

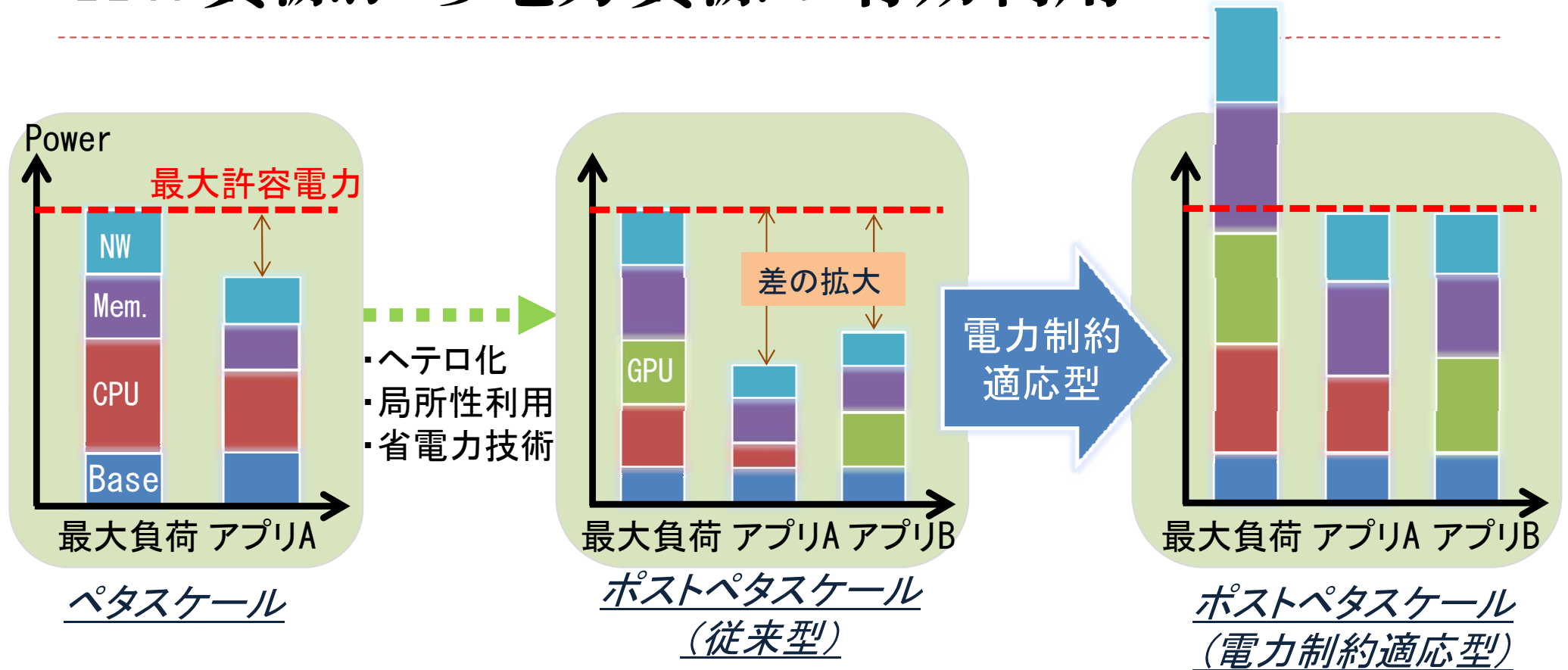
ミニアプリを用いたHPCシステムの の電力解析

電気通信大学 大学院情報システム学研究科
近藤 正章、カオ タン、和田 康孝

HPCシステムの消費電力問題

- ▶ **ポストペタ時代のシステムは消費電力が最大の設計制約**
 - ▶ 京コンピュータでは10PFLOPSを13MWで実現
 - ▶ 将来的にも20~30MWが電力供給の限度
 - ▶ エクサシステムでは同程度の電力で100倍の性能向上が必要
- ▶ **アプリケーションのシステムへの要求の多様化**
 - ▶ 計算・記憶・通信の各要素への要求が異なる
 - ▶ 電力がシステム制約となる状況下では各要素へ投入するハードウェア資源は制限せざるを得ない
- ▶ **運用時のピーク電力が制約を超えないことを保証する worst case設計ではシステムをスケールさせることは難しい**
- ▶ **JST CREST「ポストペタスケールシステムのための電力マネジメントフレームワークの開発」**

HW資源から電力資源の有効利用へ



電力制約適応型システム

- ▶ 最大負荷時電力が電力制約を超過することを積極的に許容
- ▶ **電力性能ノブ**を制御することで実効電力を制約以下に抑制
- ▶ 電力資源を計算・記憶・通信へ適応的に配分することで実効性能向上へ

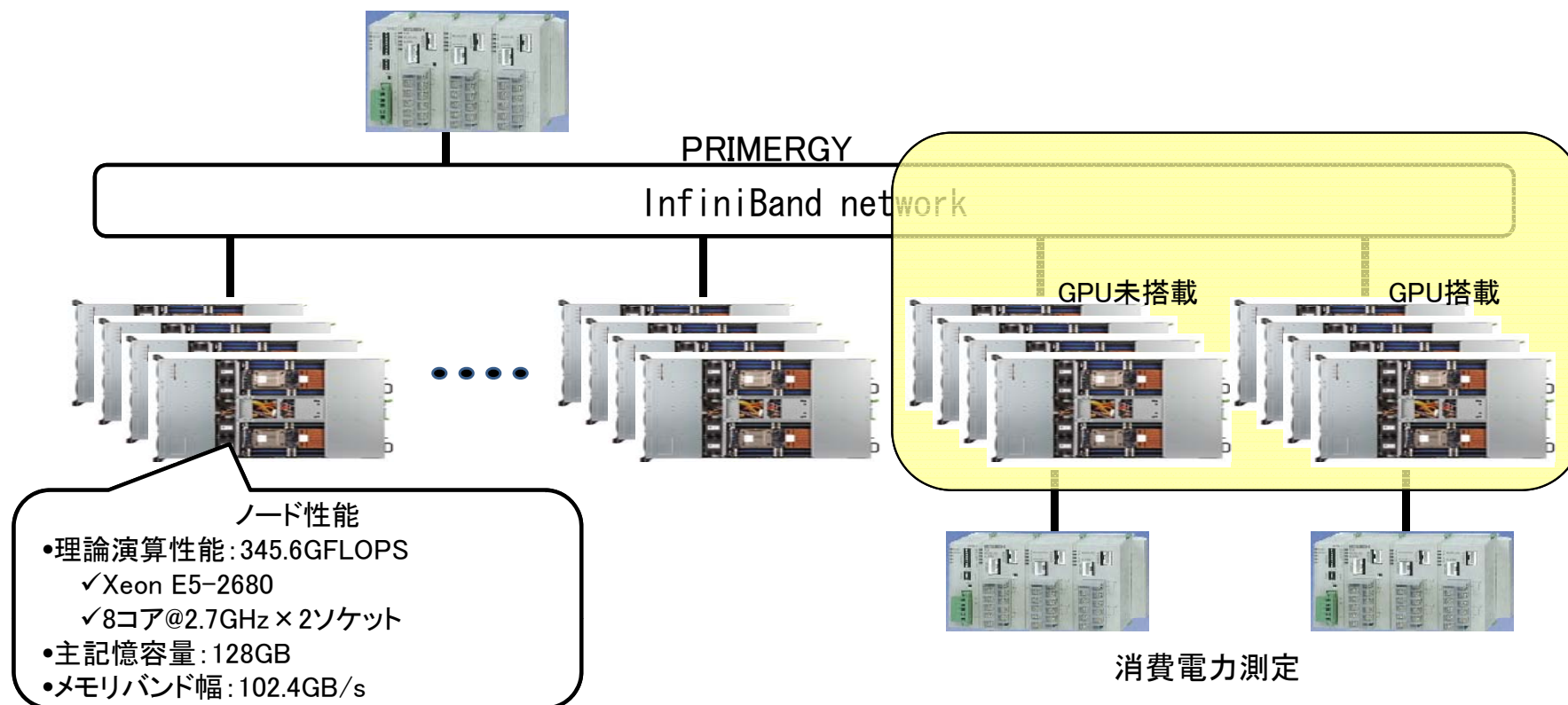
→ 種々のアプリでの電力消費状況の解析が重要

電力解析・制御環境

▶ 消費電力測定・制御環境

▶ 九州大学情報基盤研究センター CX400

- ▶ 電力メータを用いた粗粒度な消費電力測定環境
- ▶ Intel社のRAPL (Running Average Power Limiting)を用いた細粒度な消費電力測定環境

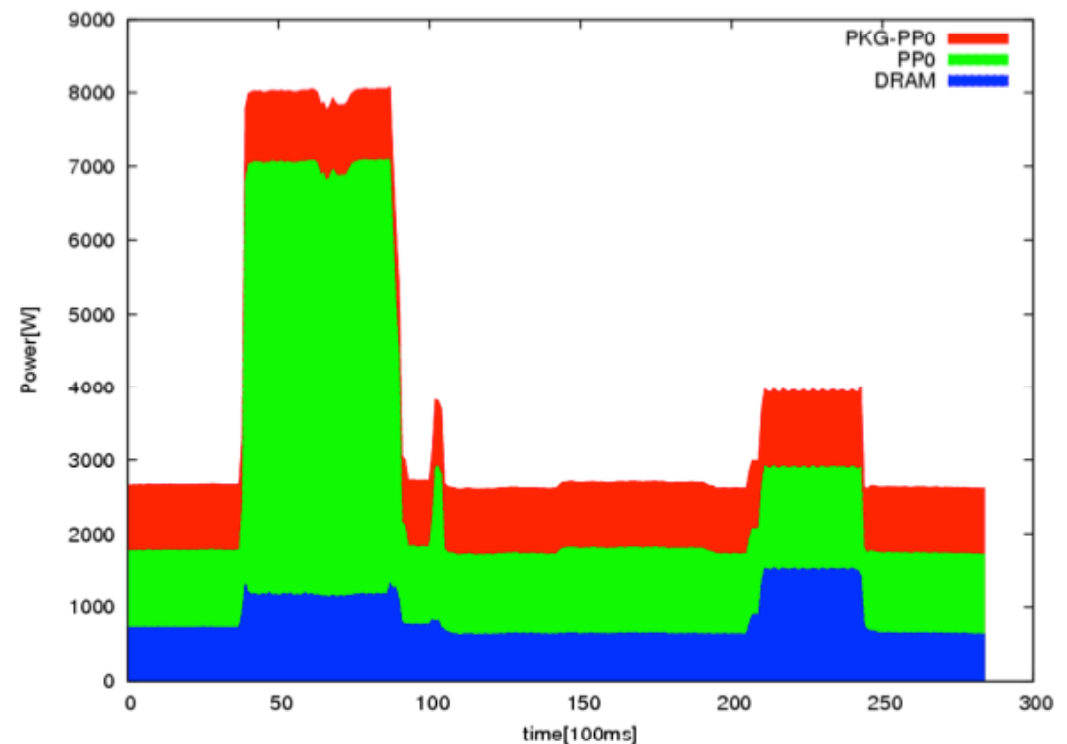
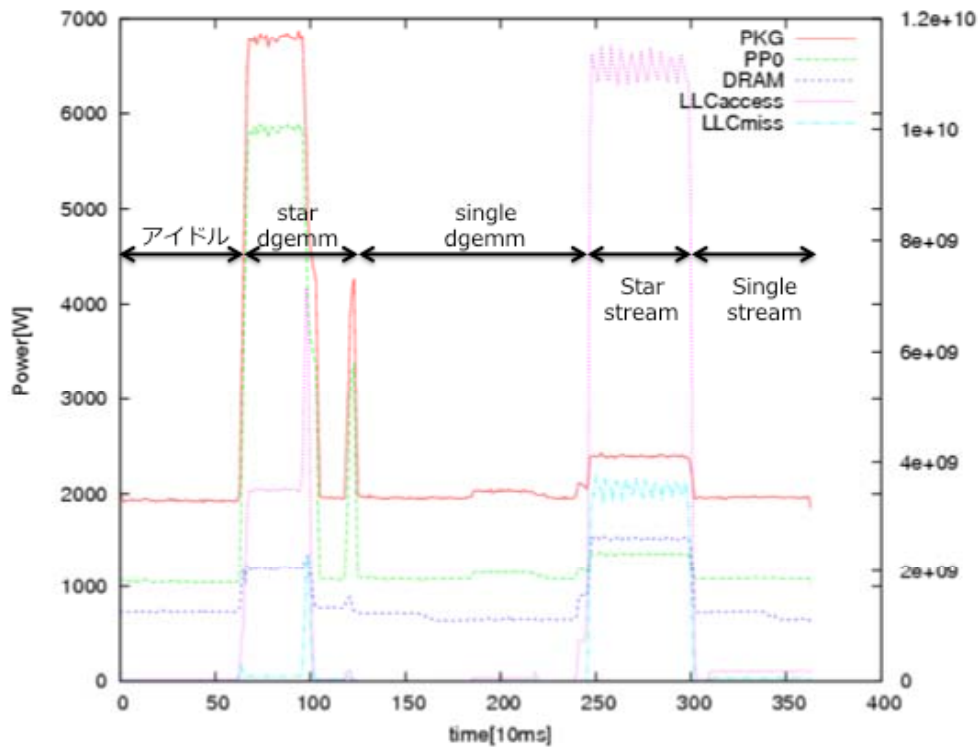


Intel RAPLおよびRIC400 (RAPL-IF-CX400)

- ▶ Intel RAPL (Runing Average Power Limit)
 - ▶ Sandy Bridge以降に搭載されている電力観測・制御インタフェース
 - ▶ MSRの読出し/書き込みにより電力値取得・制約設定
 - ▶ ソケット単位の観測と制御
- ▶ 九州大学で開発したRAPL用インタフェース
 - ▶ CX400において消費電力測定/制御を実現するための暫定API
 - ▶ アプリ・ソースコード中にてAPIを呼び出し
- ▶ 消費電力観測
 - ▶ 固定周期連続観測: 指定した時間経過毎に消費電力を記録
 - ▶ 指定区間観測: 指定した区間の平均消費電力を記録
- ▶ 消費電力制約制御
 - ▶ API呼び出し時点で消費電力制約を設定／解除

消費電力値出力表示例

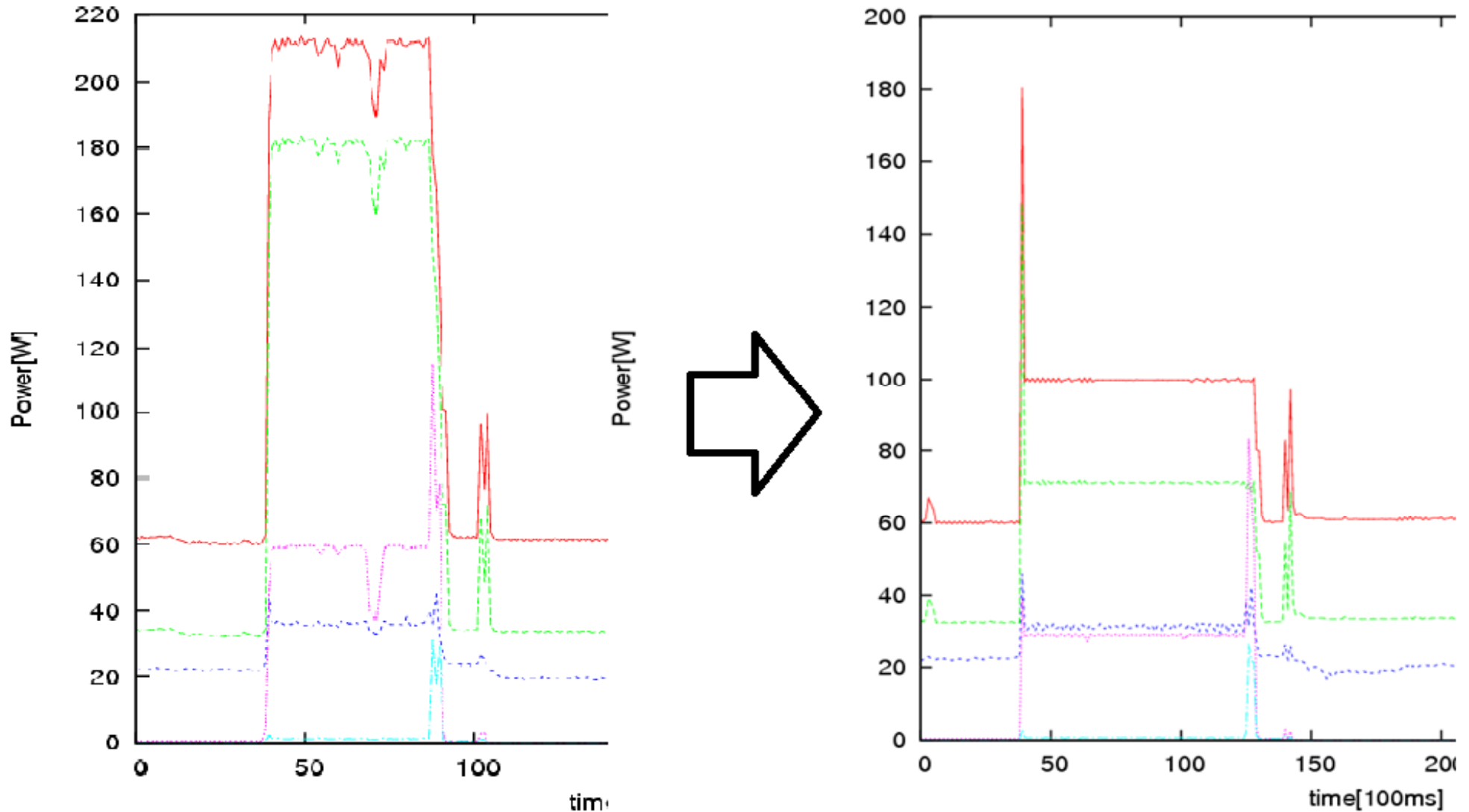
▶ HPCCベンチマークを用いた測定例



- Xeon E5-2680 32ノード
- star-dgemm→single-dgemm→star-stream→single-stream

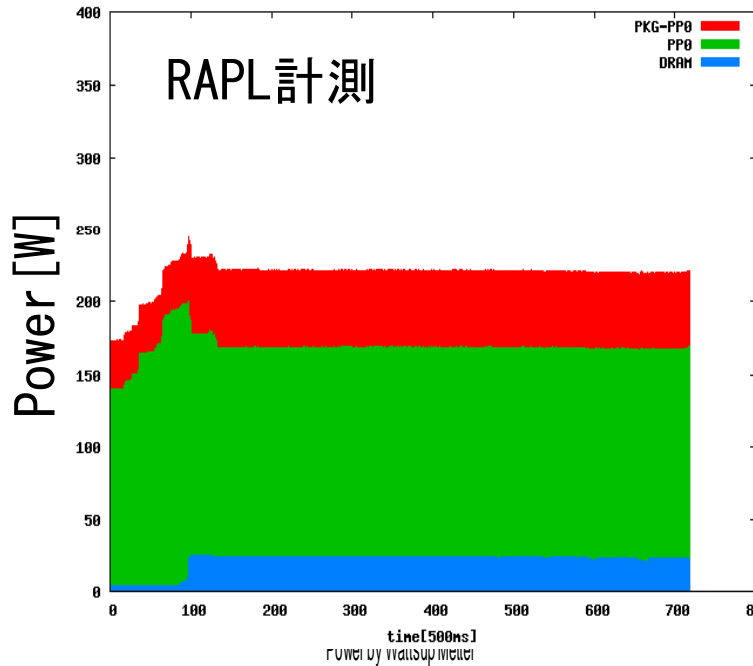
消費電力制約の設定例：CPU

▶ パッケージ電力(PKG)を50Wに設定

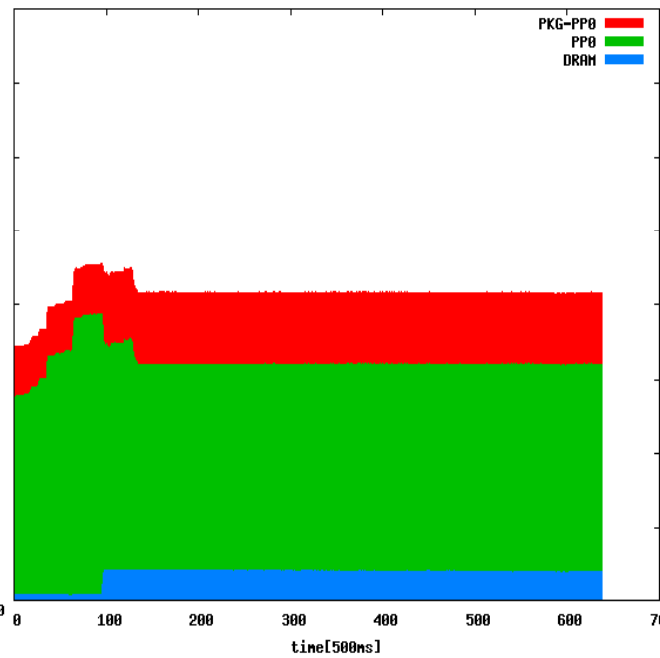


RAPLの電力計測精度検証

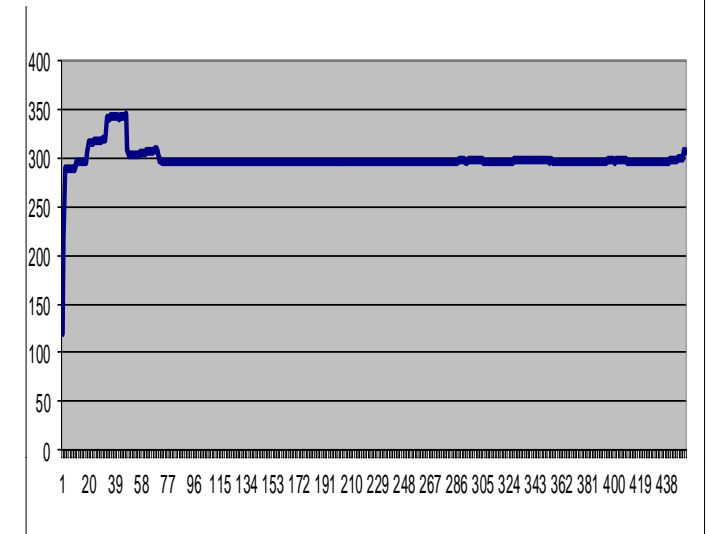
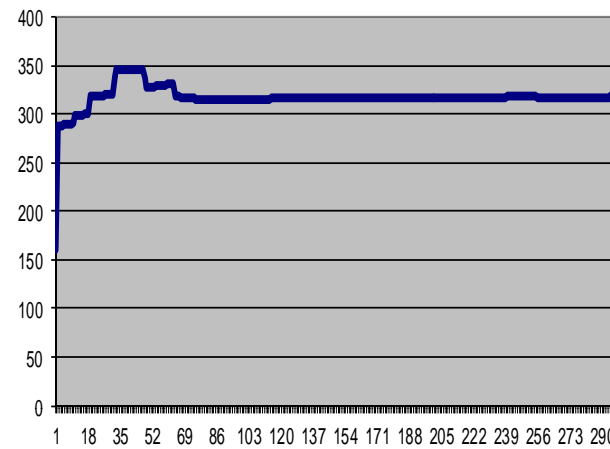
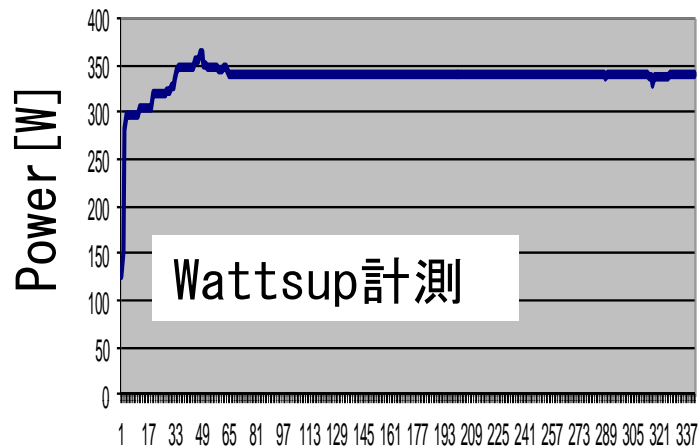
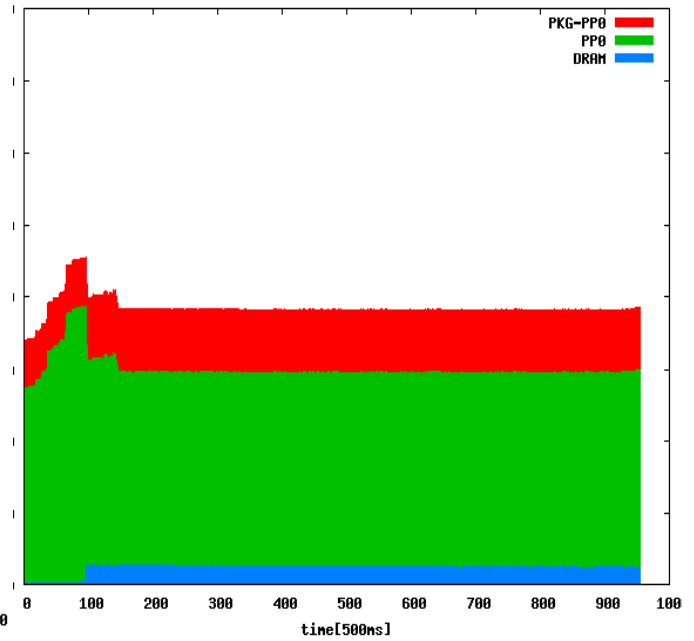
テスト1



テスト2



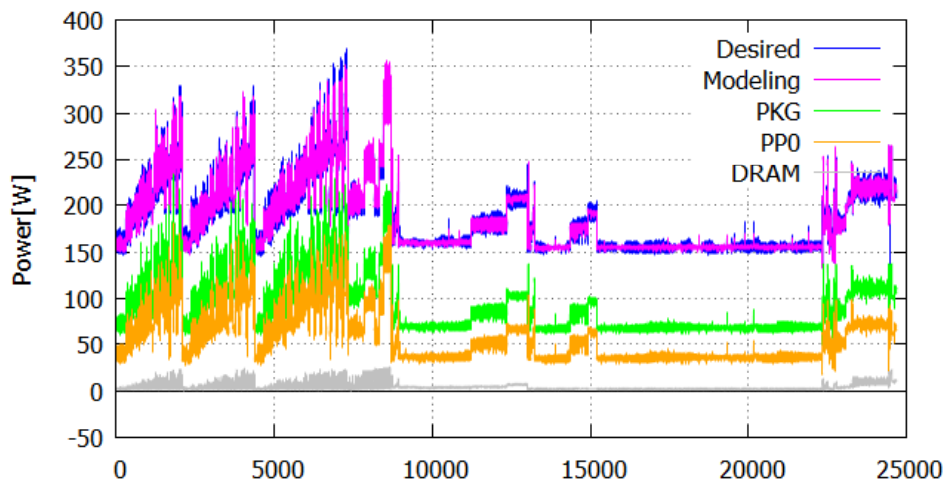
テスト3



RAPLによるシステム電力のモデリング

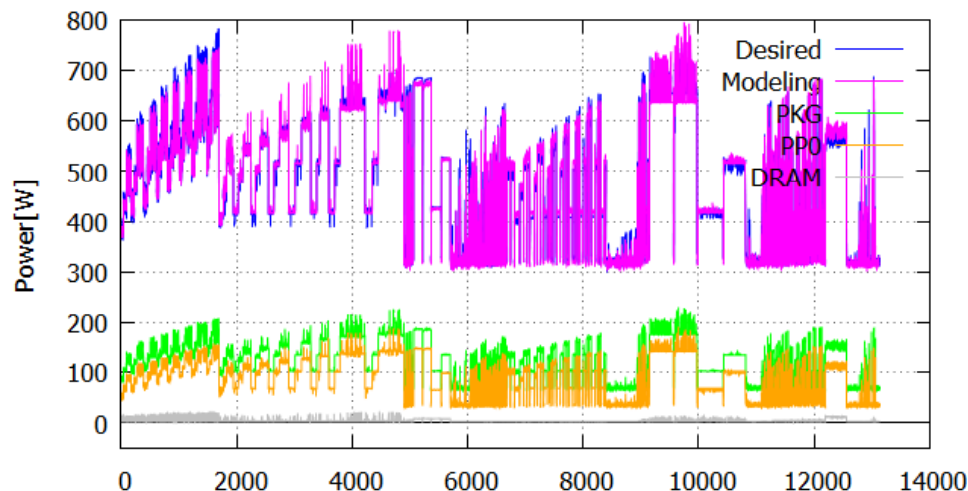
1ノードのモデリング結果

All Testing Data: MSE = 2.09E-4, Average Error = 4.21W

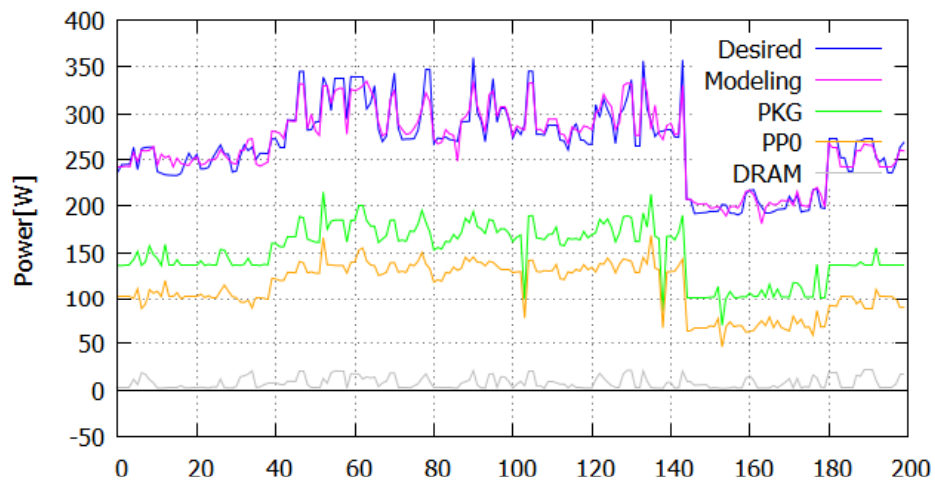


2ノード+スイッチのモデリング結果

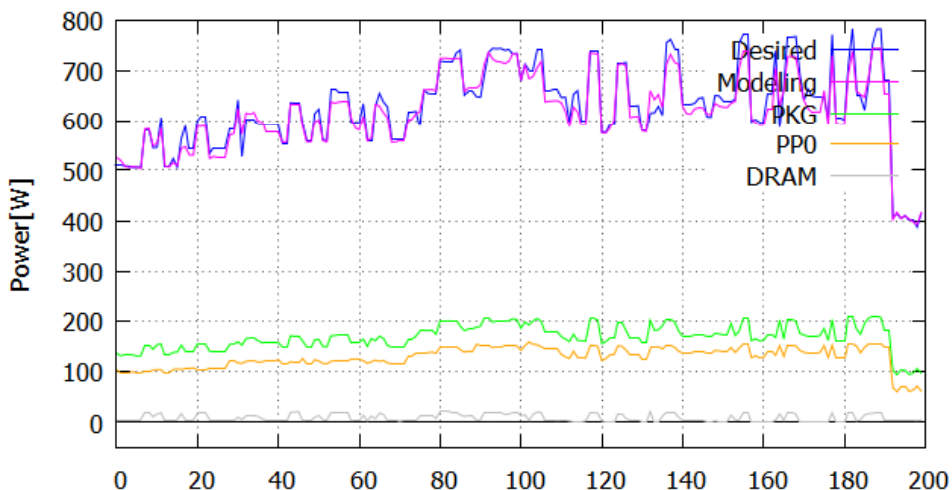
All Testing Data: MSE = 3.25E-4, Average Error = 9.66W



Detailed Data in [7000-7200]



Detailed Data in [1500-1700]



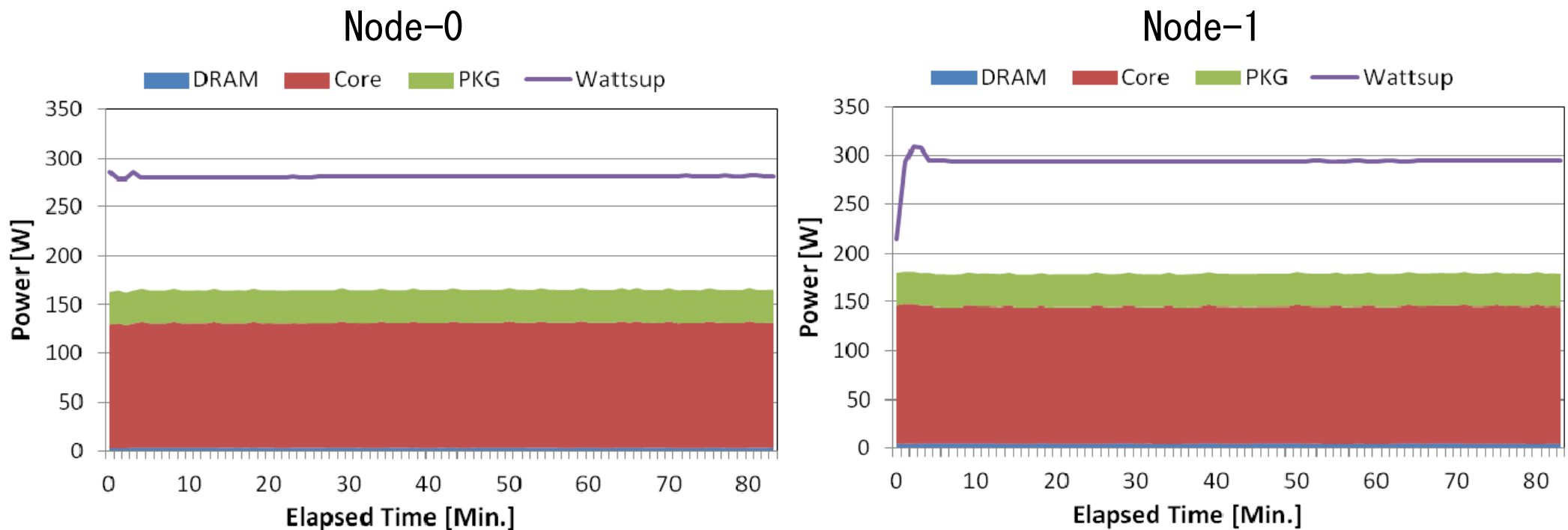
→ 高い精度でシステム全体電力がモデリング可能

使用したミニアプリ

- ▶ FFV-C (FrontFlow/VioletCartesian)
 - ▶ 次元非定常非圧縮熱流体解析ソルバー
 - ▶ 直交等間隔格子上で三次元非定常非圧縮性熱流体を解析
- ▶ CCS QCD
 - ▶ 筑波大学計算科学研究センターで開発されたQCD計算のプログラム
- ▶ Flat MPIにより実験

電力測定・制御結果 (tiny cluster system)

- ▶ FFVCミニアプリ
 - ▶ 2ノード、32コア、電力制限なし



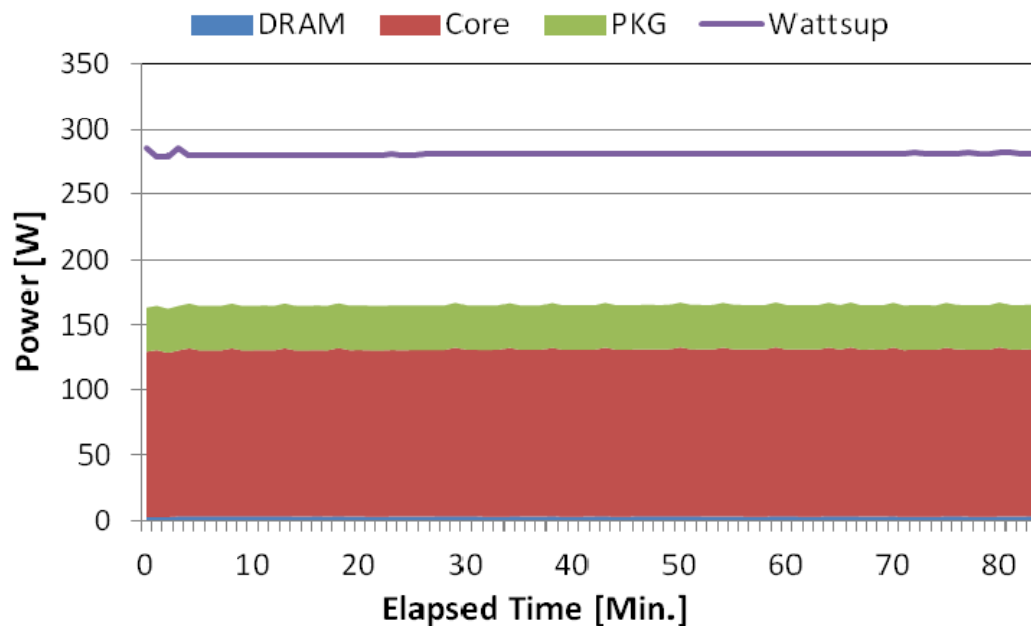
- ▶ アプリケーション内での電力変動小、ノード間のばらつき小
- ▶ 本システムではDRAM電力消費が小さい (2~6[W]/node)

電力測定・制御結果 (tiny cluster system)

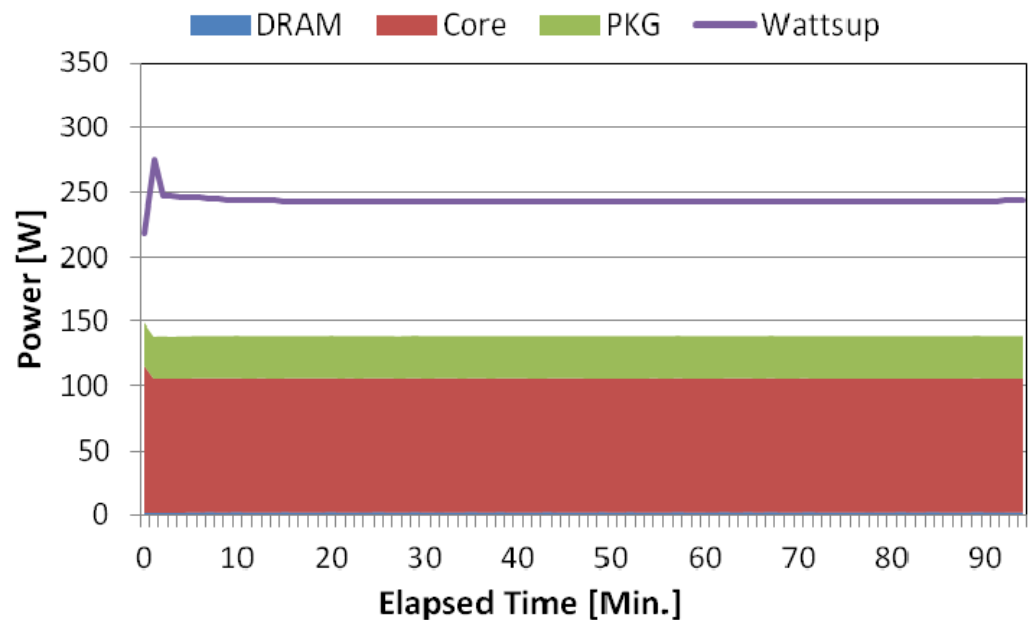
▶ FFVCミニアプリ

- ▶ 2ノード、32コア、電力制限なし vs. 電力制限あり

Node-0, 電力制限なし



Node-0, 62[W]制限/1ソケット

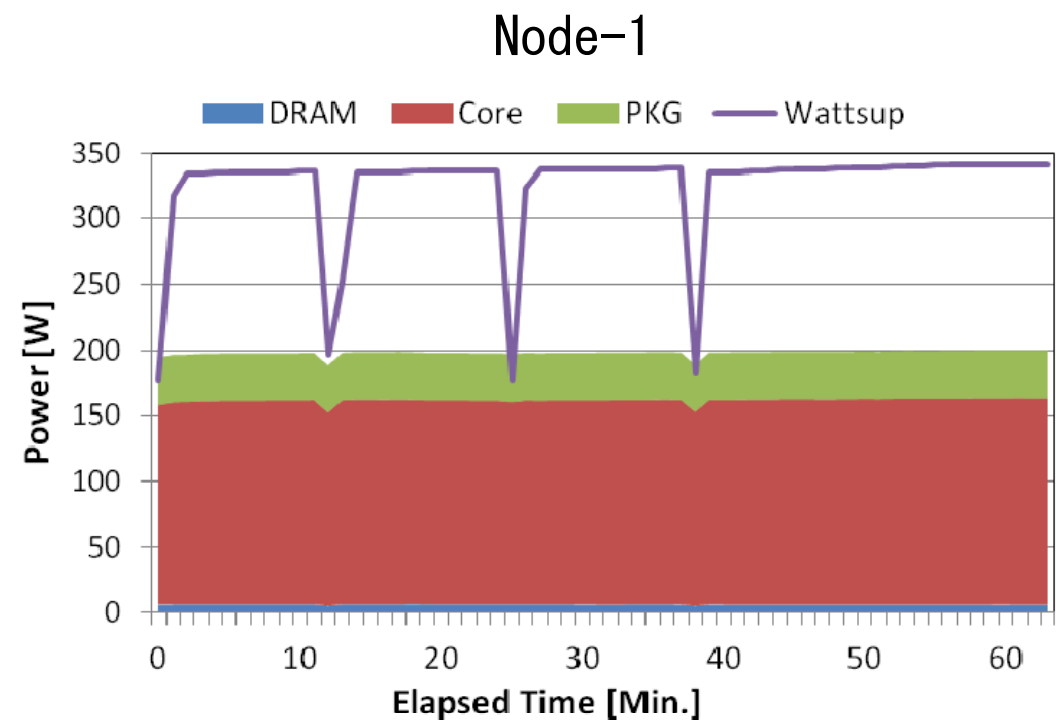
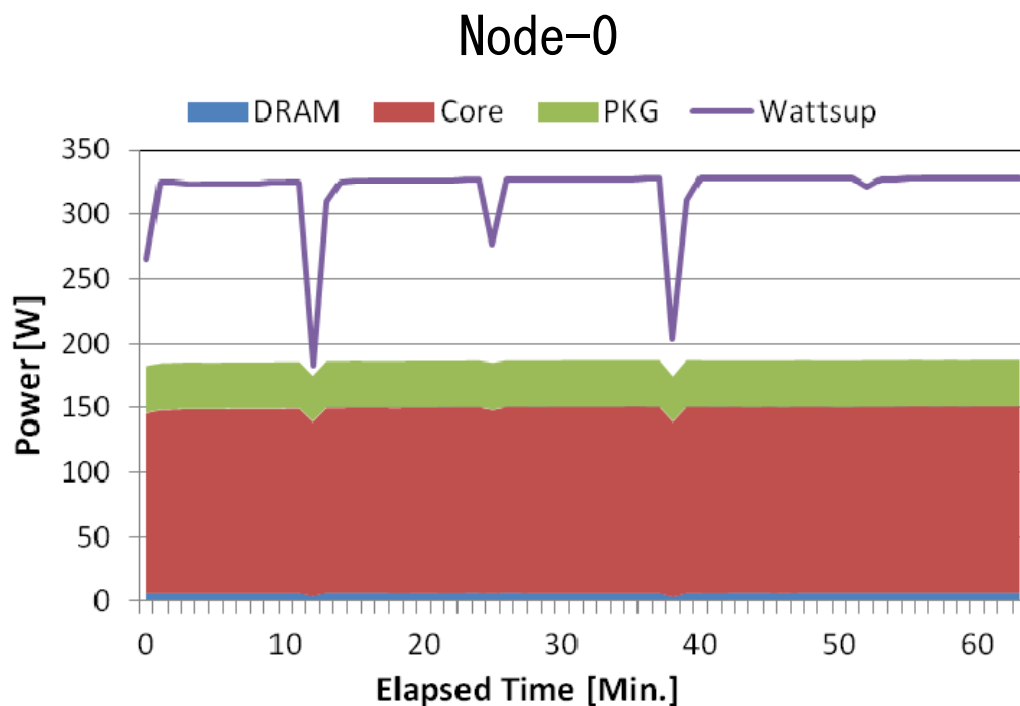


- ▶ 電力制限によりほぼ制限値まで消費電力低下
- ▶ ただし実行時間が増加

電力測定・制御結果 (tiny cluster system)

▶ CCS-QCDミニアプリ

- ▶ 2ノード、32コア、電力制限なし、5回繰り返して実行



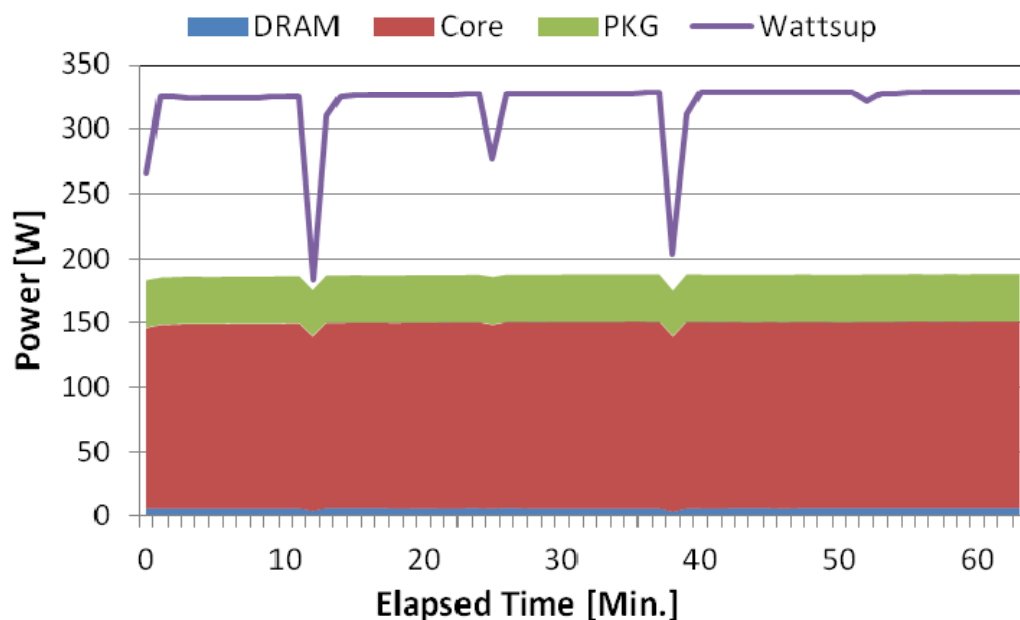
- ▶ RAPL測定での電力変動小、ノード間のばらつき小
- ▶ DRAM電力消費が小さい (6~7[W]/node)

電力測定結果 (tiny cluster system)

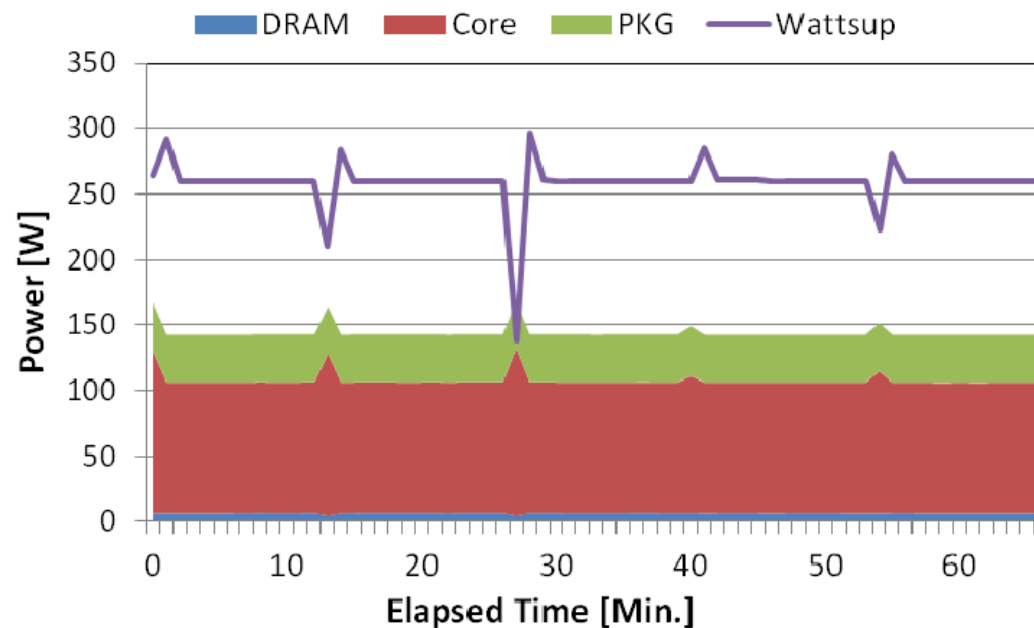
▶ CCS-QCDミニアプリ

- ▶ 2ノード、32コア、電力制限なし vs. 電力制限あり

Node-0, 電力制限なし



Node-0, 62[W]制限/1ソケット

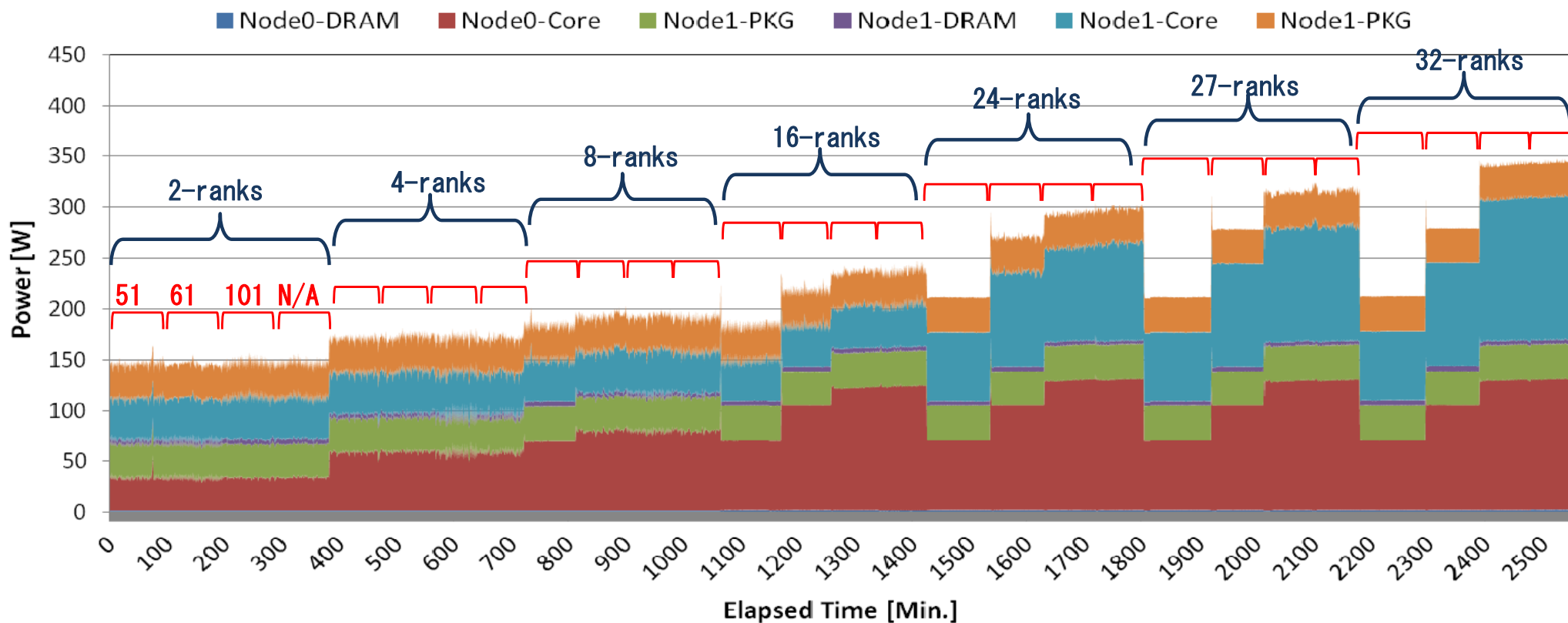


- ▶ 電力制限によりほぼ制限値まで消費電力低下
- ▶ 実行時間は多少増加

電力測定・制御結果 (tiny cluster system)

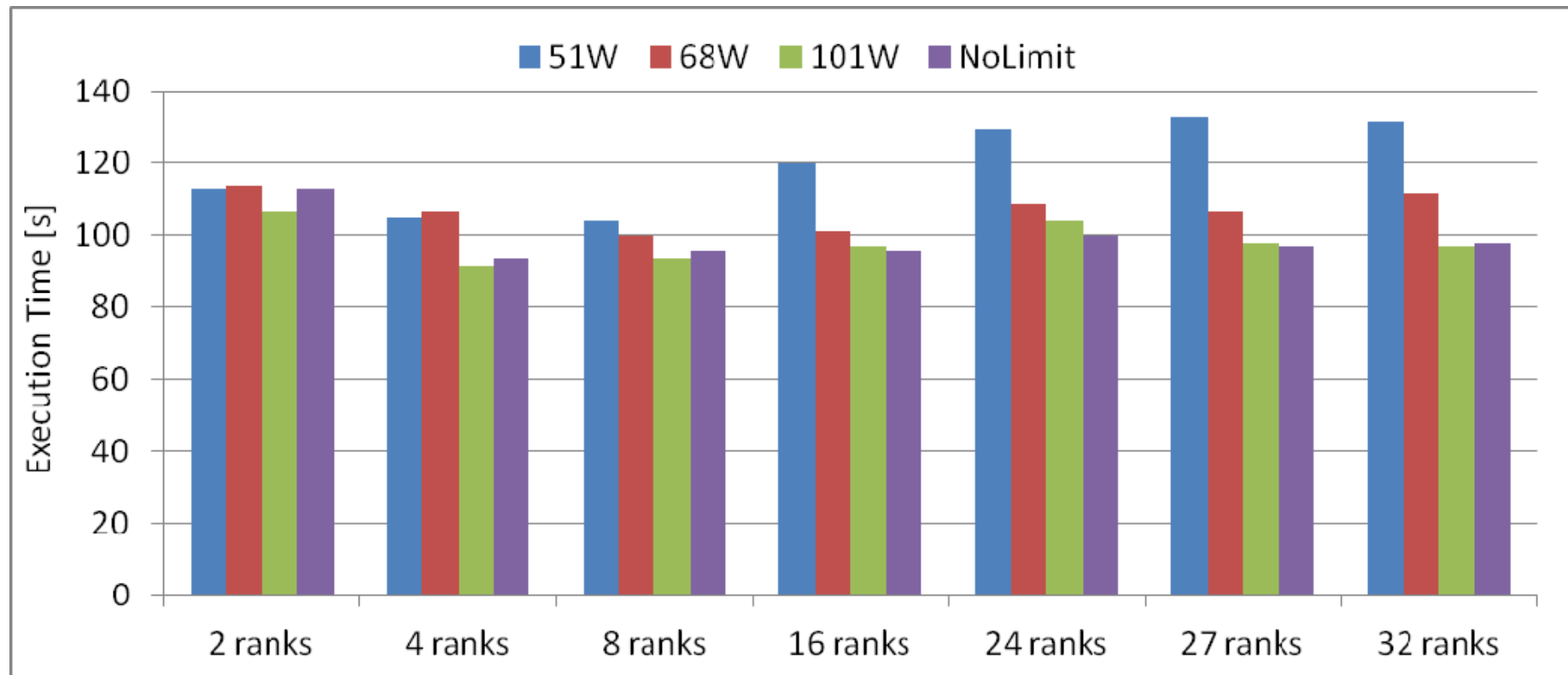
▶ FFVCミニアプリ

▶ rankおよび電力制限値を変化



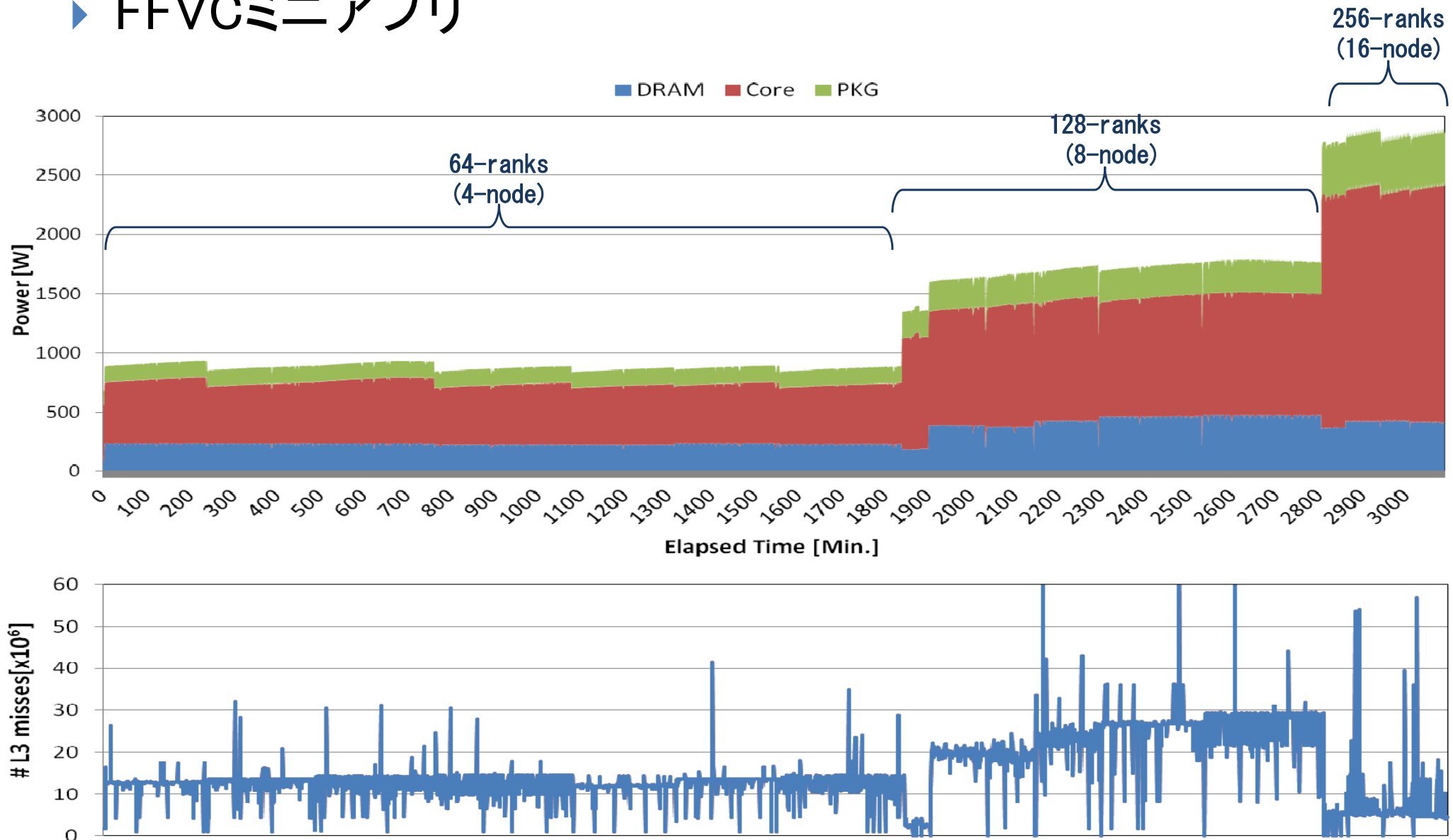
電力制約下での性能比較

- ▶ FFVCミニアプリ
 - ▶ rankおよび電力制限値を変化



電力測定結果 (九大センター CX400)

▶ FFVCミニアプリ



まとめ

- ▶ ミニアプリを用いた電力消費特性の解析
 - ▶ RAPLを用いた電力モニタリングをミニアプリに適用
 - ▶ 電力制約を与えた際の電力挙動の観測

- ▶ 将来展望
 - ▶ 種々のアプリで電力性能挙動を観測
 - ▶ 電力マネージメントフレームワークへのフィードバック

- ▶ 第141回HPC研究会@沖縄で発表予定
 - ▶ 「CPUと主記憶への電力バジェット配分を考慮したHPCアプリケーションの性能評価」
 - ▶ 「RAPLインタフェースを用いたHPCシステムの消費電力モデリングと電力評価」