On the Energy Proportionality of Distributed NoSQL Data Stores

Balaji Subramaniam and Wu-chun Feng
Department of Computer Science
Virginia Tech

VirginiaTech
Invent the Future

SyNeRG
synergy.cs.vt.edu

The Big Data Deluge

- What is Big Data?
  - High volume, velocity and variety of information assets that demand cost-effective, innovative forms of information storage and processing for enhanced insight and decision making.
- Characteristics of Big Data
  - Volume
  - Velocity
  - Variety

Big Data in Scientific Computing

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data - **volume**

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data - **volume**
- CERN – Large Hadron Collider
  - 150 million sensors delivering data 40 million times a second (approximately 700 MB/s)

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data - **volume**
- CERN – Large Hadron Collider
  - 150 million sensors delivering data 40 million times a second (approximately 700 MB/s) **- velocity**

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data - **volume**
- CERN – Large Hadron Collider
  - 150 million sensors delivering data 40 million times a second (approximately 700 MB/s) **- velocity**
- Next Generation Sequencing
  - Sequencing instruments have considerable heterogeneity in terms of cost, speed, throughput, read lengths and error rates

Big Data in Scientific Computing

- NASA Center for Climate Simulation
  - Stores 32 petabytes of climate and weather data - **volume**
- CERN – Large Hadron Collider
  - 150 million sensors delivering data 40 million times a second (approximately 700 MB/s) **- velocity**
- Next Generation Sequencing
  - Sequencing instruments have considerable heterogeneity in terms of cost, speed, throughput, read lengths and error rates **- variety**

Big Data in Scientific Computing Meets NoSQL

Balaji Subramaniam (balaji@cs.vt.edu)

Big Data in Scientific Computing Meets NoSQL

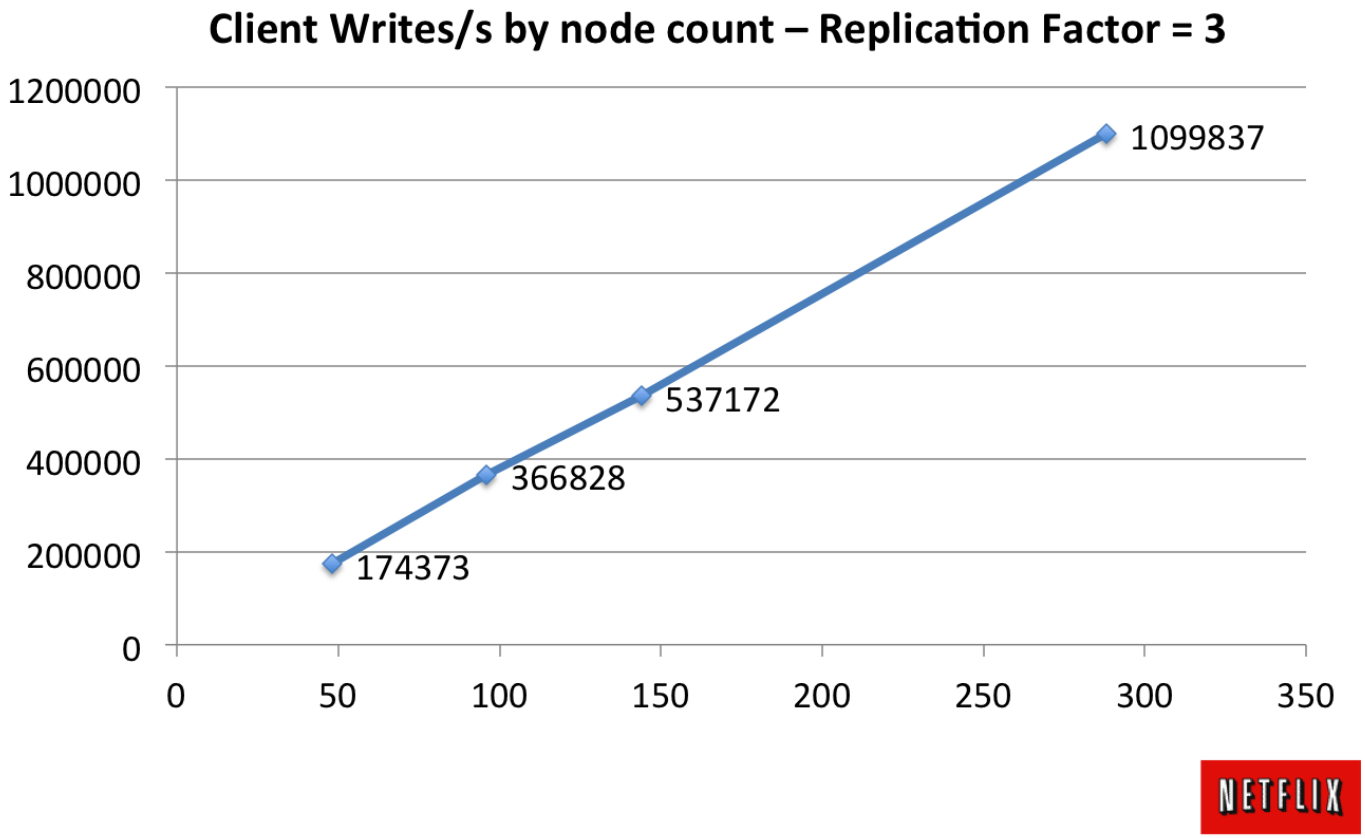
- NASA and CERN use Apache Cassandra (a NoSQL data store) to handle the volume, velocity and the variety of big data

Big Data in Scientific Computing Meets NoSQL

- Why NoSQL?

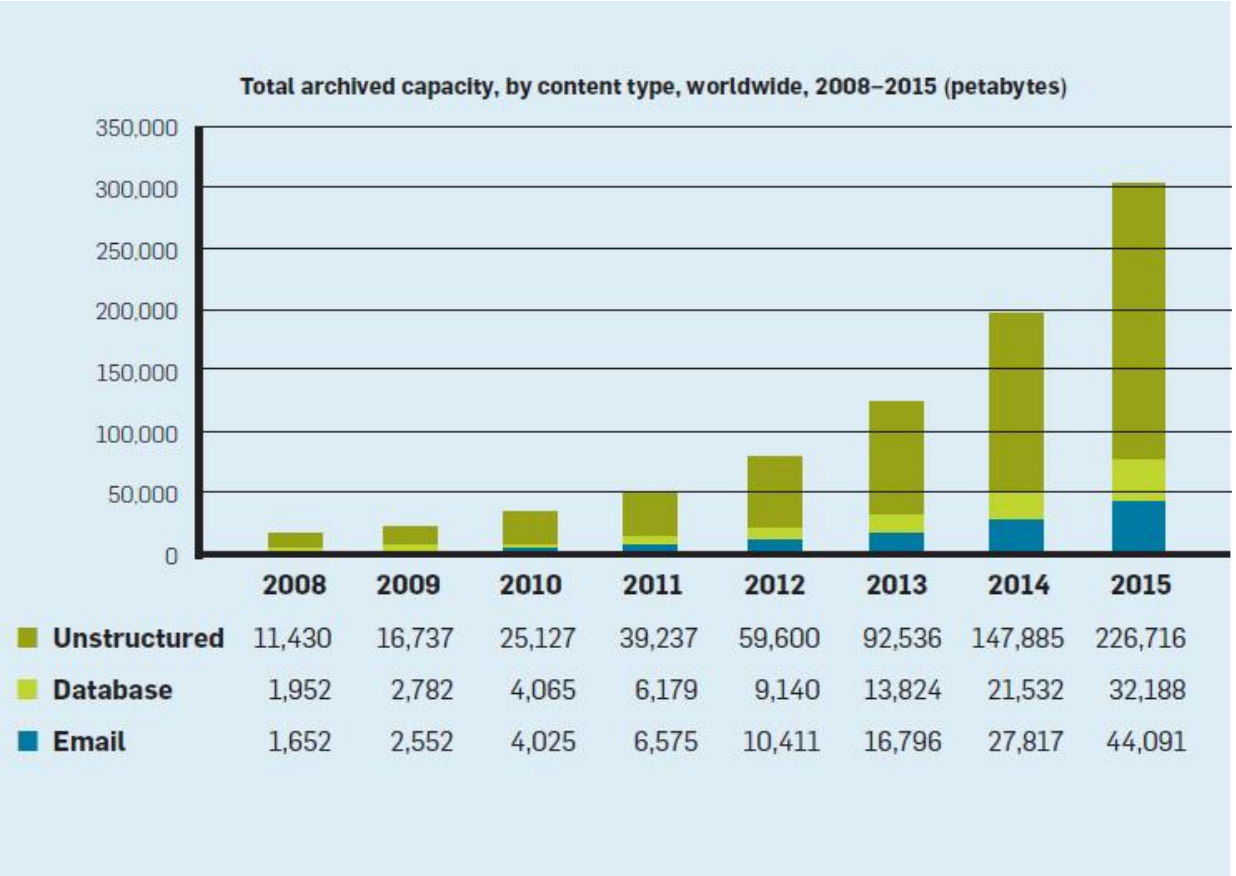
Big Data in Scientific Computing Meets NoSQL

- Why NoSQL?
  - Scalability



Big Data in Scientific Computing Meets NoSQL

- Why NoSQL?
  - Growth in unstructured data



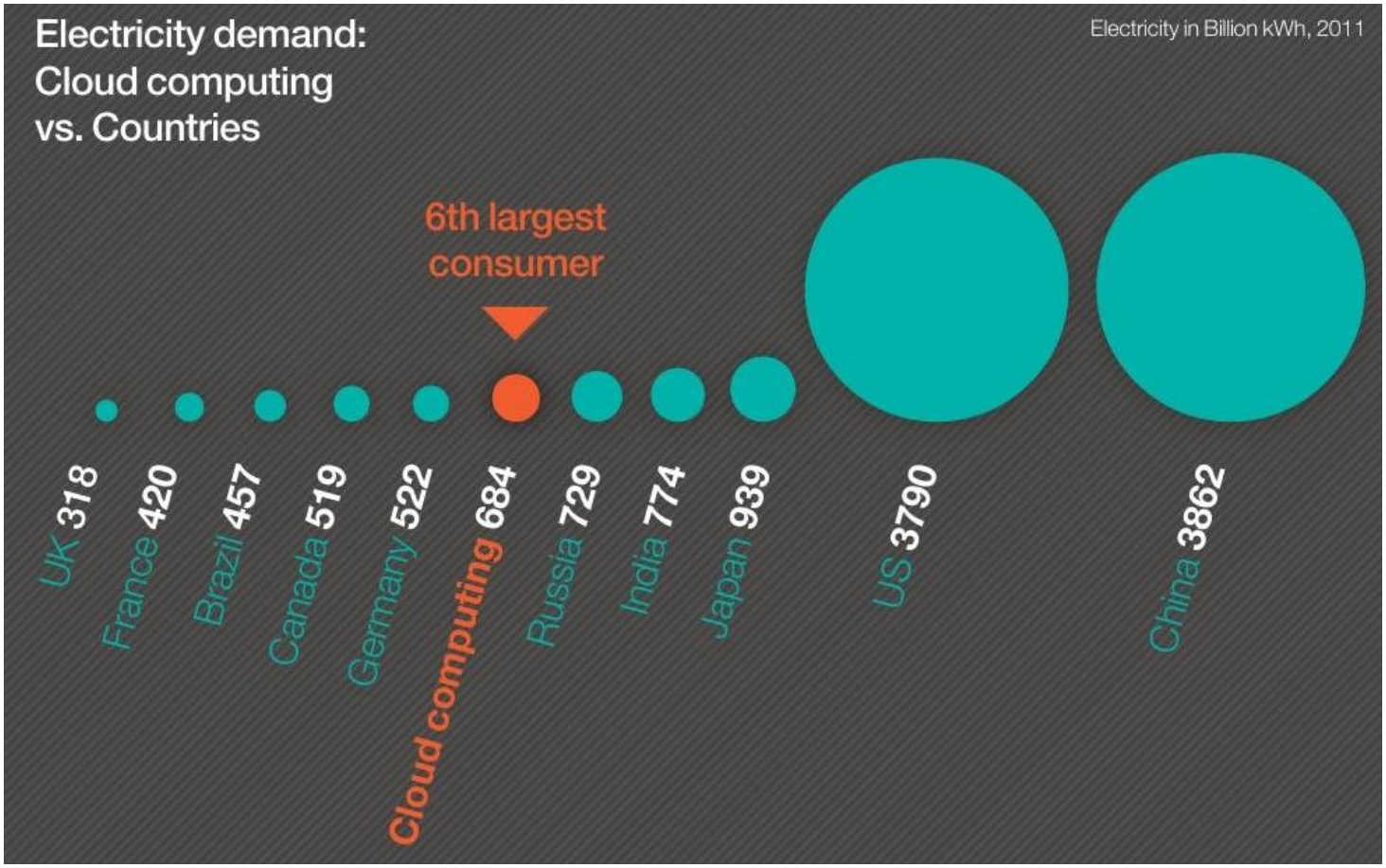
	2008	2009	2010	2011	2012	2013	2014	2015
Unstructured	11,430	16,737	25,127	39,237	59,600	92,536	147,885	226,716
Database	1,952	2,782	4,065	6,179	9,140	13,824	21,532	32,188
Email	1,652	2,552	4,025	6,575	10,411	16,796	27,817	44,091

Source: http://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext

Big Data in Scientific Computing Meets NoSQL

- But storing, retrieving and managing such volumes of data requires massive computing infrastructure
  - Apple's Cassandra deployment uses over 75000 nodes to store 10PB of data

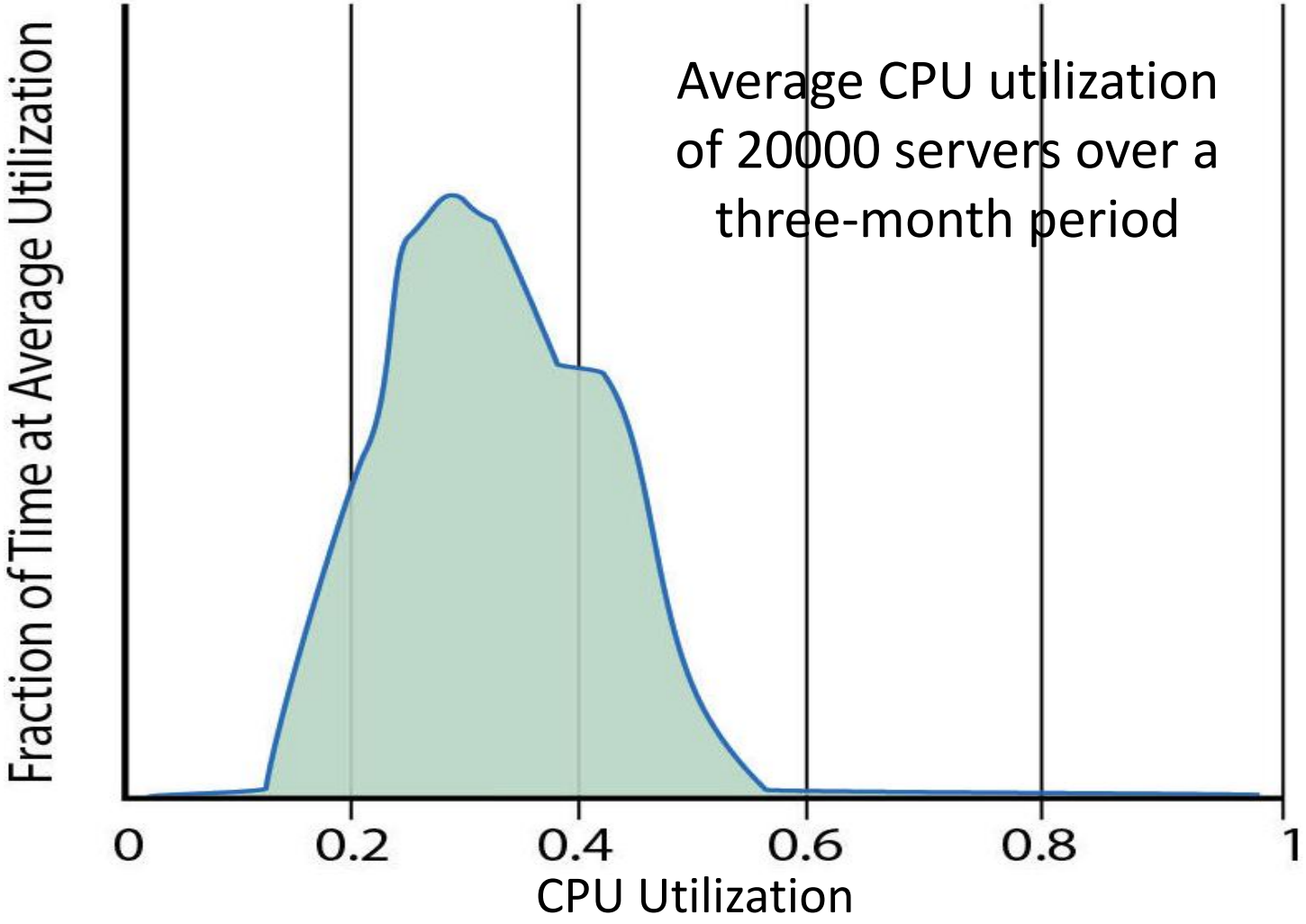
Power: First-Order Design Constraint in Data Centers



"Data centers will be the fastest growing part of the global IT sector energy footprint as our online world rapidly expands; their energy demand will increase 81% by 2020."

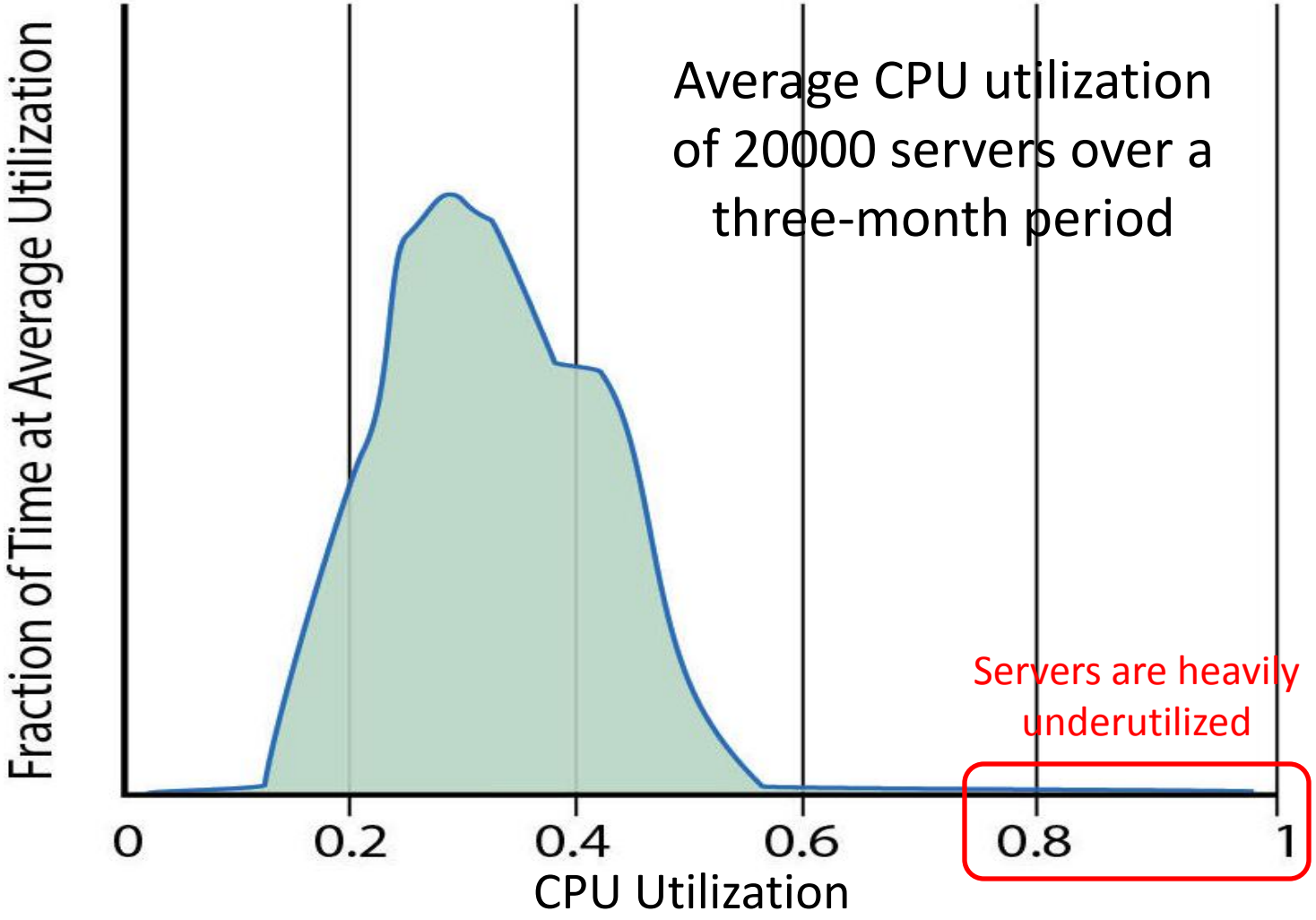
Source: Clicking Clean: How Companies are Creating the Green Internet, GreenPeace, April, 2014

Severe Underutilization in Data Centers



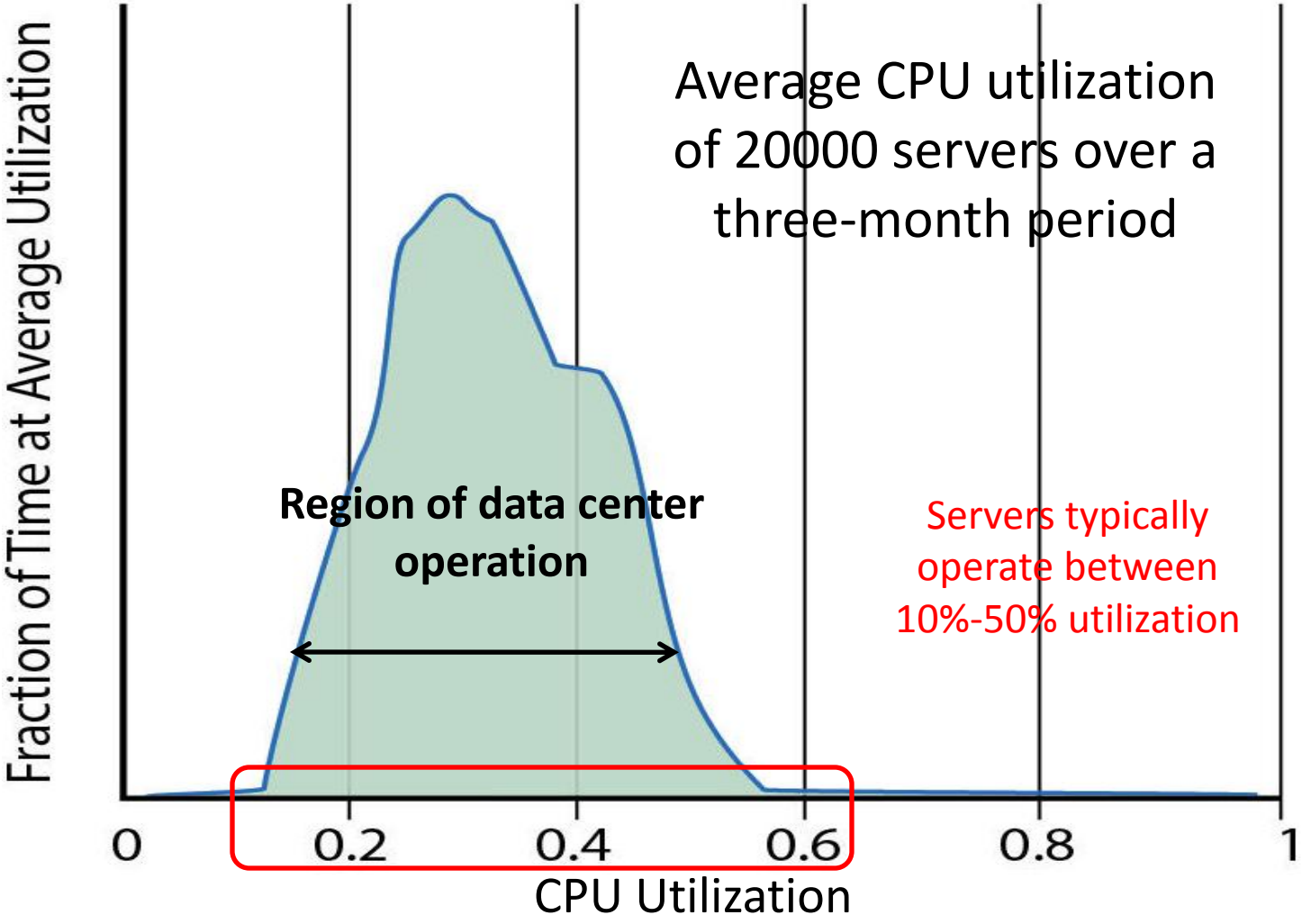
Source: L. A. Barroso et al., The Datacenter as a Computer, 2013

Severe Underutilization in Data Centers

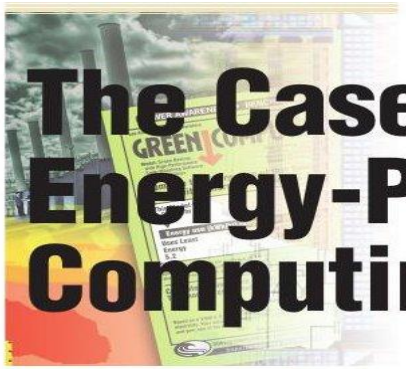


Source: L. A. Barroso et al., The Datacenter as a Computer, 2013

Severe Underutilization in Data Centers



Source: L. A. Barroso et al., The Datacenter as a Computer, 2013

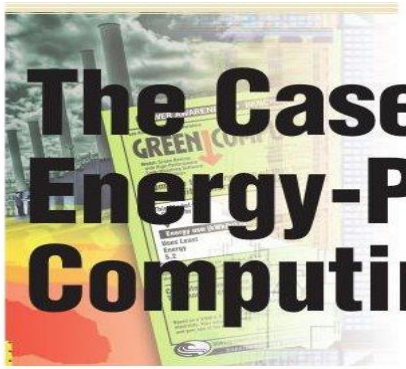


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

- Consume power proportional to utilization

Source: L. A. Barroso et al., The Case for Energy-Proportional Computing, IEEE Computer, 2007

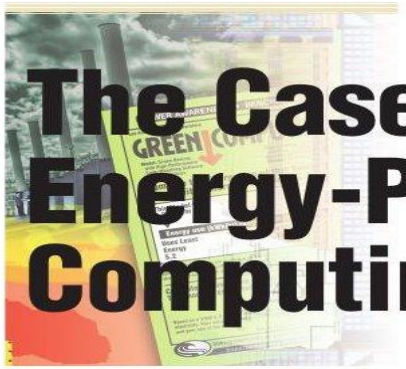


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

- Consume power proportional to utilization
- Advocate improvements in non-peak power efficiency

Source: L. A. Barroso et al., The Case for Energy-Proportional Computing, IEEE Computer, 2007

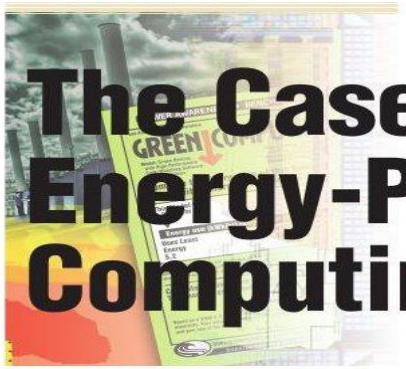


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

- Consume power proportional to utilization
- Advocate improvements in non-peak power efficiency
- Enhance data center energy efficiency via non-peak power efficiency improvements

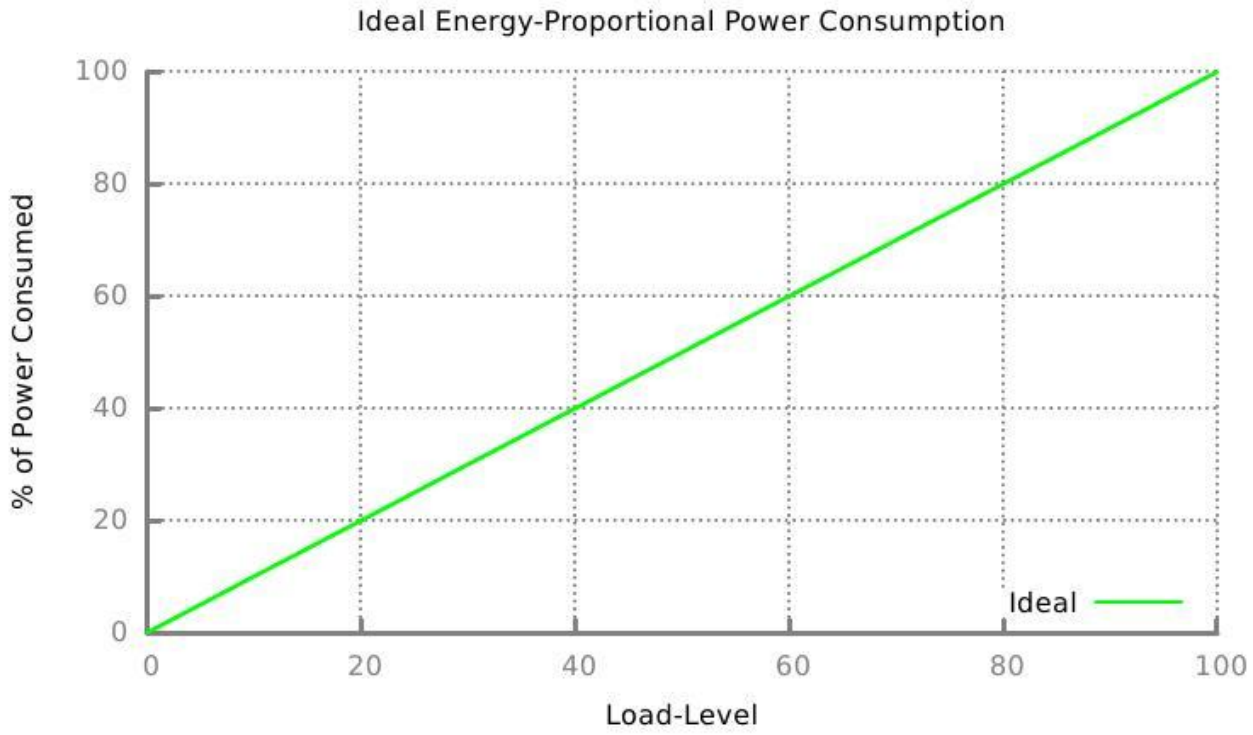
Source: L. A. Barroso et al., The Case for Energy-Proportional Computing, IEEE Computer, 2007



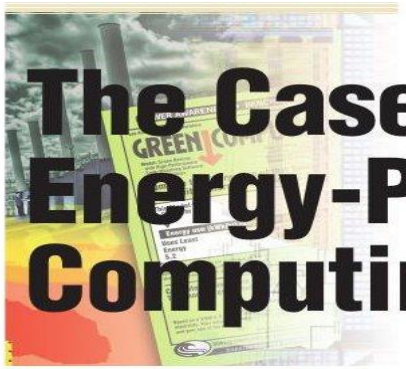
The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)



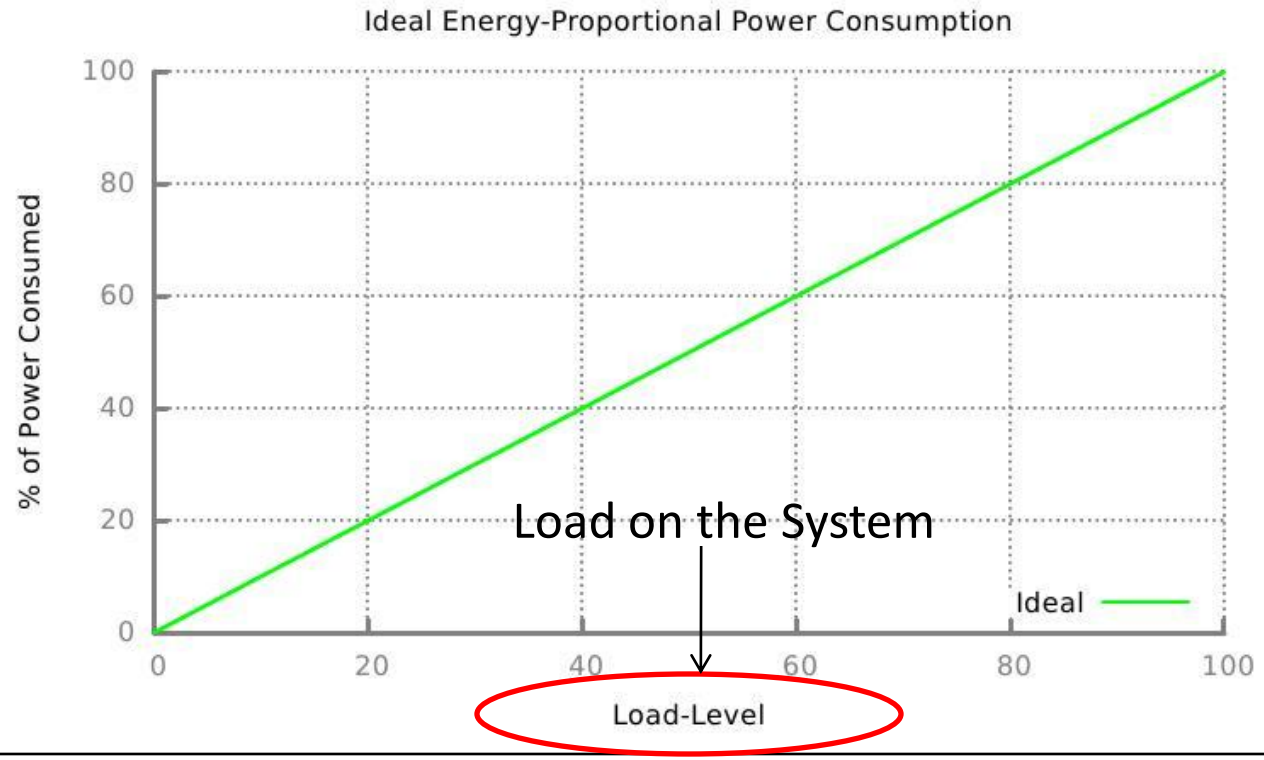
Balaji Subramaniam (balaji@cs.vt.edu)

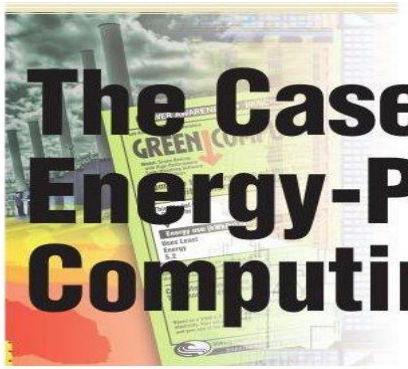


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

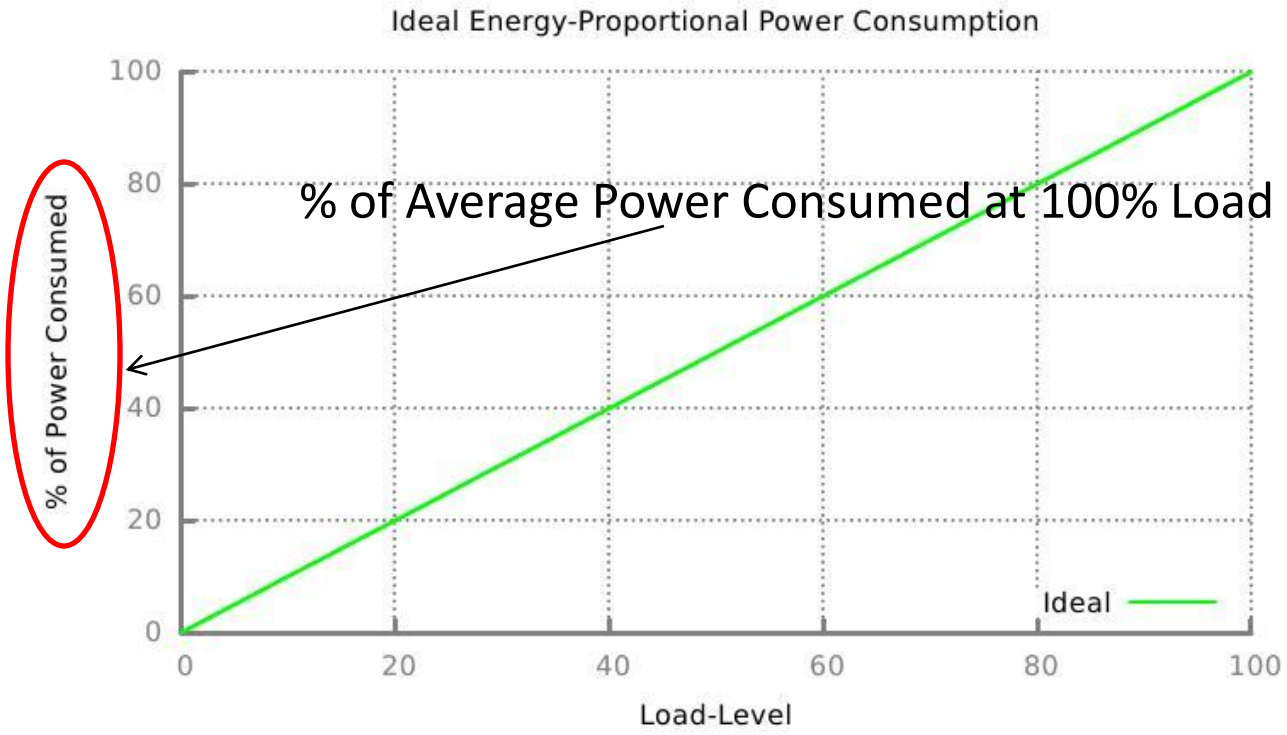


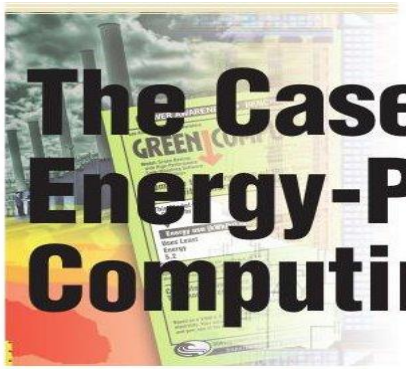


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

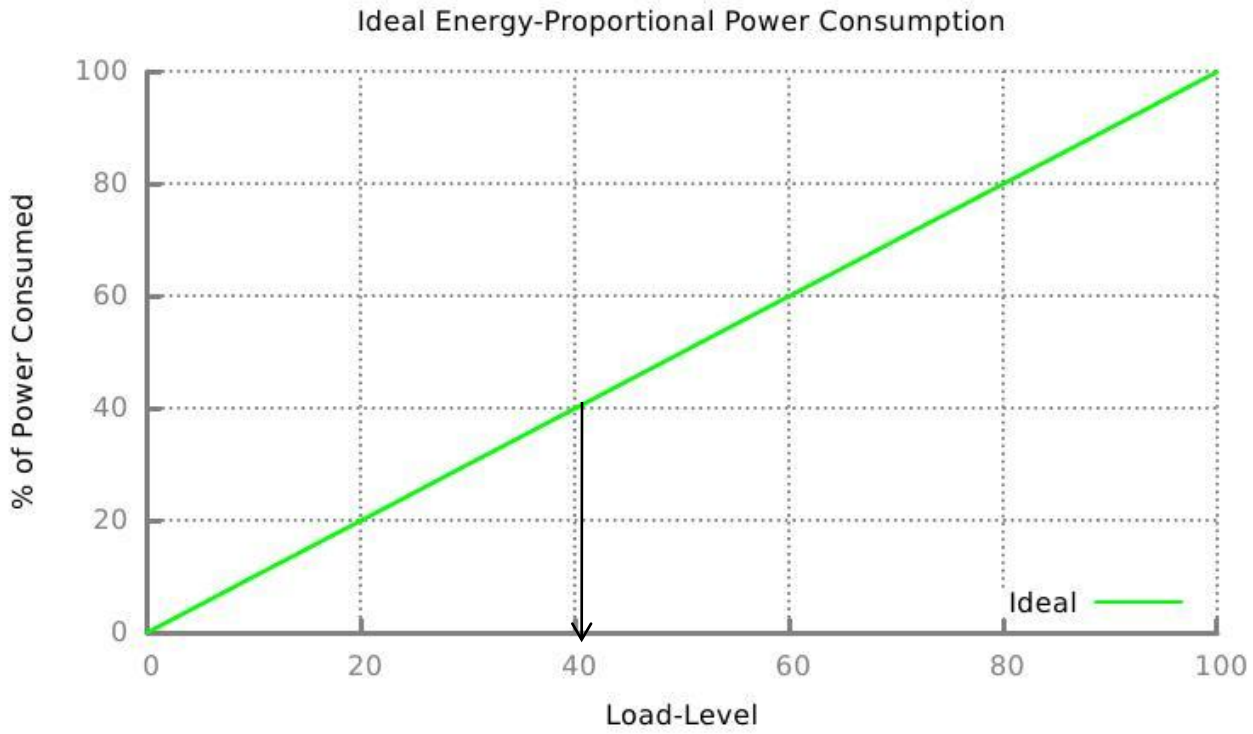


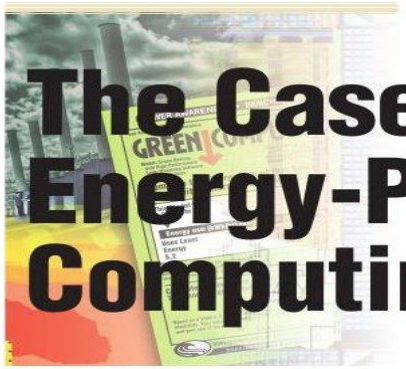


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

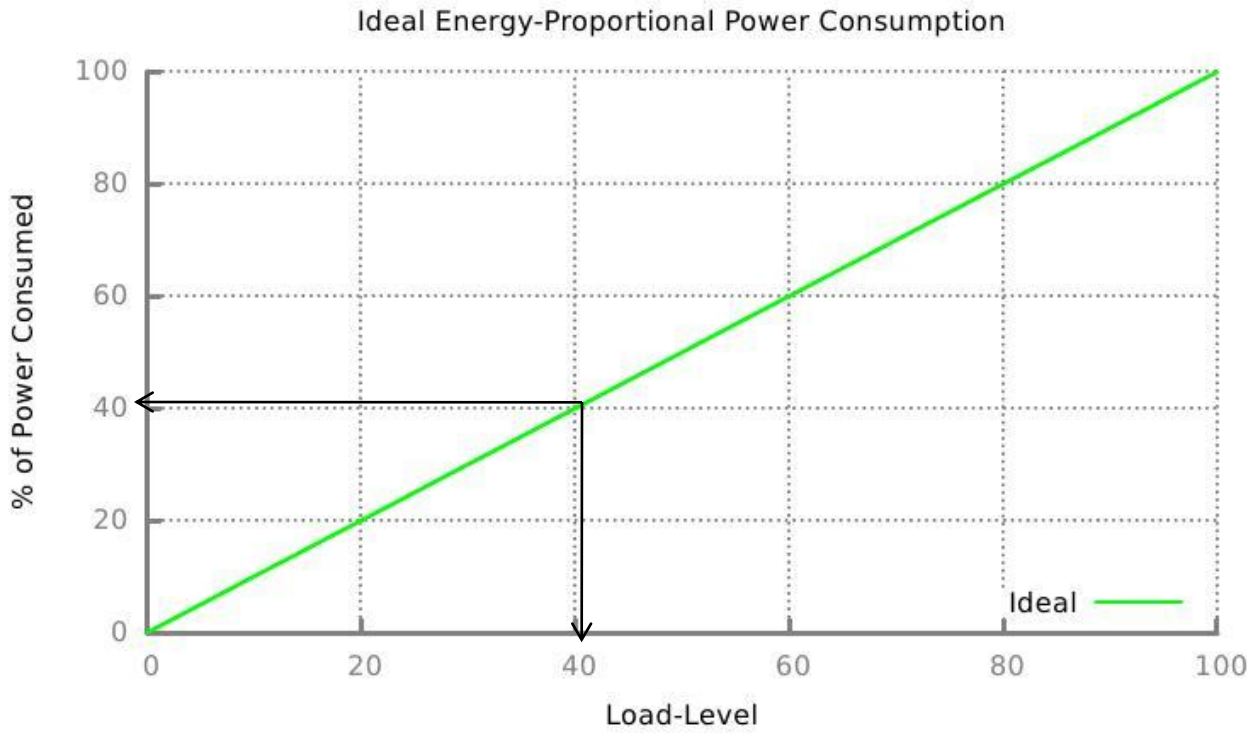


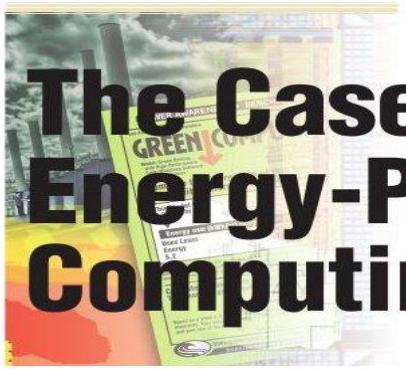


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

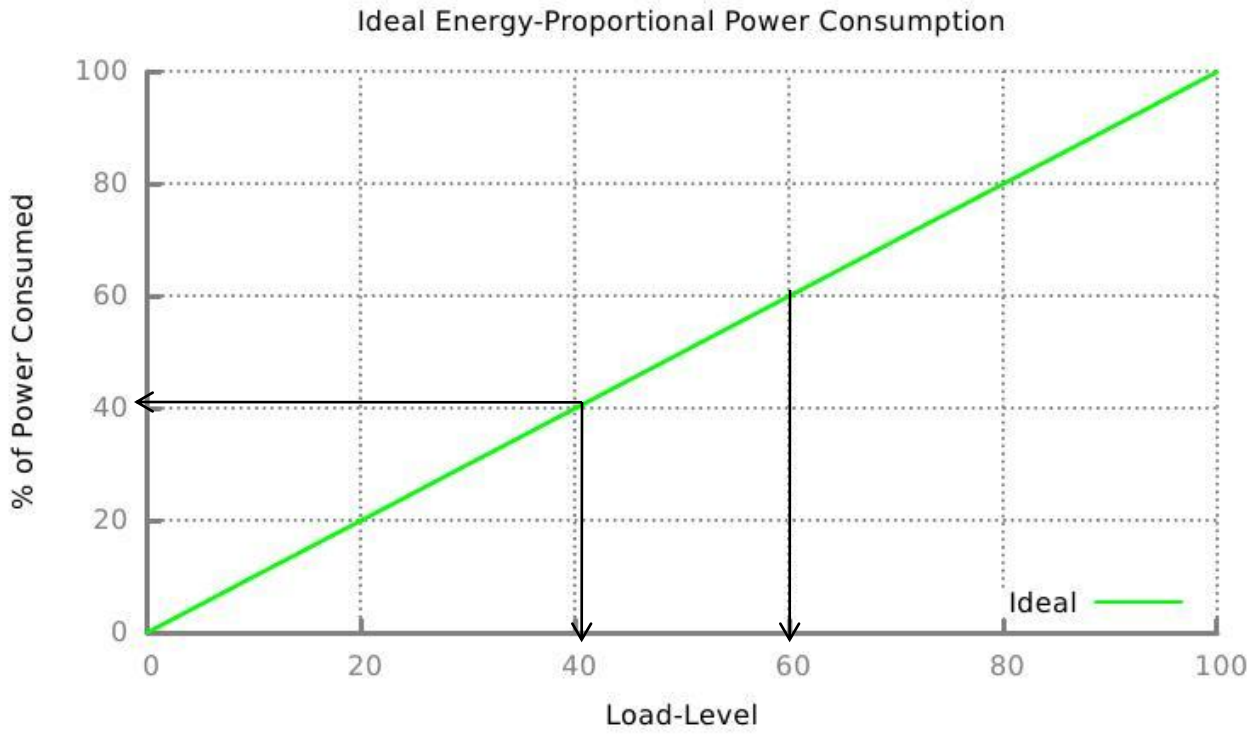


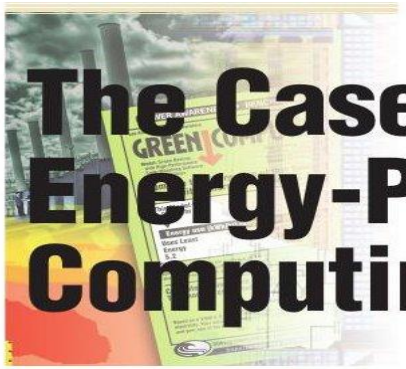


The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

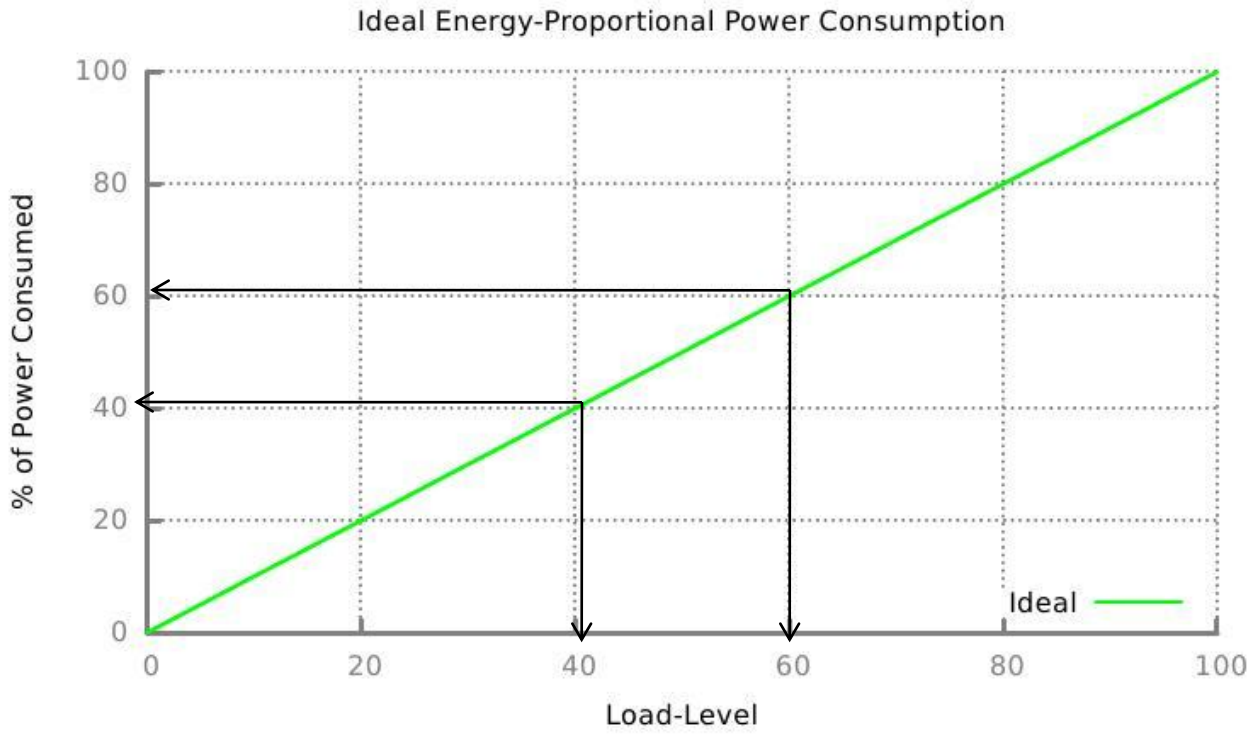




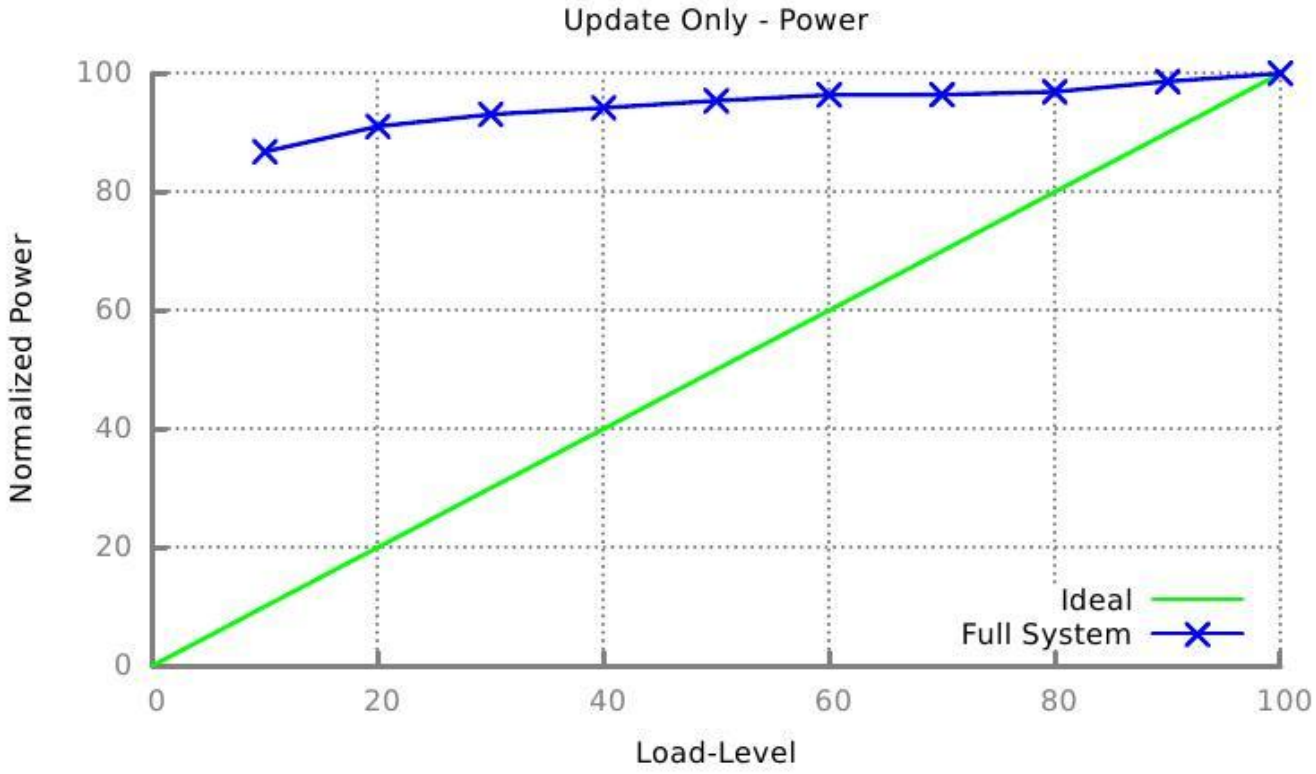
The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Consume power proportional to load on the system (ideal case)

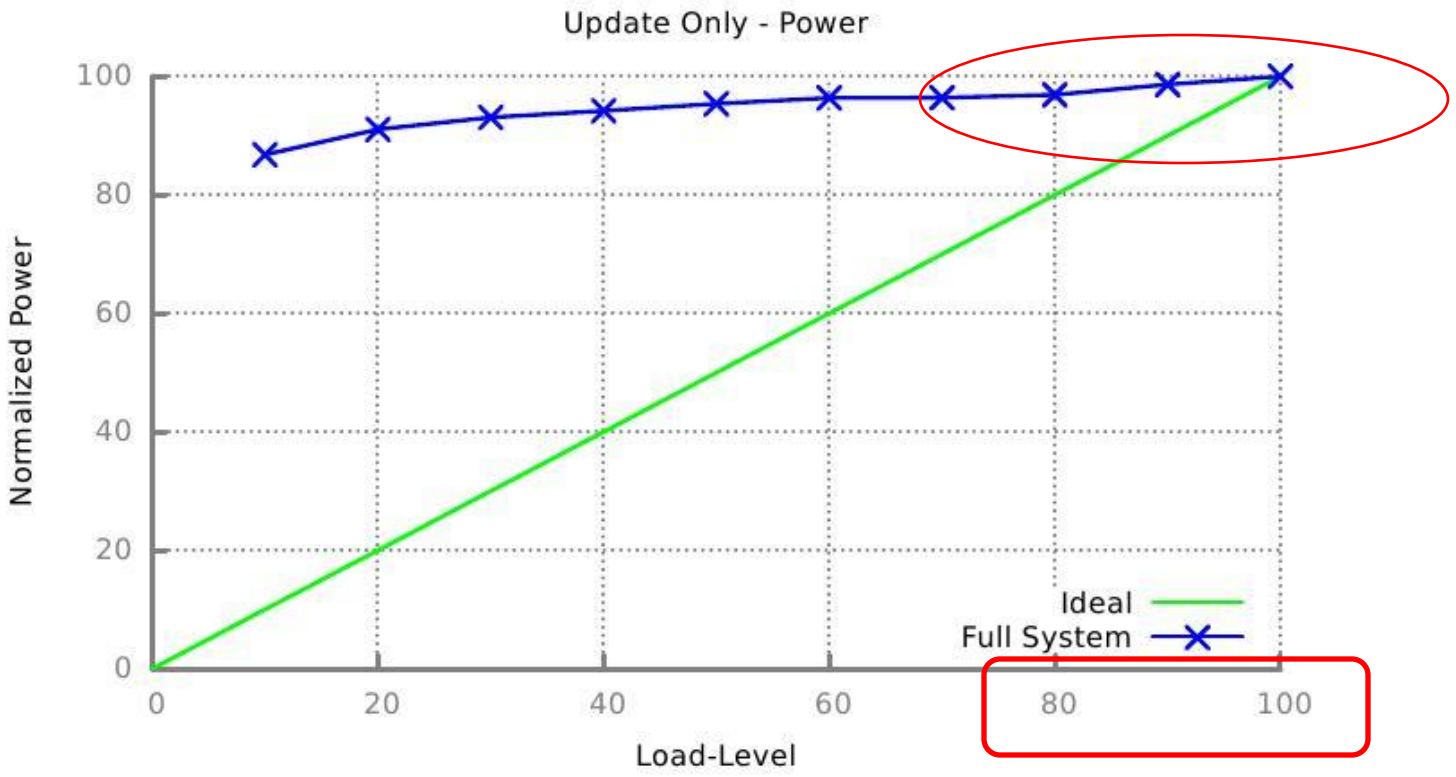


Power Efficiency at Different Levels of Utilization



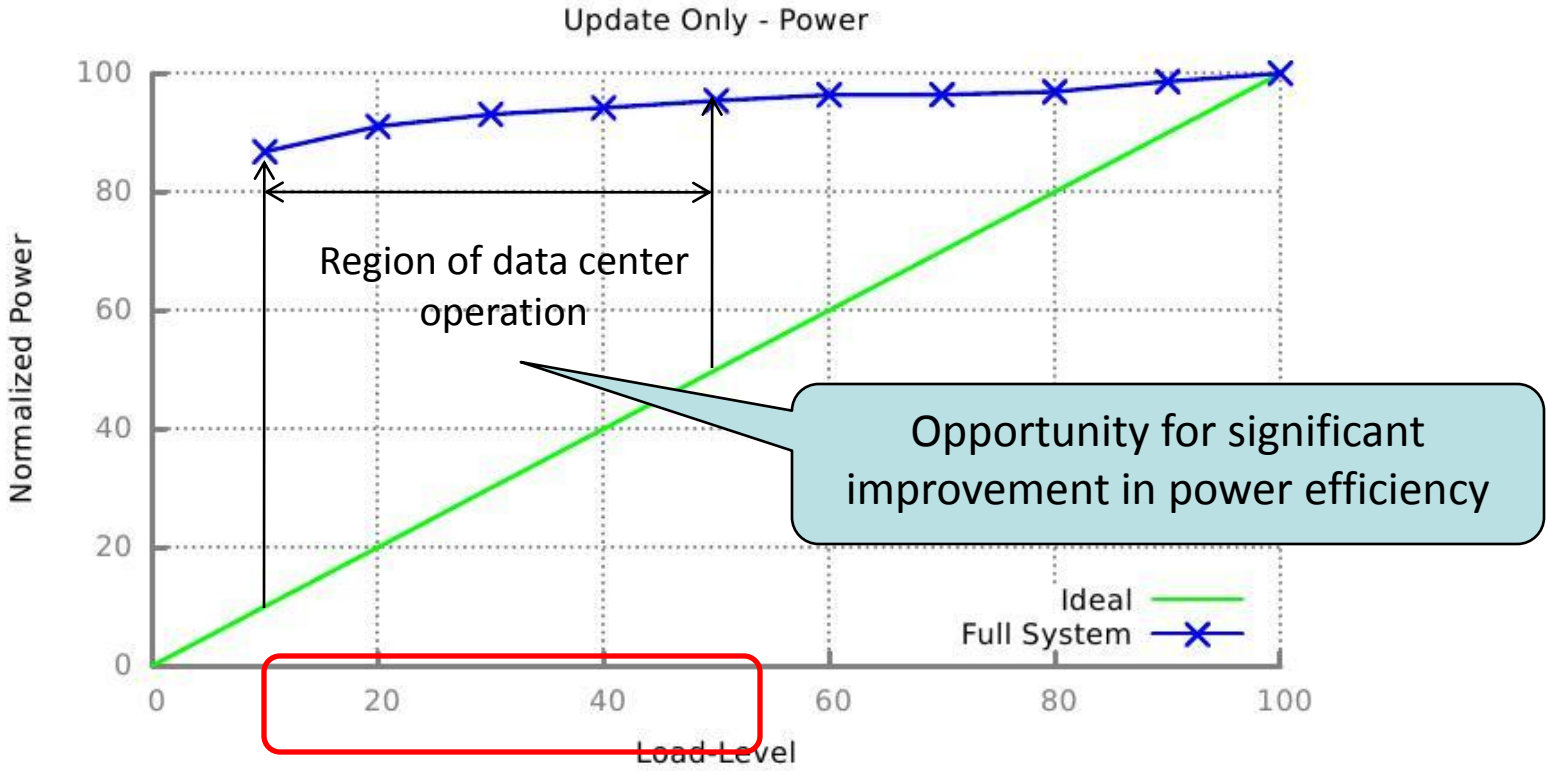
- Apache Cassandra full system power consumption at different load-levels

Power Efficiency at Different Levels of Utilization



- Near energy-proportional power consumption at high load-levels

Power Efficiency at Different Levels of Utilization



- Near energy-proportional power consumption at high load-levels
- Significant gap between ideal and energy-proportional power consumption in the region of typical data center operation

Improving Energy Proportionality

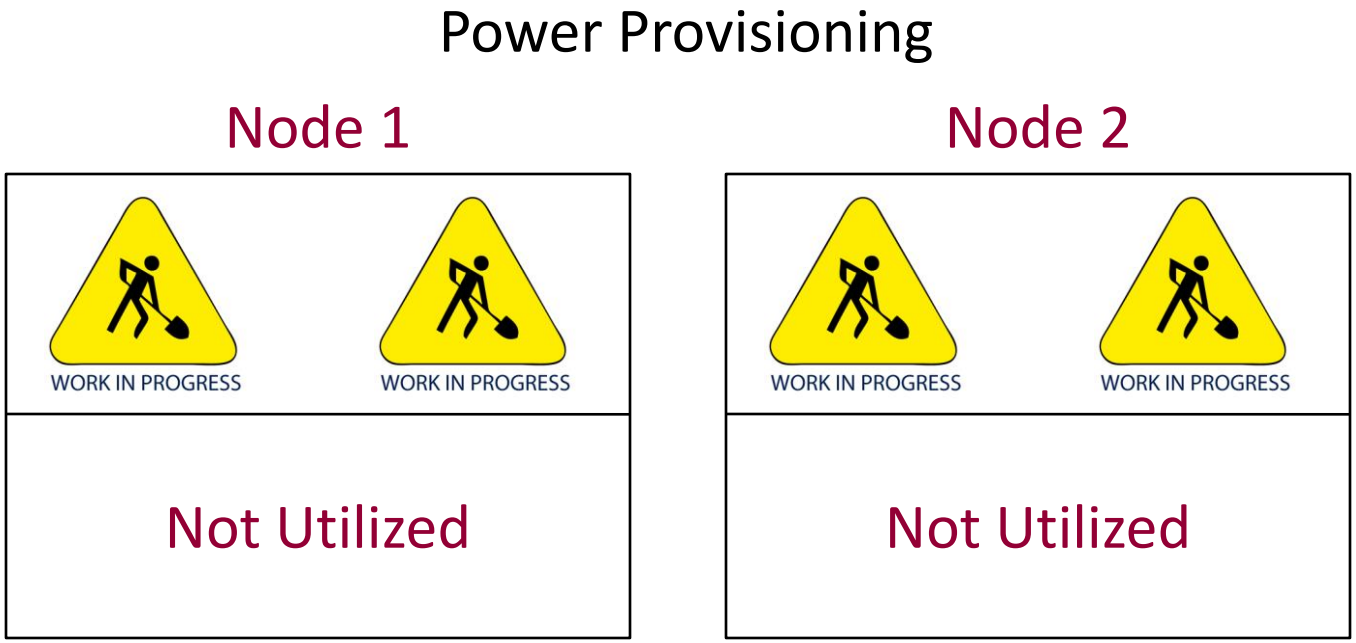
- Techniques to improve energy proportionality
  - Two schools of thought exists

Improving Energy Proportionality

- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power

Improving Energy Proportionality

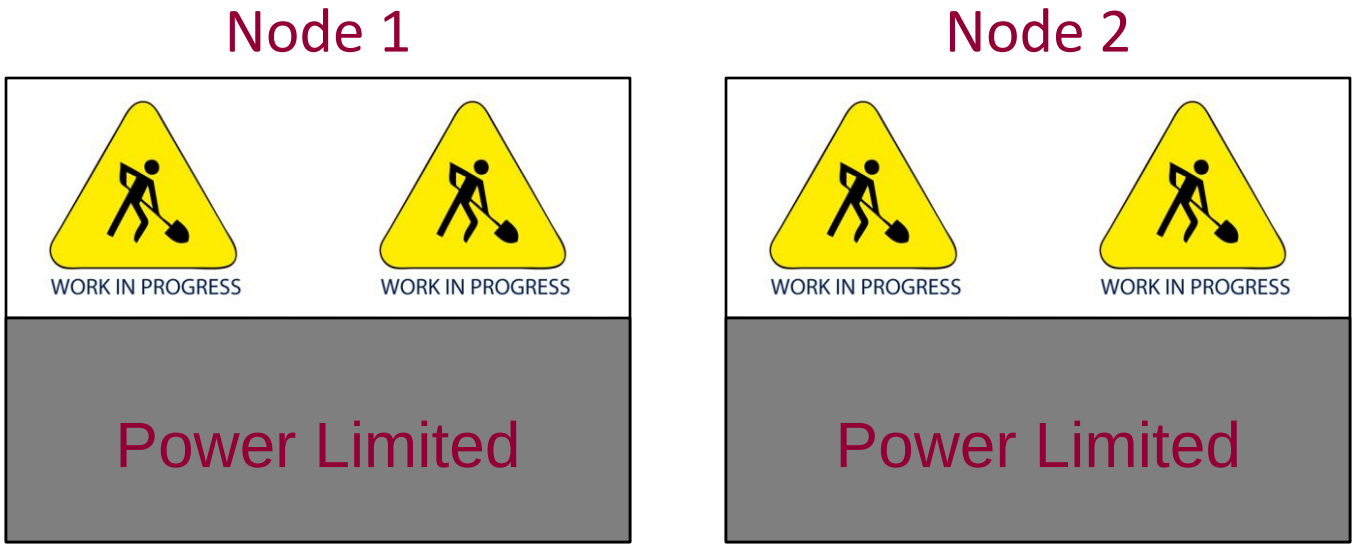
- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power



Improving Energy Proportionality

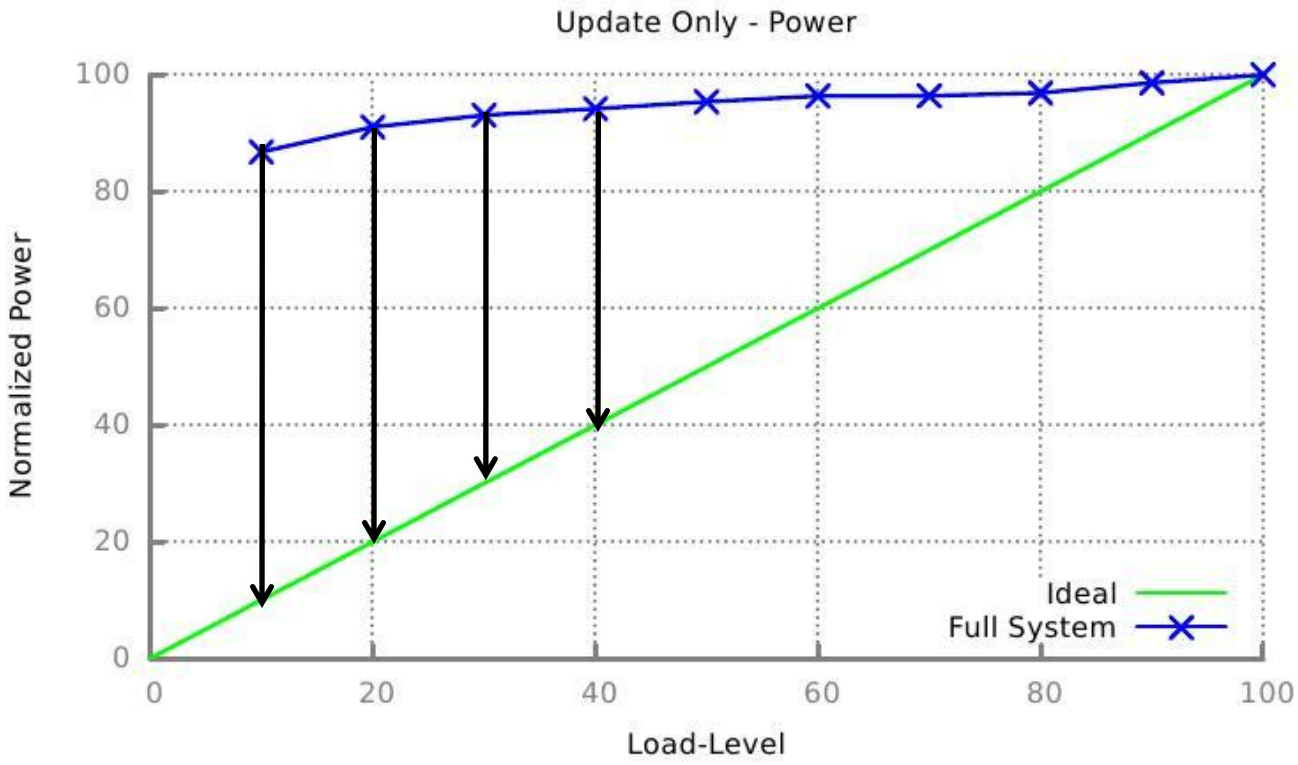
- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power

Power Provisioning



Improving Energy Proportionality

- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power



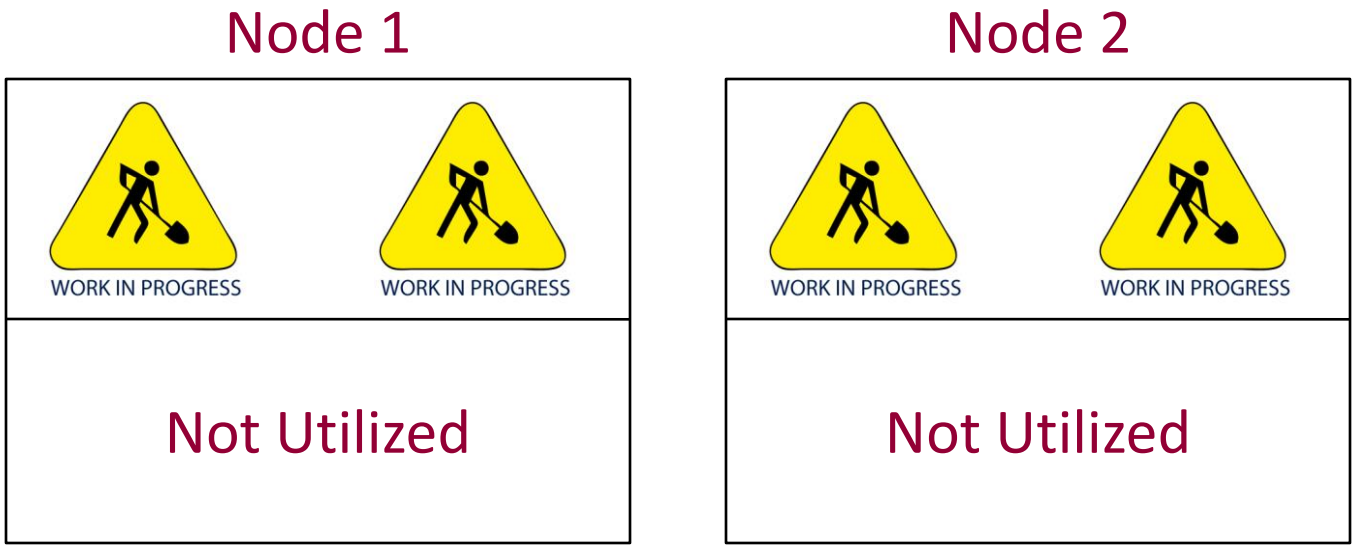
Improving Energy Proportionality

- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power
    - Improve energy proportionality by provisioning resource

Improving Energy Proportionality

- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power
    - Improve energy proportionality by provisioning resource

Resource Provisioning

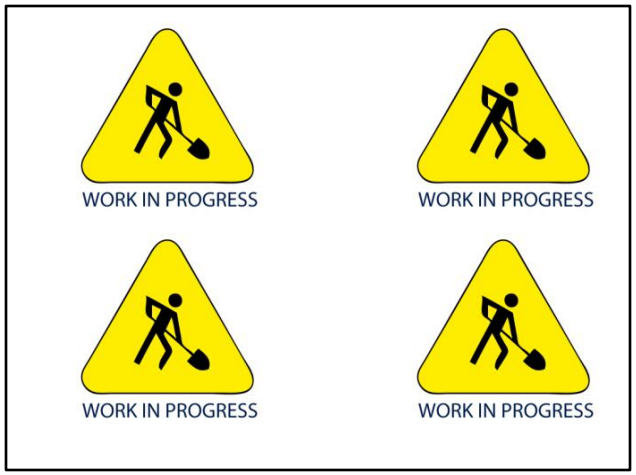


Improving Energy Proportionality

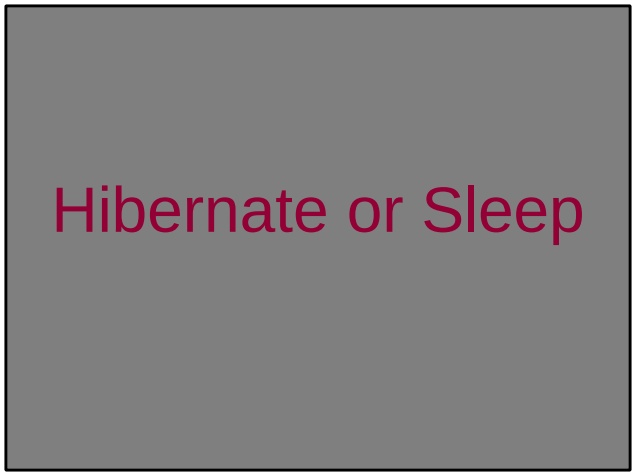
- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power
    - Improve energy proportionality by provisioning resource

Resource Provisioning

Node 1



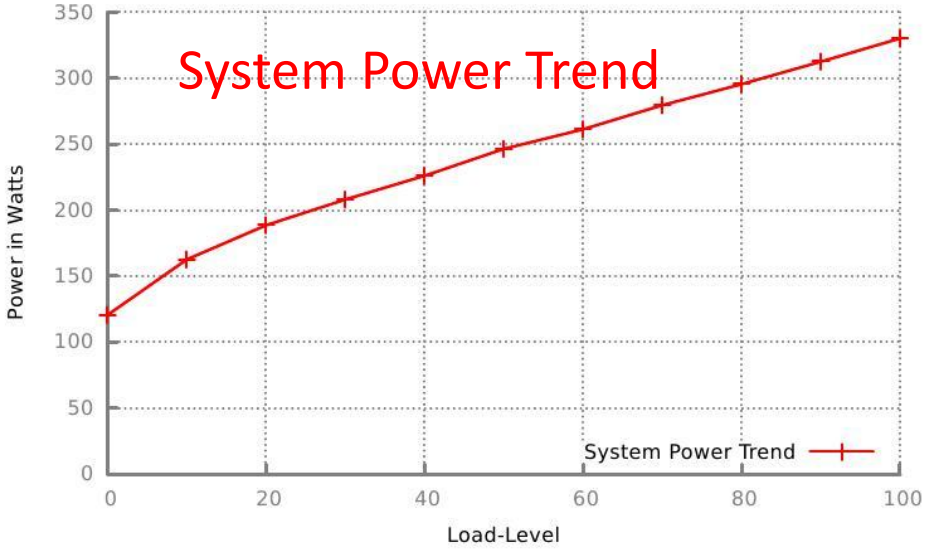
Node 2



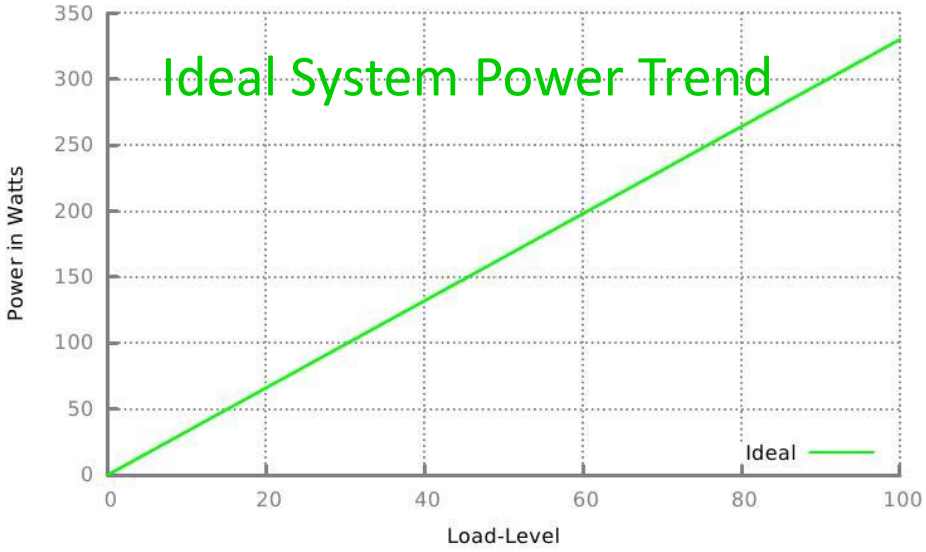
Improving Energy Proportionality

- Techniques to improve energy proportionality
  - Two schools of thought exists
    - Improve energy proportionality by provisioning power
    - Improve energy proportionality by provisioning resource

Our Vision

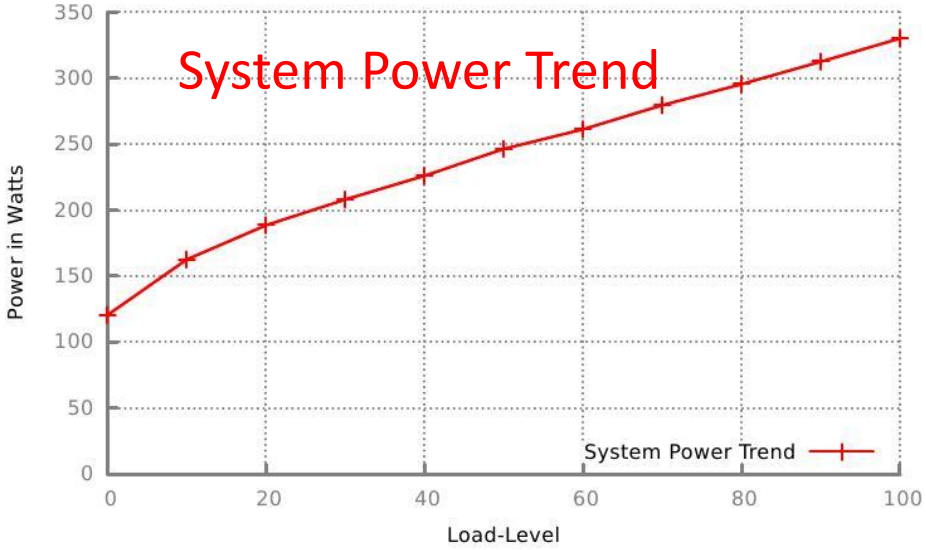


Power and/or
Resource Provisioning

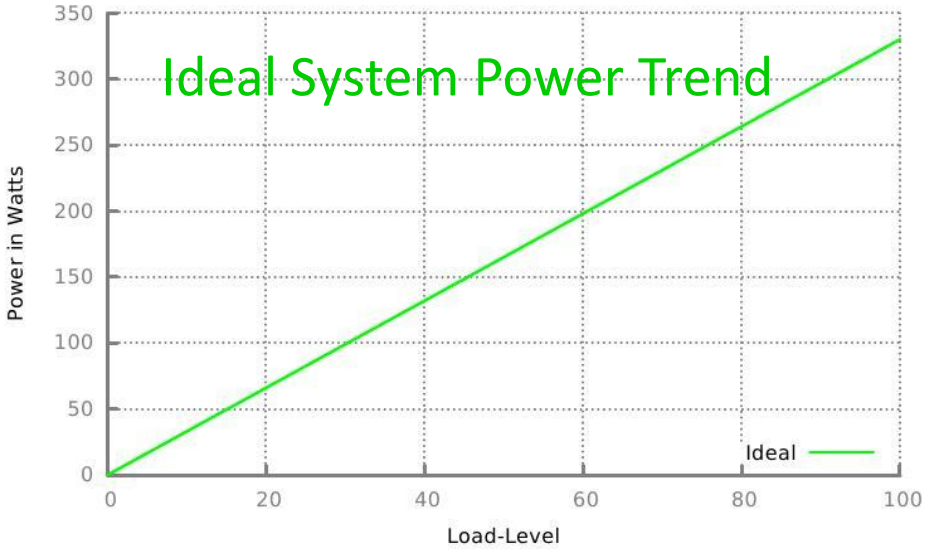


Balaji Subramaniam (balaji@cs.vt.edu)

Our Vision



Power and/or
Resource Provisioning



Can we improve energy proportionality?

Balaji Subramaniam (balaji@cs.vt.edu)

Evaluation Setup

- Data-Server: Apache Cassandra
  - Distributed key-value store

Evaluation Setup

- Data-Server: Apache Cassandra
  - Distributed key-value store
- Evaluation setup
  - Yahoo Cloud Serving Benchmark (YCSB)
  - Workload: Read-only and update-only
  - 10 million records
  - Replication factor of 3

Evaluation Setup

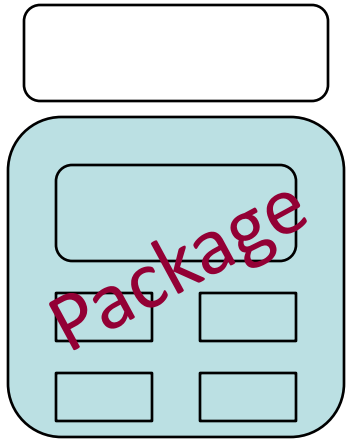
- Data-Server: Apache Cassandra
  - Distributed key-value store
- Evaluation setup
  - Yahoo Cloud Serving Benchmark (YCSB)
  - Workload: Read-only and update-only
  - 10 million records
  - Replication factor of 3
- Evaluation platform
  - Four Intel Xeon E5-2620 nodes
  - 16 GB of memory per node
  - 256 GB hard disk per node

Power Management Interface:
Intel's Running Average Power Limit (RAPL)

- Introduced in Intel Sandy Bridge processors
- Capabilities exposed
  - Energy metering
  - Power limiting
- Subsystems over which we have control via RAPL

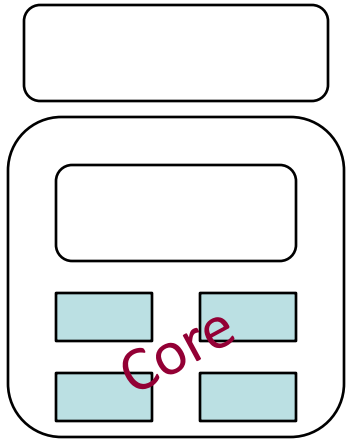
Power Management Interface:
Intel's Running Average Power Limit (RAPL)

- Introduced in Intel Sandy Bridge processors
- Capabilities exposed
  - Energy metering
  - Power limiting
- Subsystems over which we have control via RAPL
  - Package (processor package) subsystem



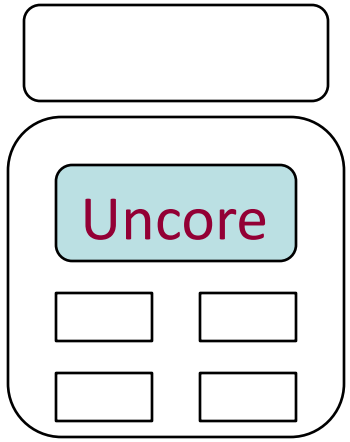
Power Management Interface:
Intel's Running Average Power Limit (RAPL)

- Introduced in Intel Sandy Bridge processors
- Capabilities exposed
  - Energy metering
  - Power limiting
- Subsystems over which we have control via RAPL
  - Package (processor package) subsystem
  - PP0 (processor *core*) subsystem



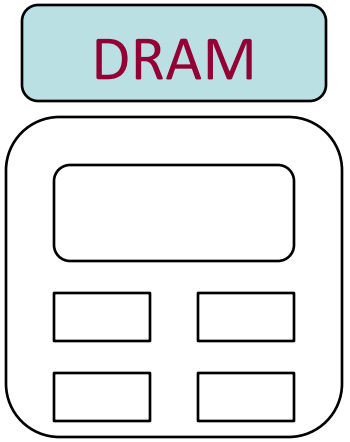
Power Management Interface:
Intel's Running Average Power Limit (RAPL)

- Introduced in Intel Sandy Bridge processors
- Capabilities exposed
  - Energy metering
  - Power limiting
- Subsystems over which we have control via RAPL
  - Package (processor package) subsystem
  - PP0 (processor *core*) subsystem

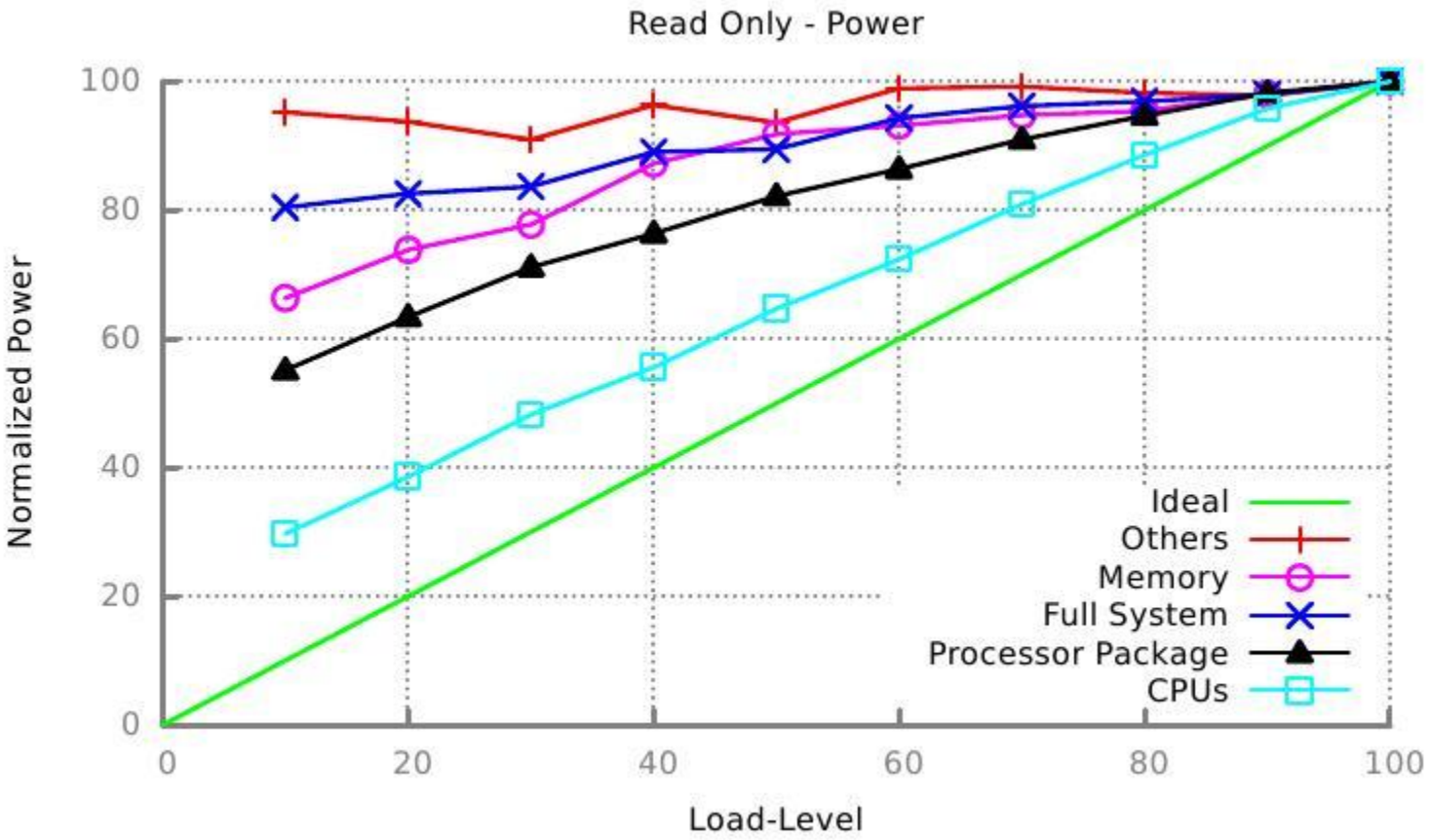


Power Management Interface:
Intel's Running Average Power Limit (RAPL)

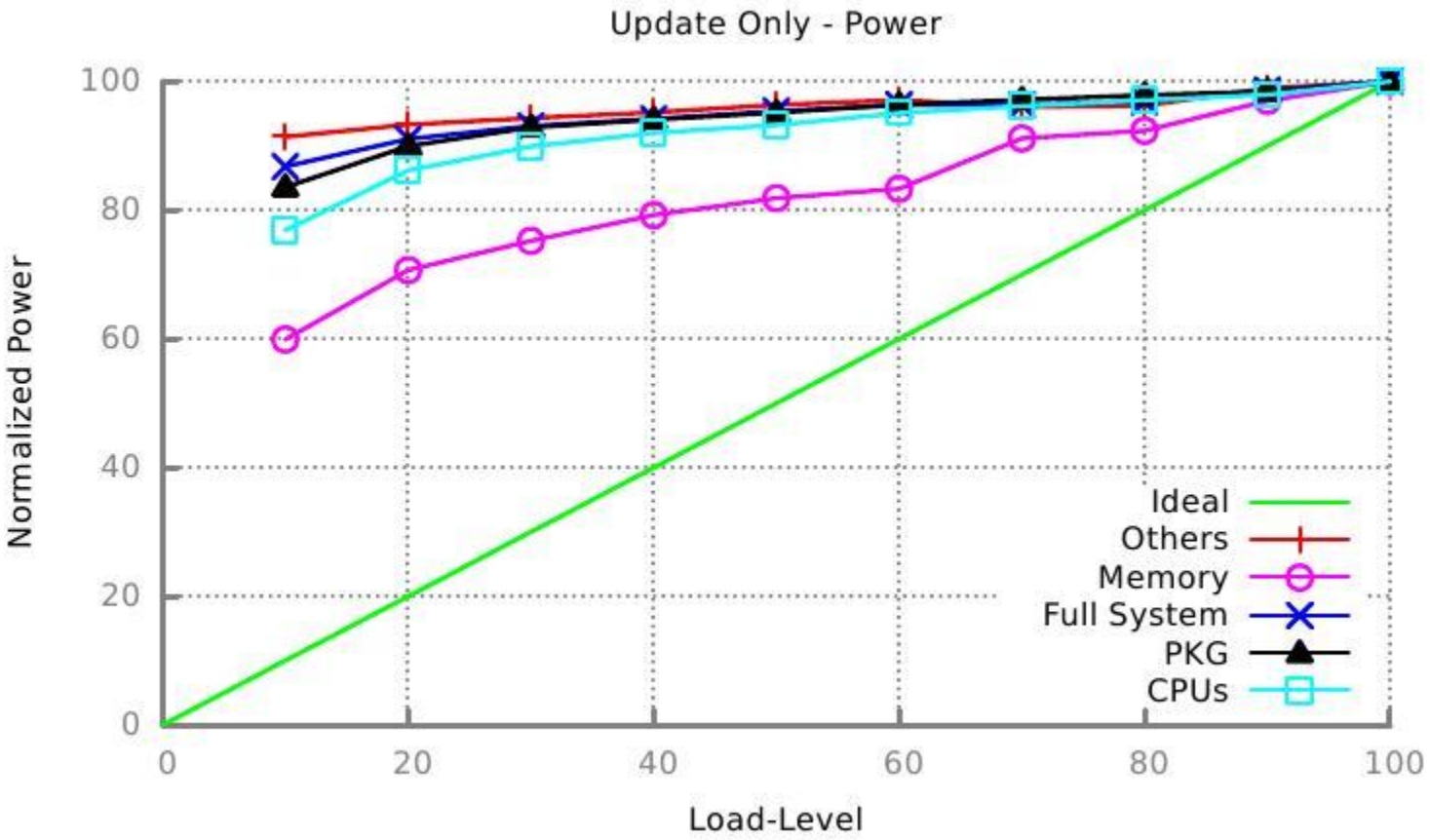
- Introduced in Intel Sandy Bridge processors
- Capabilities exposed
  - Energy metering
  - Power limiting
- Subsystems over which we have control via RAPL
  - Package (processor package) subsystem
  - PP0 (processor *core*) subsystem
  - DRAM (memory) subsystem



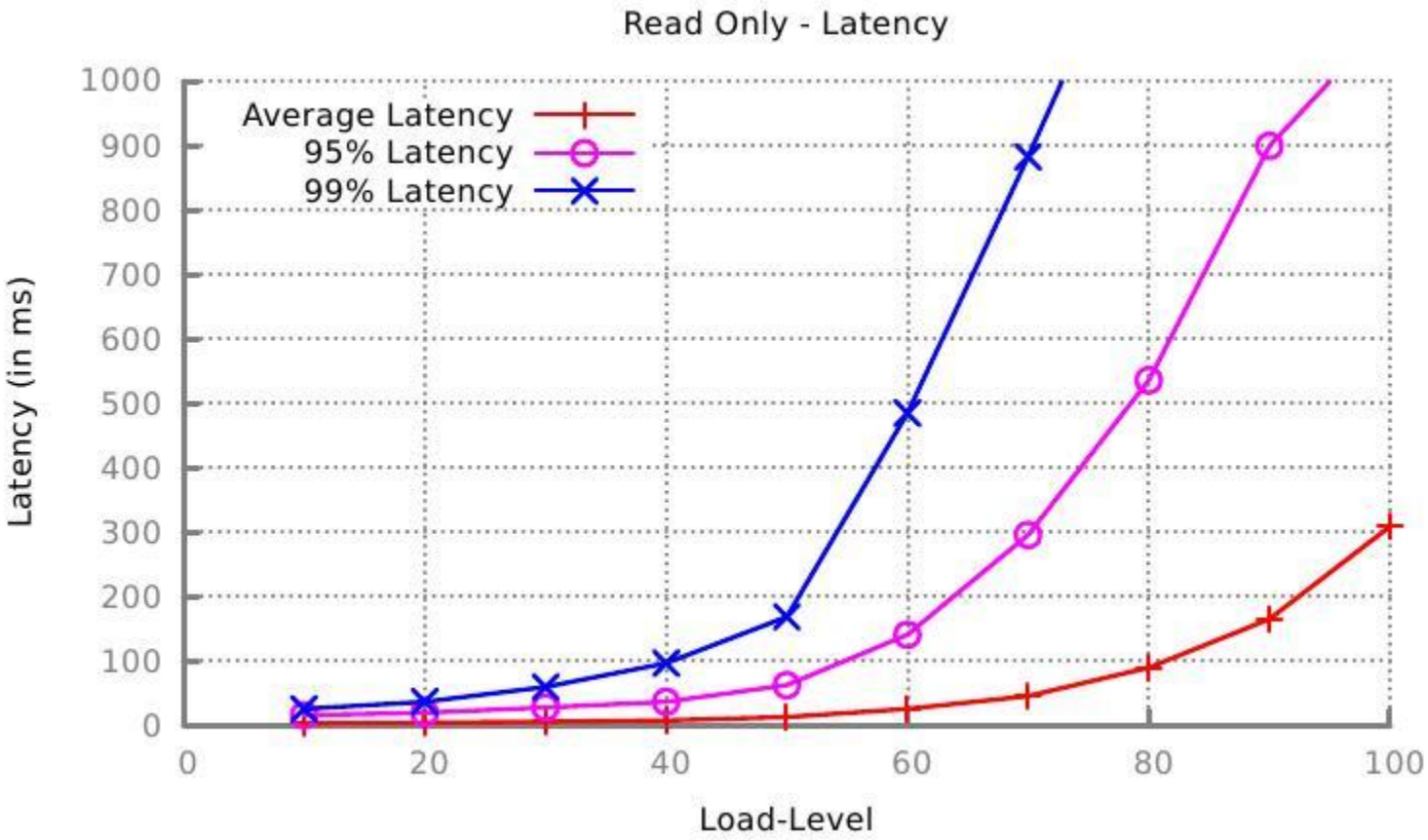
Baseline Energy Proportionality



Baseline Energy Proportionality

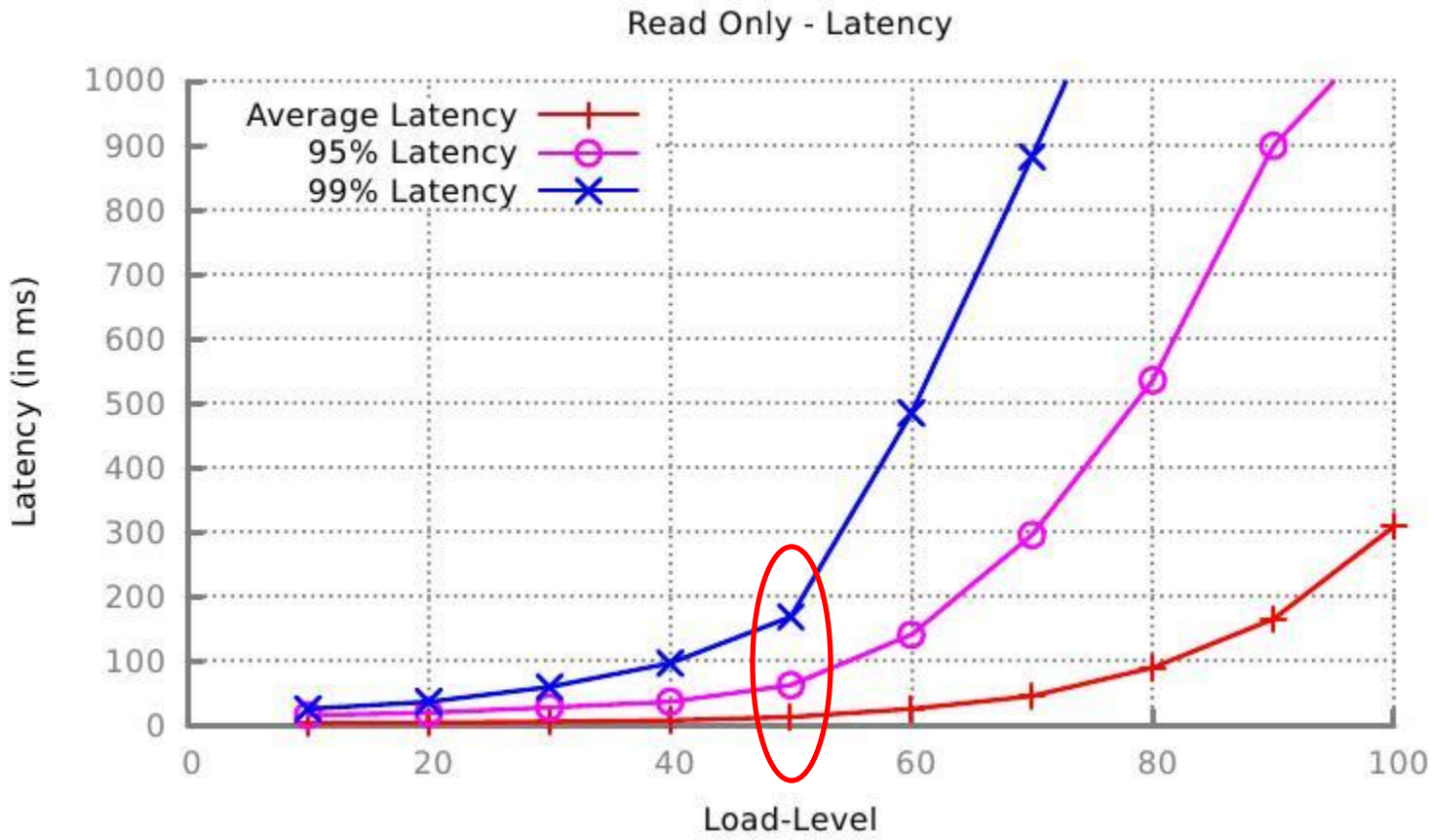


Load-Level vs. Latency



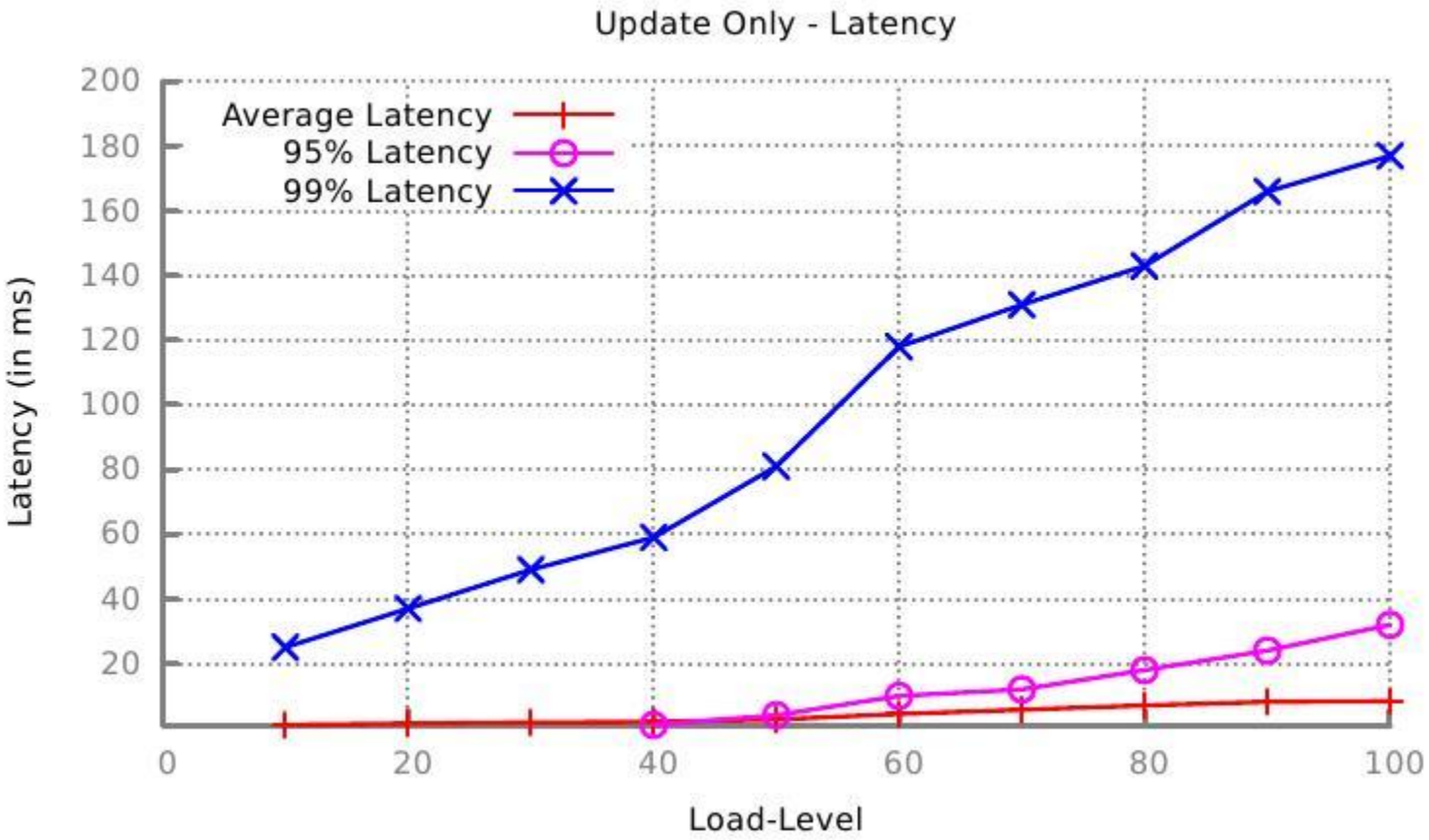
- Large differences in average, 95% and 99% latency curves

Load-Level vs. Latency



- Typical knee in latency curve

Load-Level vs. Latency

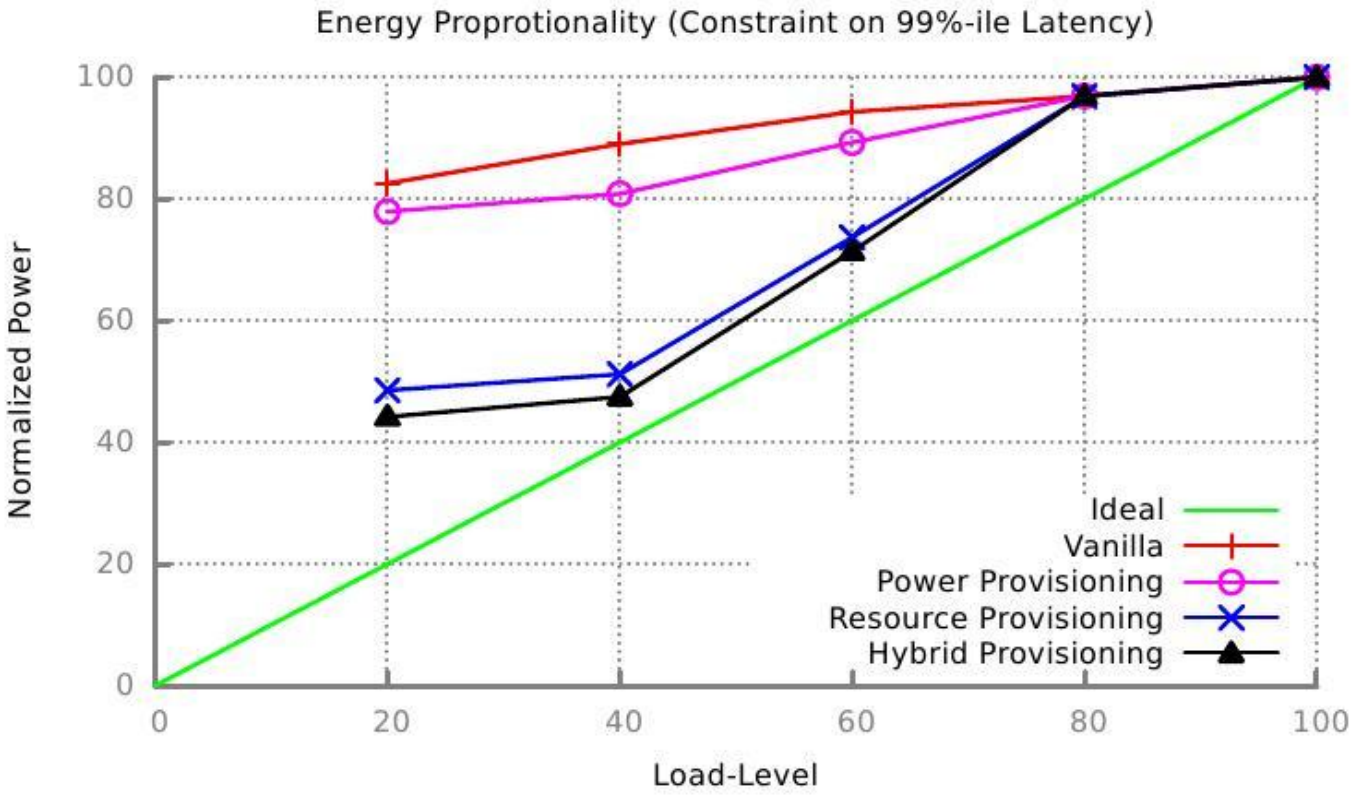


Balaji Subramaniam (balaji@cs.vt.edu)

Improving Energy Proportionality

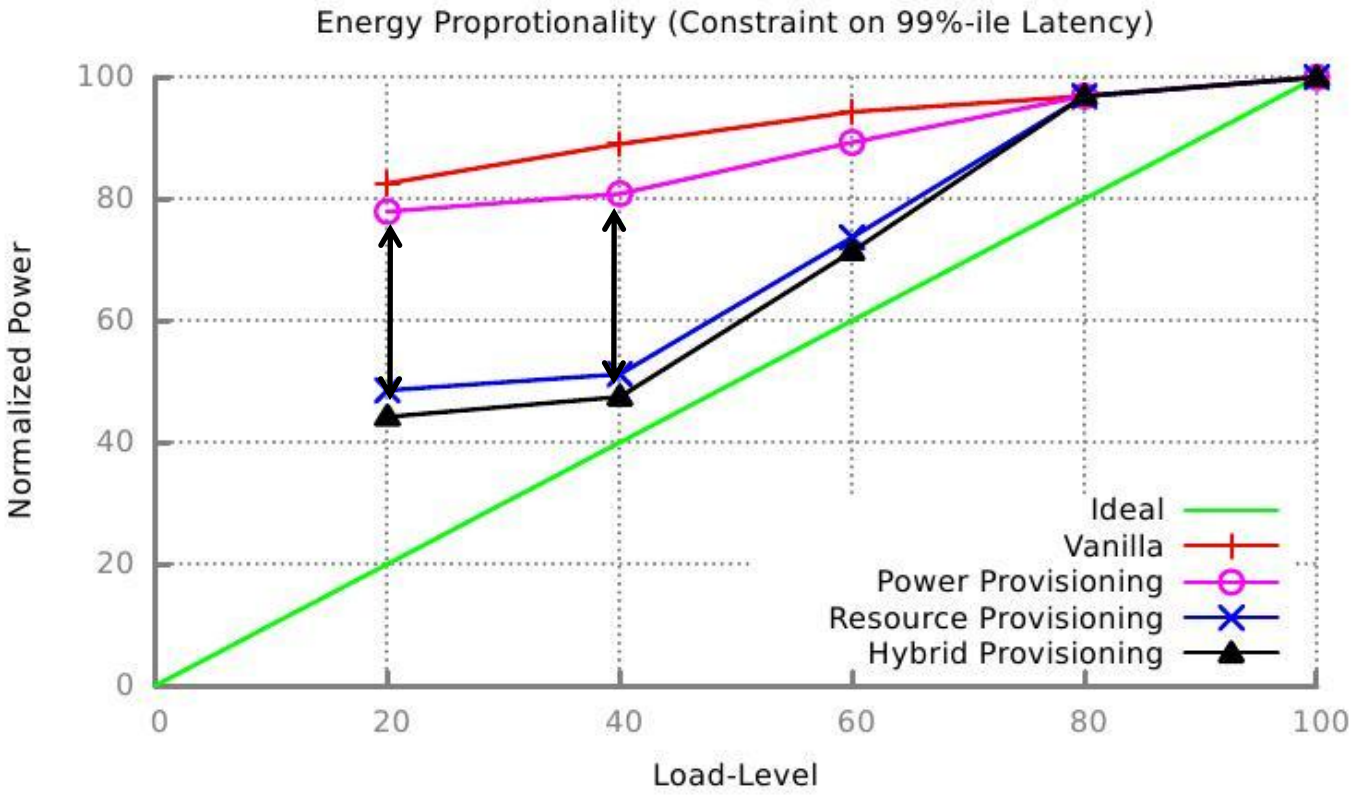
- Constraints on 99% latency = 1000ms
- Constraints on 99% latency = 600ms
- Three provisioning techniques evaluated
  - Power provisioning
  - Resource provisioning
  - Hybrid provisioning

Improving Energy Proportionality
Read-Only Workload



- Hybrid provisioning performs the best

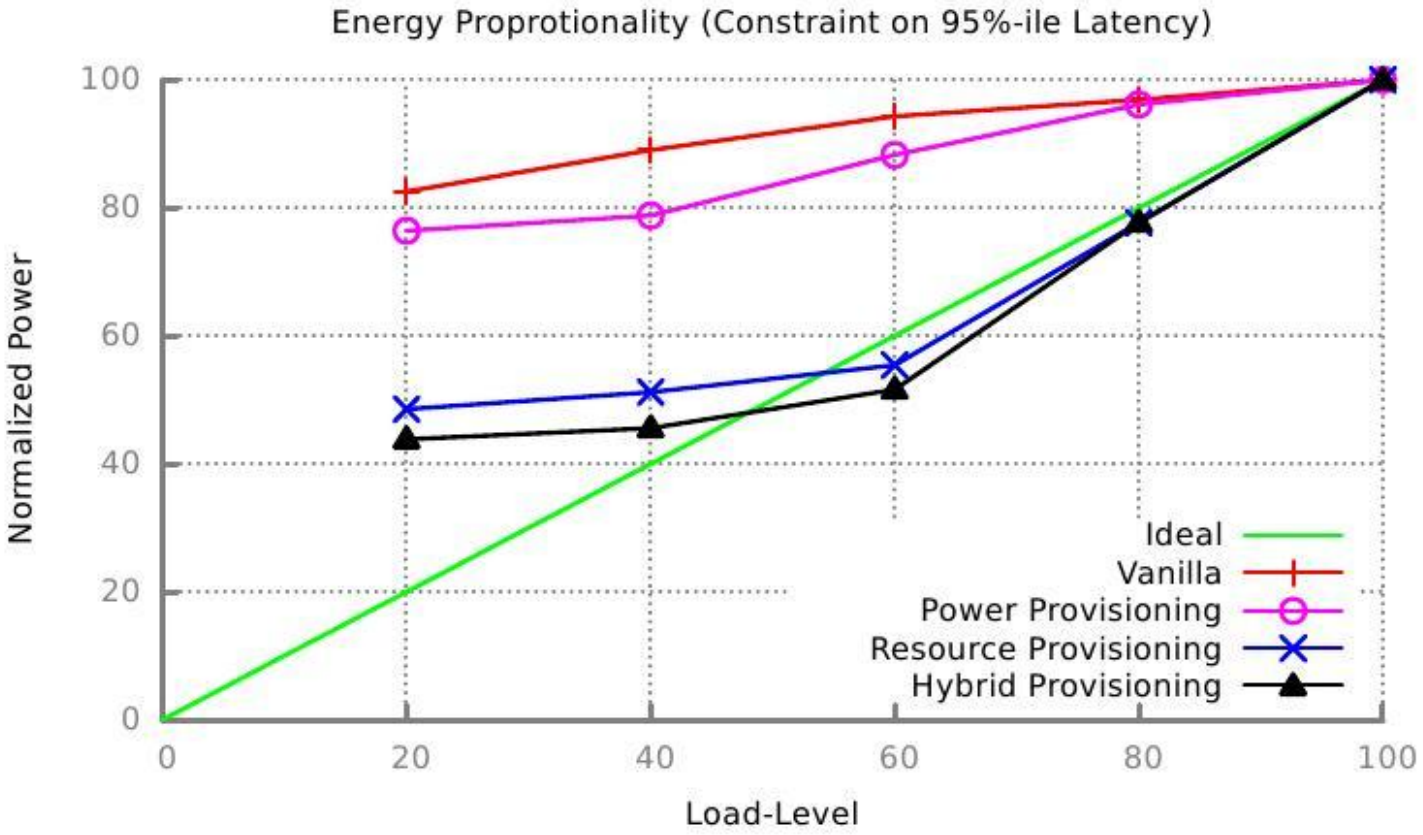
Improving Energy Proportionality
Read-Only Workload



- Differences due to resource provisioning

Balaji Subramaniam (balaji@cs.vt.edu)

Improving Energy Proportionality
Read-Only Workload



Balaji Subramaniam (balaji@cs.vt.edu)

Power Savings

Balaji Subramaniam (balaji@cs.vt.edu)

Contributions

- Baseline energy proportionality and latency
  - Power inefficiencies exists at different load-levels
  - Large differences between average, 95%-ile and 99%-ile latencies at different load-levels

Contributions

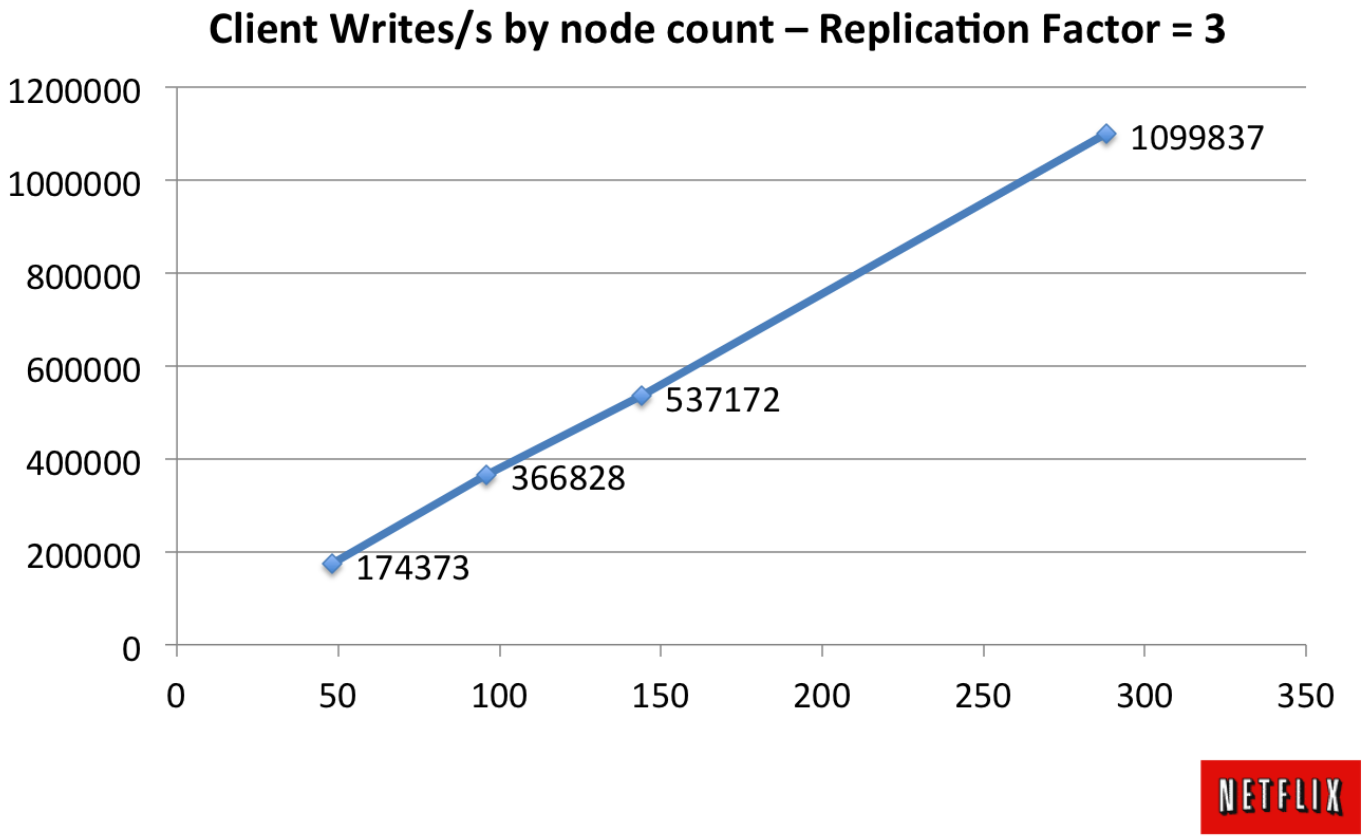
- Baseline energy proportionality and latency
  - Power inefficiencies exists at different load-levels
  - Large differences between average, 95%-ile and 99%-ile latencies at different load-levels
- Power vs. latency trade-offs
  - Latency can be traded for power consumption
  - Proper resource provisioning provides large power savings
  - Hybrid provisioning provides the best power savings

Big Data in Scientific Computing Meets NoSQL

- Why NoSQL?

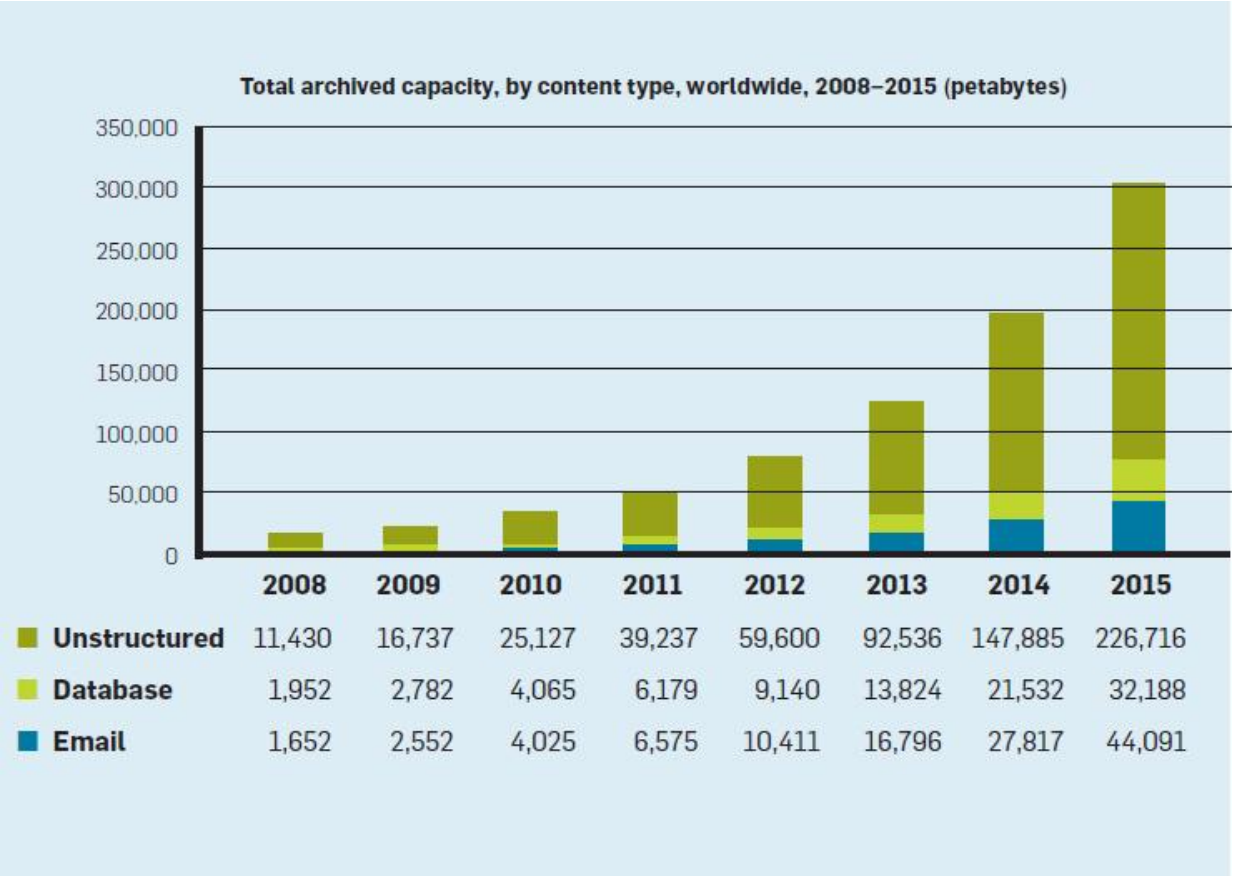
Big Data in Scientific Computing Meets NoSQL

- Why NoSQL?
  - Scalability



Big Data in Scientific Computing Meets NoSQL

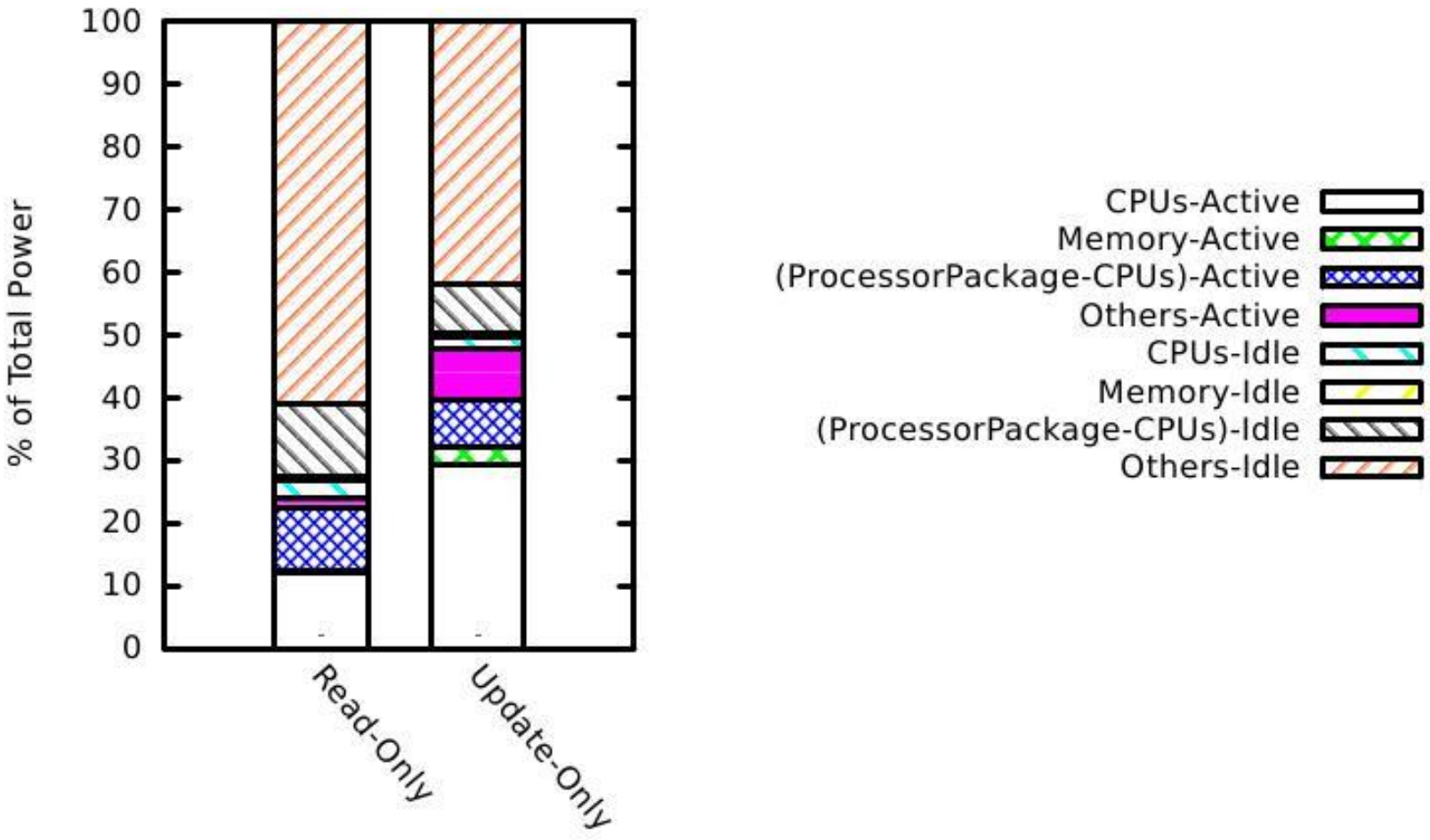
- Why NoSQL?
  - Growth in unstructured data



	2008	2009	2010	2011	2012	2013	2014	2015
Unstructured	11,430	16,737	25,127	39,237	59,600	92,536	147,885	226,716
Database	1,952	2,782	4,065	6,179	9,140	13,824	21,532	32,188
Email	1,652	2,552	4,025	6,575	10,411	16,796	27,817	44,091

Source: http://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext

Baseline Power Consumption



Balaji Subramaniam (balaji@cs.vt.edu)