

Energy-Efficient Cluster Computing



<http://www.eeclust.de>

April 2009 – März 2012

Timo Minartz

Universität Hamburg

Arbeitsbereich Wissenschaftliches Rechnen



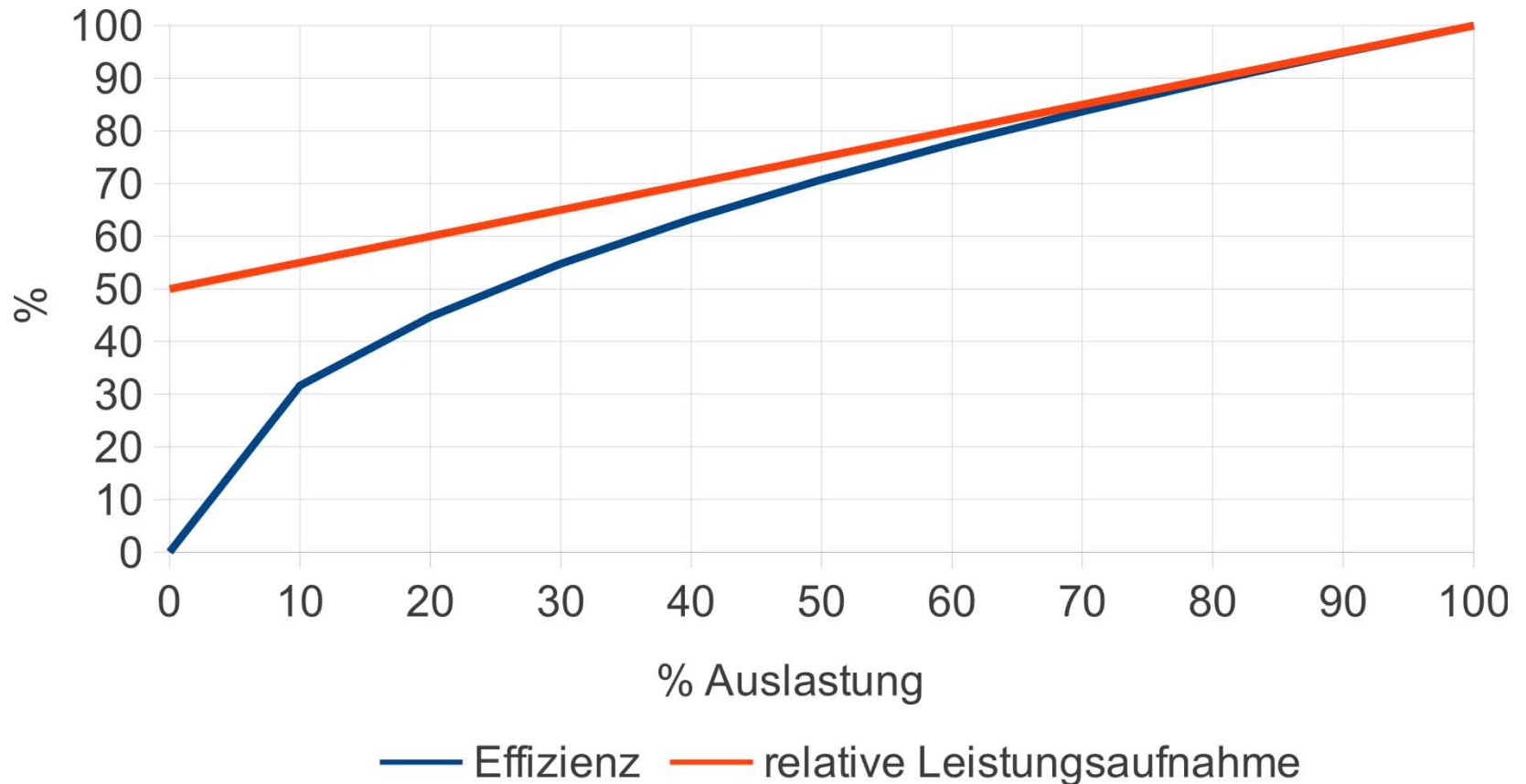
Bundesministerium
für Bildung
und Forschung

Konsortium

- Universität Hamburg (Koordinator)
 - Thomas Ludwig, Timo Minartz
- Jülich Supercomputing Center
 - Bernd Mohr, Michael Knobloch
- Technische Universität Dresden
 - Wolfgang Nagel, Matthias Müller, Daniel Molka
- ParTec Cluster Competence Center München
 - Hugo Falter, Stephan Krempel



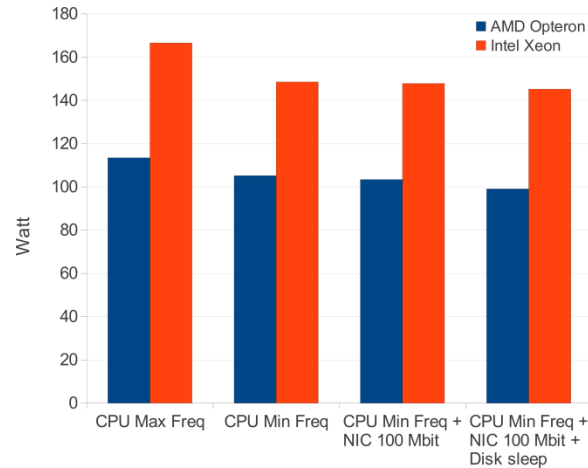
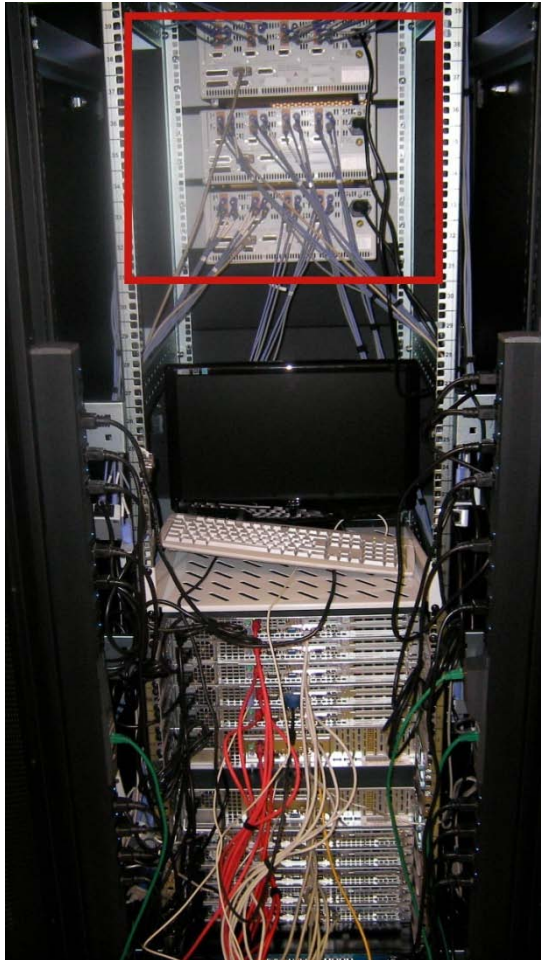
Problem: Leistungsproportionalität



Projekt-Ziel

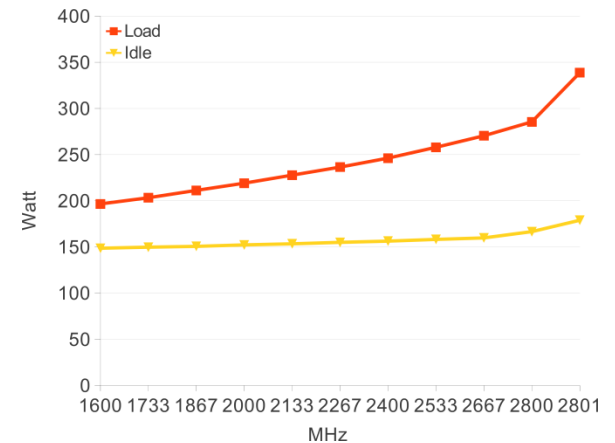
Reduzierung des Energieverbrauchs eines Rechenclusters ohne die Verlängerung von Programmlaufzeiten durch das temporäre Umschalten von Betriebsmodi

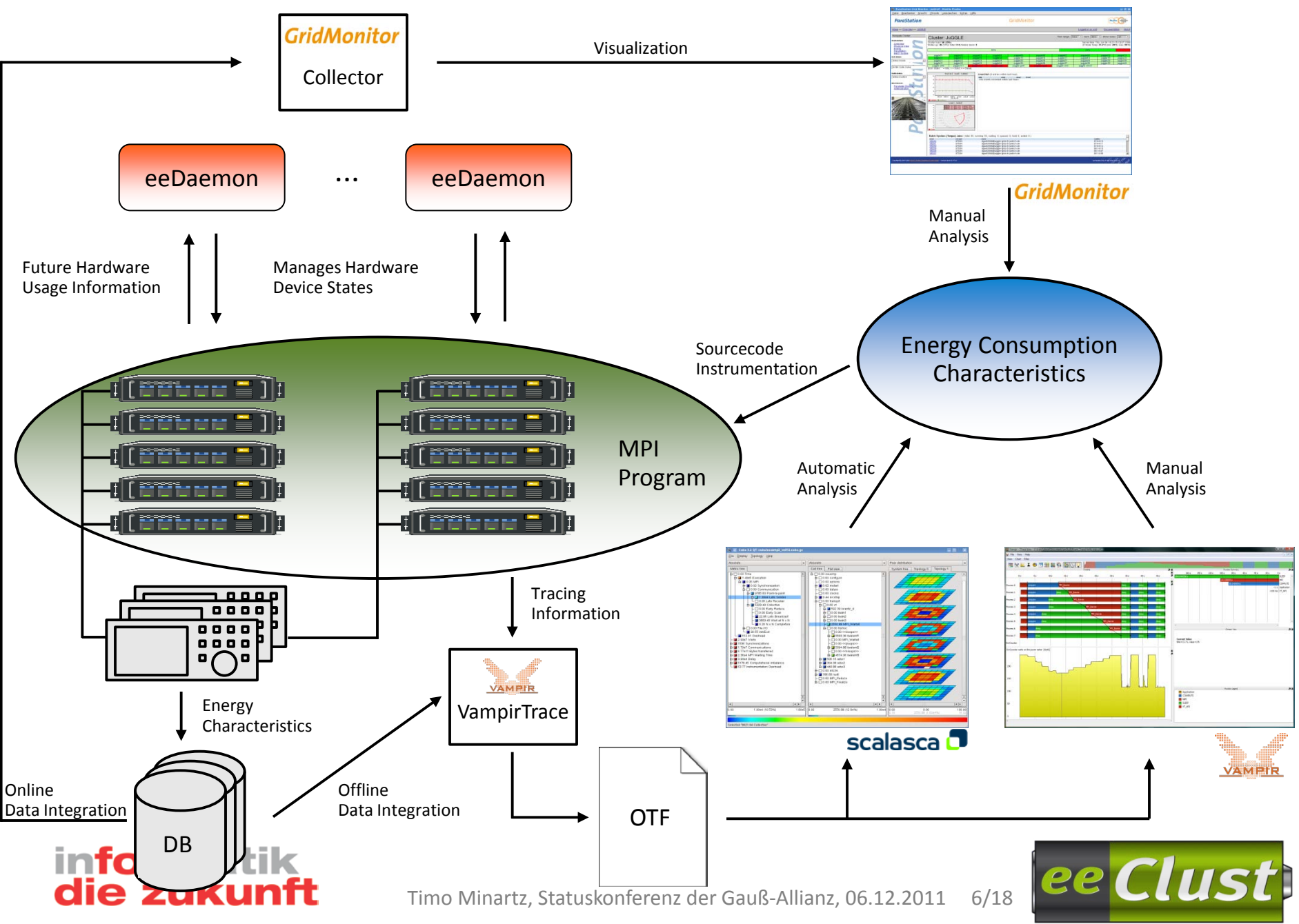
Stromsparpotential



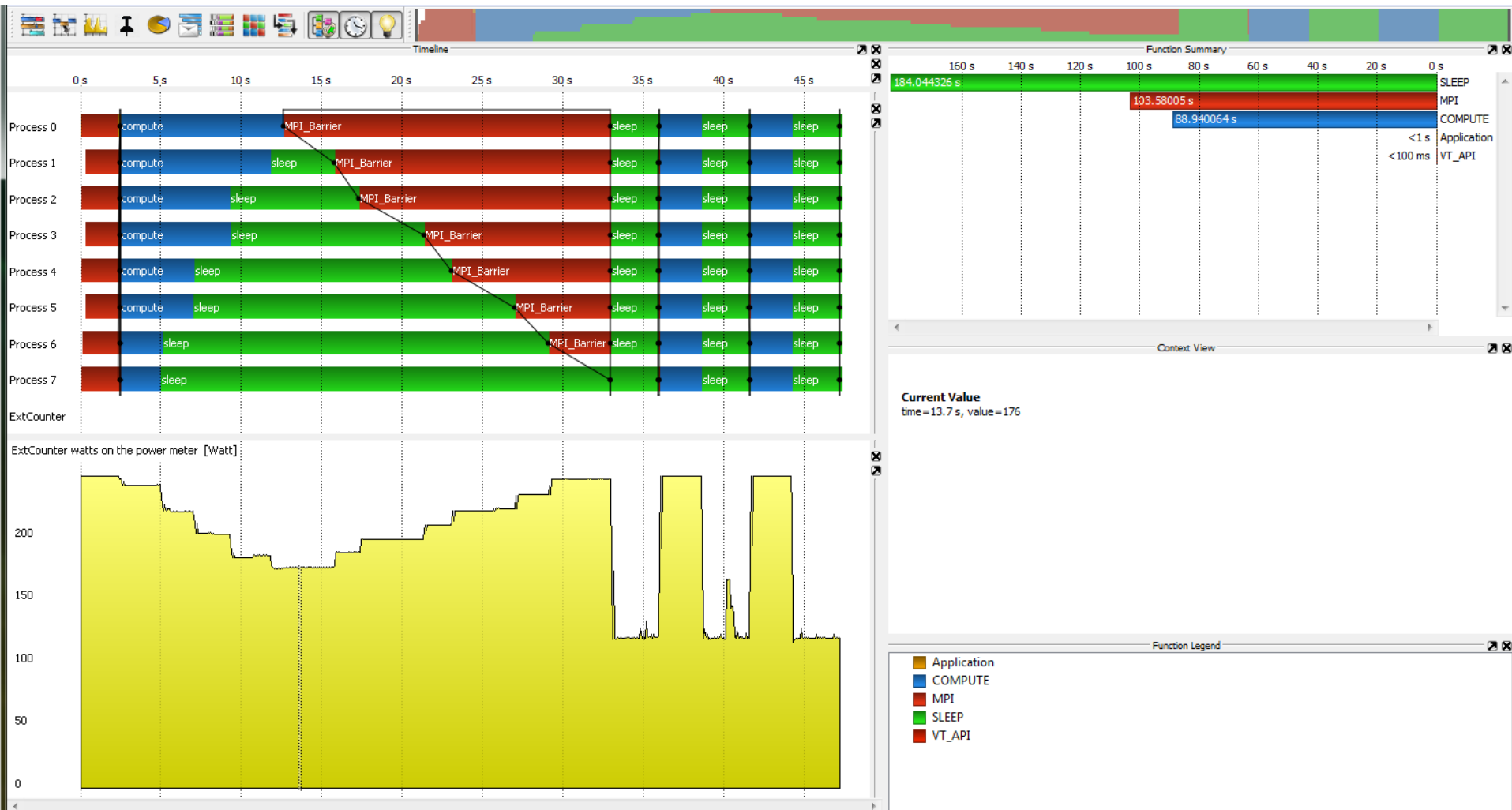
Leistungsaufnahme in Abhängigkeit von den Betriebsmodi der Komponenten

Leistungsaufnahme in Abhängigkeit von der Prozessorfrequenz



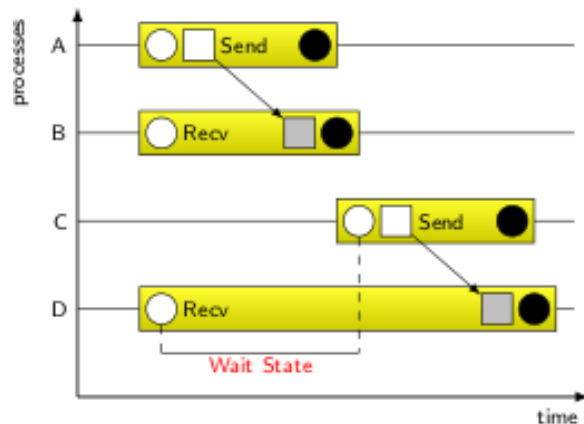


VampirTrace-Erweiterung



Scalasca-Erweiterung

- Bestimmung des Energiesparpotentials
- Basierend auf der Scalasca “Wait-State Detection”
 - z.B.. Late Sender, Late Receiver, Wait at Barrier, Late Broadcast, ...



- Berechnung der Energie in einem bestimmten Energiesparmodus
 - Inkl. Umschaltzeiten

Knobloch et al.: Determine Energy-Saving Potential in Wait-States of Large-Scale Parallel Programs (EnA-HPC'11)

Steuerprozess: eeDaemon

- Dienst, der den Betriebsmodus der Komponenten explizit steuert
 - Prozessor(kerne)
 - Frequenz (P-States)
 - Festplatten
 - Rotation
 - Pufferung
 - Netzwerkkarten
 - Geschwindigkeit
- Wird durch Quellcodeinstrumentierung gesteuert
 - Definierter Zustand für jede Phase im Programmablauf

Minartz et al.: Managing Hardware Power Saving Modes for High Performance Computing (IGCC'11)

Energieeffizienzbenchmarks

- Engagement in SPEC High Performance Group
 - MPI2007 Verbrauchsanalysen
 - Entwicklung einer Effizienzmetrik für OMP 2012 (SC 2011 BoF)
- eeMark - Energieeffizienzbenchmark
 - Bewertung der Energieeffizienz von HPC Systemen
 - Lastprofile erzeugen definierte Last für Prozessoren, Speicher, Verbindungsnetzwerk und Dateisystem
 - Evaluierung von Durchsatz und Effizienz
 - Untersuchung der Einflüsse von Betriebsmodi
 - Bestimmung des Einsparpotentials und der Laufzeitverlängerung

Molka et al.: Flexible Workload Generation for HPC Cluster Efficiency Benchmarking (EnA-HPC'11)

Navigate :

Power Details:

- [Overview](#)
- [Power & Load](#)

Cluster Details:

- [Overview](#)
- [Physical View](#)
- [Events](#)
- [ParaStation](#)
- [Batch System](#)
- [Power info](#)

Node Details:

Select node ▼

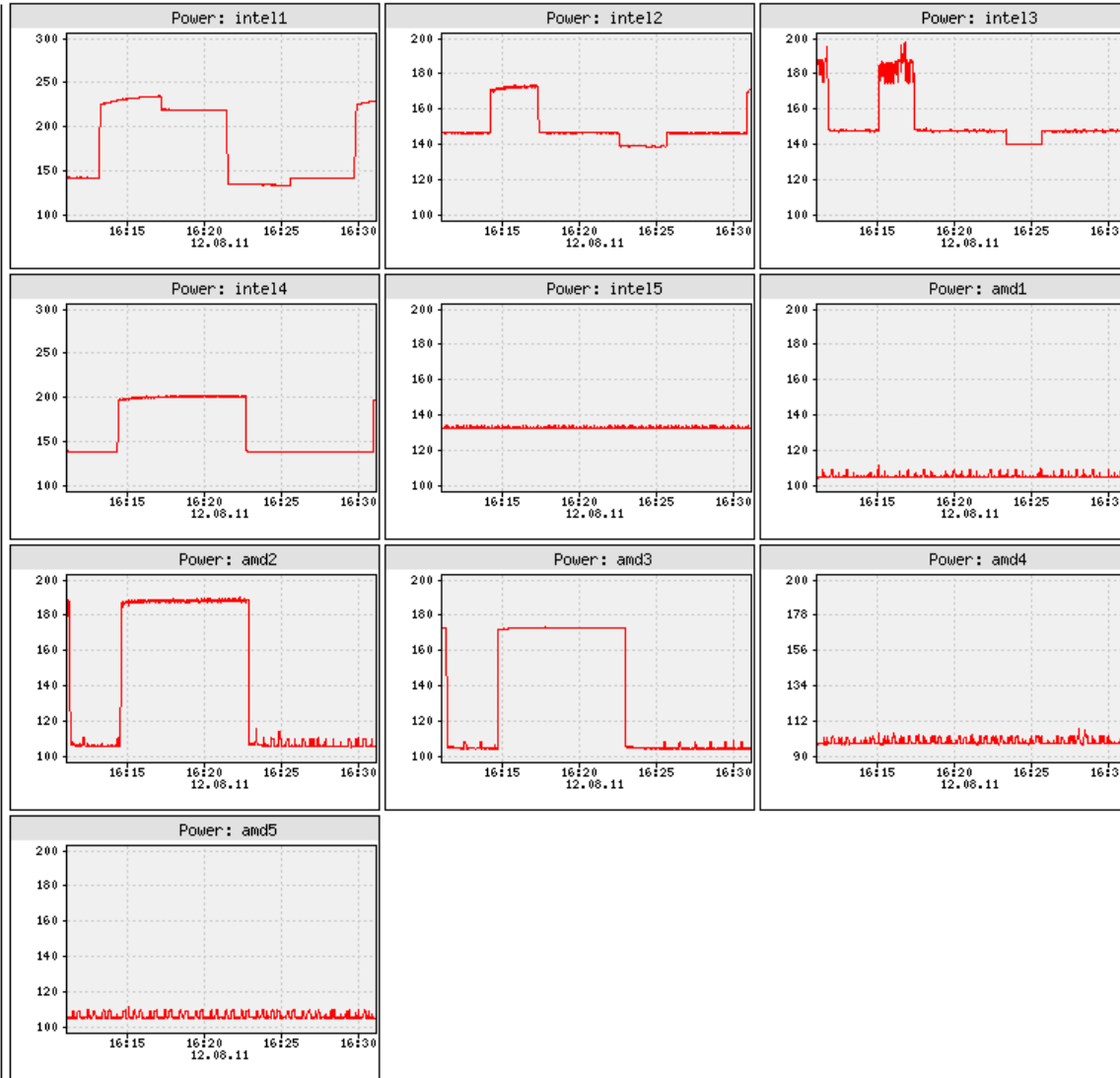
Enter node name

Switch Details:

Select switch ▼

Miscellaneous:

- [Parameter Browser](#)
- [Administration](#)

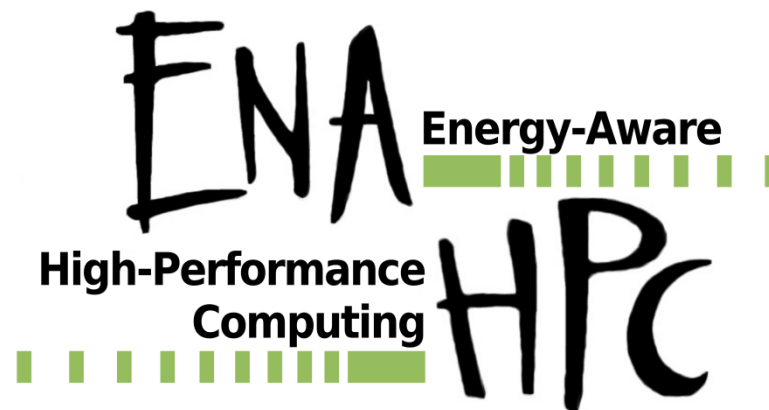


Ergebnisse

- Abhängig vom Zugriffsmuster / der Hardware kann Energie gespart werden
 - Bis zu 10 % für reale Anwendungen
 - Ohne oder mit geringer Laufzeitverlängerung
- Integration der Leistungsaufnahme in Clusterverwaltung
 - Monetäre Bewertung von Programmläufen möglich
- Möglichkeiten für Folgeprojekte
 - Automatische Phasenerkennung im Programmablauf
 - Instrumentierung (Online / Offline)
 - Energieeffiziente Bibliotheken

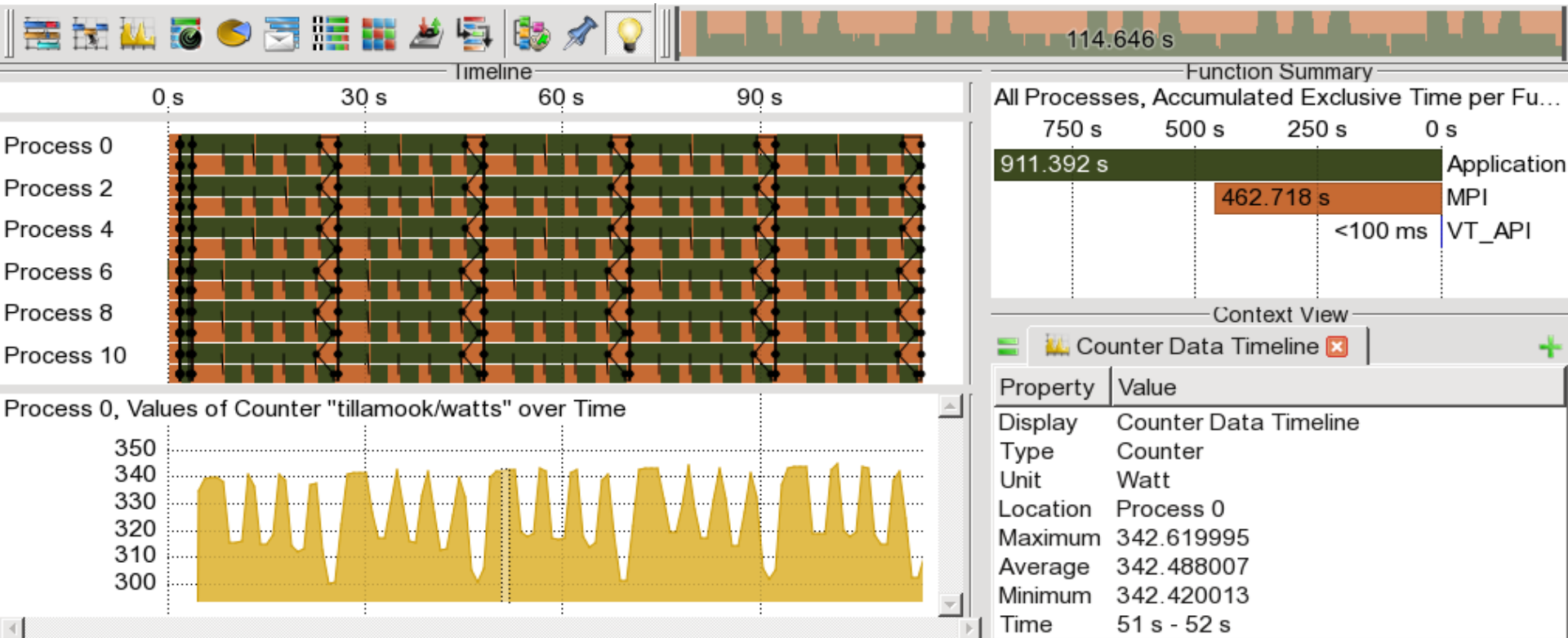
Abschlussworkshop

- Detaillierte Präsentation der Ergebnisse
- Demonstration der entwickelten Werkzeuge
- Direkt vor der “Energy-Aware High-Performance Computing”-Konferenz
 - 12.-14. September 2012 in Hamburg





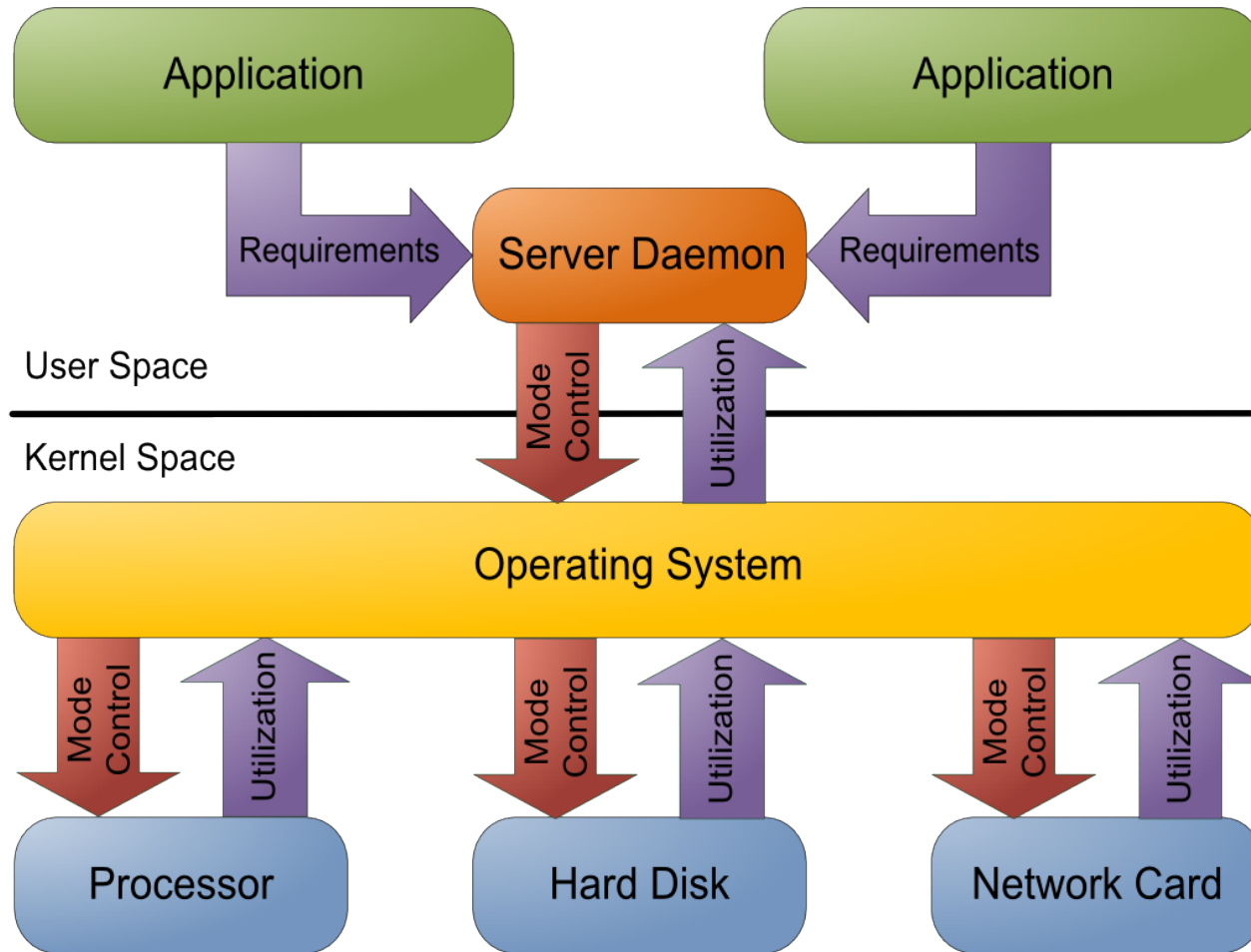
eeMark



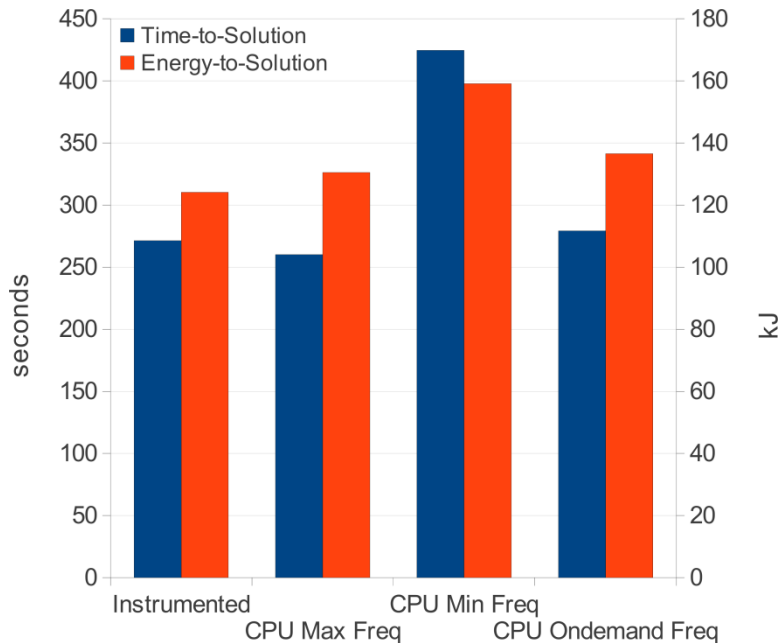
Producer-Consumer-Beispiel: Alle Prozesse haben anteilig Rechen- und Kommunikationszeit

Molka et al.: Flexible Workload Generation for HPC Cluster Efficiency Benchmarking (EnA-HPC'11)

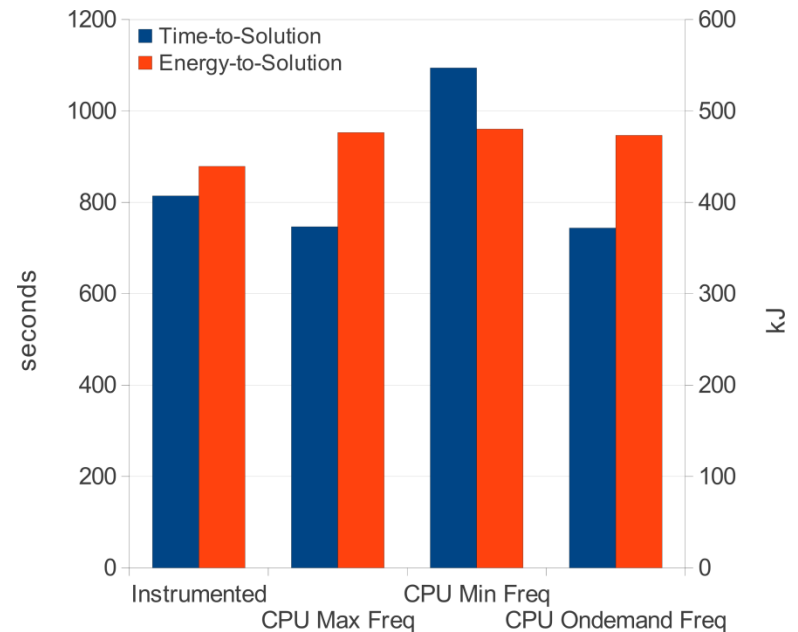
eeDaemon



Ergebnisse



- Intel Xeon-Knoten
- 5 % Energieersparnis (Energy-to-Solution)
- Leistungsverlust (Time-to-Solution) 4 %



- AMD Opteron-Knoten
- 8 % Energieersparnis
- Leistungsverlust 9 %

Minartz et al.: Managing Hardware Power Saving Modes for High Performance Computing (IGCC'11)