployment Toolkit」(DTK) を無償でダウンロードすることができます。

DTK は、Linux OS から BIOS 設定を変更できるようにする「syscfg」というユーティリティを提供しています。このコマンドラインユーティリティは、クラスターノード上で並行して実行したり、スクリプトを書くことが可能で、HPC 環境で使いやすくなっています。今回の調査ではこれが幅広く利用されています。

syscfg は以下のように使います。

```
# cd /opt/dell/toolkit/bin
# ./syscfg < コマンドラインのヘルプが表示されます >
```

BIOS オプション	現在の設定を表示させる	設定を変更する
Node Interleaving	<code>./syscfg --nodeinterleave</code>	<code>./syscfg --nodeinterleave=enable   disable</code>
SMT	<code>./syscfg --logicproc</code>	<code>./syscfg --logicproc=enable   disable</code>
Turbo Mode	<code>./syscfg --turbomode</code>	<code>./syscfg --turbomode=enable   disable</code>
C-states	<code>./syscfg --cstates</code>	<code>./syscfg --cstates=enable   disable</code>
Power Profile	<code>./syscfg power</code>	<code>./syscfg power --profile=&lt;profile&gt; --setuppwdoverride</code> <profile> は maxperformance、osctl、apc、custom のいずれか

syscfg を使って BIOS オプションを変更してそのオプションの現在値を読むと設定された新しい値が表示されます。しかし、この設定は再起動するまで反映されません。たとえば次のようになります。

SMT は現在「無効」に設定されています。./syscfg --logicproc では「logicproc=disable」と表示されます。

SMT を「有効」に変更します。./syscfg --logicproc=enable。

ここで SMT の値をもう一度読みます。すると、./syscfg --logicproc では「logicproc=enable」と表示されます。

しかし SMT の有効状態は再起動するまで反映されません。

このことは上述のすべての BIOS オプションに当てはまりますが、消費電力プロファイルだけは例外です。消費電力プロファイルは現在機能しているプロファイルを常時表示します。syscfg を使ってこのプロファイルを変更したときは、システムを再起動しないと syscfg は新たに有効になったプロファイルを表示しません。

DTK に関する詳細は <http://support.dell.com/support/edocs/software/dtk/> を参照して下さい。

## 付録 C - 各種ベンチマーク

### 1.Stream - v5.8

<http://www.cs.virginia.edu/stream/>

### 2.lat\_mem\_rd v1.13 from LMBench3

<http://www.bitmover.com/lmbench/>

### 3.DGEMM from Intel MKL 10.1.2

<http://software.intel.com/en-us/intel-mkl/>

### 4.HPL v2.0

<http://www.netlib.org/benchmark/hpl/>

### 5.Fluent: Ansys Fluent v12, Fluent benchmarks v6.3.26

<http://www.ansys.com/products/fluid-dynamics/fluent/>

### 6.Ansys: Distributed Ansys Workbench v12, Distributed Ansys benchmarks v12.

<http://www.ansys.com/products/workbench/default.asp>

### 7.ECLIPSE : Schlumberger の Reservoir Simulation Suite 2008.2 (2008 年 11 月)

### 8.WRF v3.1. Data set Conus 12km.

<http://www.mmm.ucar.edu/wrf/users/>

### 9.LU: NAS Parallel Benchmarks. NPB - MPI v3.3, LU class D.

<https://www.nas.nasa.gov/cgi-bin/software/start>

