Measuring the Energy Consumption of HPC Systems

Anne-Cécile Orgerie

ORAP Forum
9th December 2021
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Paris Agreement: 1.5° C

Objective in 2019: reducing global greenhouse gas emissions by 8% each year
ICT energy consumption grows by \(~9\%\) each year.

Anne-Cécile Orgerie

[HDR 2020]
My scientific context

• Energy consumption
• Large-scale distributed systems
• Computing and networking parts
• Use phase

Started with Grid computing some years ago…
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Energy efficiency: business as usual?

Computing faster?

Computing slower?
Energy efficiency: business as usual?

Computing faster?

Temperature matters.

Computing slower?

Anne-Cécile Orgerie
How to measure energy efficiency in DCs?

PUE: Power usage effectiveness

\[
PUE = \frac{Total\ Facility\ Power}{IT\ Equipment\ Power}
\]

Wrong idea #1

*Idle server consumes nothing or little.*

Nova node: 2 x Intel Xeon E5-2620 v4, 8 cores/CPU, 64 GiB RAM, 598 GB HDD (2016)
Wrong idea #2

This server model consumes that amount of power.

10% difference in idle and more at maximal consumption.
No chance for naive modeling

Naive model:
\[ 5 \times P_{\text{idle}} + 8 \times P_{\text{process}} = X \text{ Watts} \]

Best configuration for power consumption?
It depends.

Anne-Cécile Orgerie
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Energy consumption: a complex phenomenon

Need for **wattmeters** and sound experimental campaigns

- To understand
- To build robust models
- To get solid instantiations
- To obtain realistic algorithms
Performing measurements

Intel’s RAPL (Running Average Power Limit) interface

Warning: RAPL counters ignore a large part of the power consumption of servers.

Anne-Cécile Orgerie
Wrong idea #3

**RAPL counters capture most of the power consumption of a server.**

Big difference
Reproducibility?

Idle power consumption varies over time.

[Cluster 2017]
Methodology for measuring server consumption

Hardware tuning
Turbo Boost, hyper-threading, frequency drivers, governors.

Logging service
Timestamps, RAPL, MSRs, wattmeters, thermometers, configuration.

Experiment
Clock frequency, pinning, logging, benchmarking, sleeping.

Fixing all possible uncertainty sources.

Raw data
Configuration, timestamps, power, temperature, logs.

[ISCC 2021]
Wrong idea #4

The relation between power and CPU load is linear/quadratic/cubic.

17% difference in consumption for applications fully loading the server.

[Cluster 2017]
Wrong idea #5
For a given application, there is a least consuming configuration.

Faster with Turboboost, but consuming more energy.

Faster with Turboboost, and consuming less energy.

Anne-Cécile Orgerie
Wrong idea #6

Low power processors consume less energy.

BW_l: Xeon E5-2630L v4 (Broadwell) -> low power processor (orange)
BW: Xeon E5-2630 v4 (Broadwell) (green)

Anne-Cécile Orgerie [ISCC 2021]
Process placement onto cores

Up to 8% difference in average power consumption between unpaired and pairwise.

[JOCS 2020]
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Simulating energy consumption
Simulating energy consumption

Platforms + Experimental scenarios = Simulation models

Measurements
- Model ideas
- Instantiation values
- Validation results

Validation
- Instantiation
- Validity limits

Scientific results

Anne-Cécile Orgerie
Server profiling

To do for each computing kernel.
At each frequency.
And each time we want to compare the model to real life.

[Cluster 2017]
Simulating server clusters

Reproducible results: https://gitlab.inria.fr/fheinric/paper-simgrid-energy

Anne-Cécile Orgerie
Wasted energy at all levels of data centers

Cooling
Power generators
Batteries
...

Unused servers
Overprovisioning
Redundancy
...

Power non-proportionality
Dark silicon
Unused components
...

Anne-Cécile Orgerie
Models and simulation tools for what?

Capacity and energy planning
What-if scenarios
Algorithm analysis
Estimating VM energy consumption
Estimating end-to-end energy consumption
Closing doors
Outline

• Context

• Understanding the energy consumption of HPC systems

• Measuring accurately the energy consumption of HPC systems

• Modeling energy consumption of HPC systems

• Concluding broader remarks
Increasing energy efficiency ≠ reducing consumption
Increasing energy efficiency ≠ reducing consumption

Energy optimization → Resource cost reduction → Usage increase → Energy consumption increase

Beware of rebound effects!
Full life cycle of servers

Dell PowerEdge R430 (Nova cluster)

Estimated carbon footprint (by Dell): 8,150 kgCO2e

Assumptions for calculating product carbon footprint:

<table>
<thead>
<tr>
<th>Product Weight</th>
<th>26.3 kg</th>
<th>Server Type</th>
<th>Rack</th>
<th>Assembly Location</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Lifetime</td>
<td>4 years</td>
<td>Use Location</td>
<td>EU</td>
<td>Energy Demand (Yearly TEC)</td>
<td>1760.3 kWh</td>
</tr>
<tr>
<td>HDD/SSD Quantity</td>
<td>x2 1TB 3.5” HDD</td>
<td>DRAM Capacity</td>
<td>16GB</td>
<td>CPU Quantity</td>
<td>2</td>
</tr>
</tbody>
</table>

Anne-Cécile Orgerie

Source: Dell PowerEdge R430 carbon footprint, 2019.
Life cycle of end devices

iPad Pro (12.9-inch) life cycle
136 kg carbon emissions

Use

Source
Materials

Make

Package
and Ship

Recover

ENERGY STAR limit

iPad Pro (12.9-inch)

83% Production
11% Transport
6% Use
<1% End-of-life processing

4 years of use

Numerous other environmental impacts

Product Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model name</td>
<td>SM-N950U (Galaxy Note8)</td>
</tr>
<tr>
<td>Processor</td>
<td>Qualcomm 2.35GHz, 1.9GHz Octa-Core 64bit</td>
</tr>
<tr>
<td>Dimension</td>
<td>162.5 x 74.8 x 8.6 mm</td>
</tr>
<tr>
<td>Display</td>
<td>6.3” 2960 x 1440, 16M In-Cell Touch LCD</td>
</tr>
<tr>
<td>Battery</td>
<td>Li-ion 3300 mAh</td>
</tr>
<tr>
<td>Camera</td>
<td>12 MP / 5MP</td>
</tr>
<tr>
<td>Wt.(g)</td>
<td>189.34g</td>
</tr>
</tbody>
</table>

Material Use

Characterized Environment Impact

Thank you for your attention

http://people.irisa.fr/Anne-Cecile.Orgerie
Citations


Saving energy

- Low power processors (big.LITTLE)
- Multi-core architectures
- Energy-efficient dedicated architectures (FPGA, GPU)
- Dynamic Voltage Frequency Scaling
- Workload consolidation techniques
- On/off policies
- Hot spot management
- Workload peak reduction
- Dynamic adaptation
Designing energy efficient algorithms

5 DCs with 20 homogeneous servers each, no migration

Optimal solution (dynamic programming algorithm) => 2 weeks of computation on 30 Grid’5000 servers

SAGITTA is close to the optimal solution.
VM migration algorithm

1. **Pre-allocation**: incoming VM requests

   [...] 

2. **Migration**: moving running VMs between DCs with network constraints
   
   a. Evaluate energy costs (VM migrations) and gains (expected remaining green energy on DCs)
   
   b. Schedule the VM migrations between DCs

3. **Consolidation**: packing VMs inside DCs

4. **Allocation**: actually send the commands to the servers

   a. Switch ON/OFF servers

   b. Deploy and migrate the VMs
Energy-efficient algorithm dissection

9 DCs and 1,035 servers in total

Theoretical lower bound => best-fit on a single DC

State-of-the-art: MBFD, OOD-MARE

[SBAC-PAD2018]