# Proportional Computing

Georges Da Costa, IRIT, Toulouse University

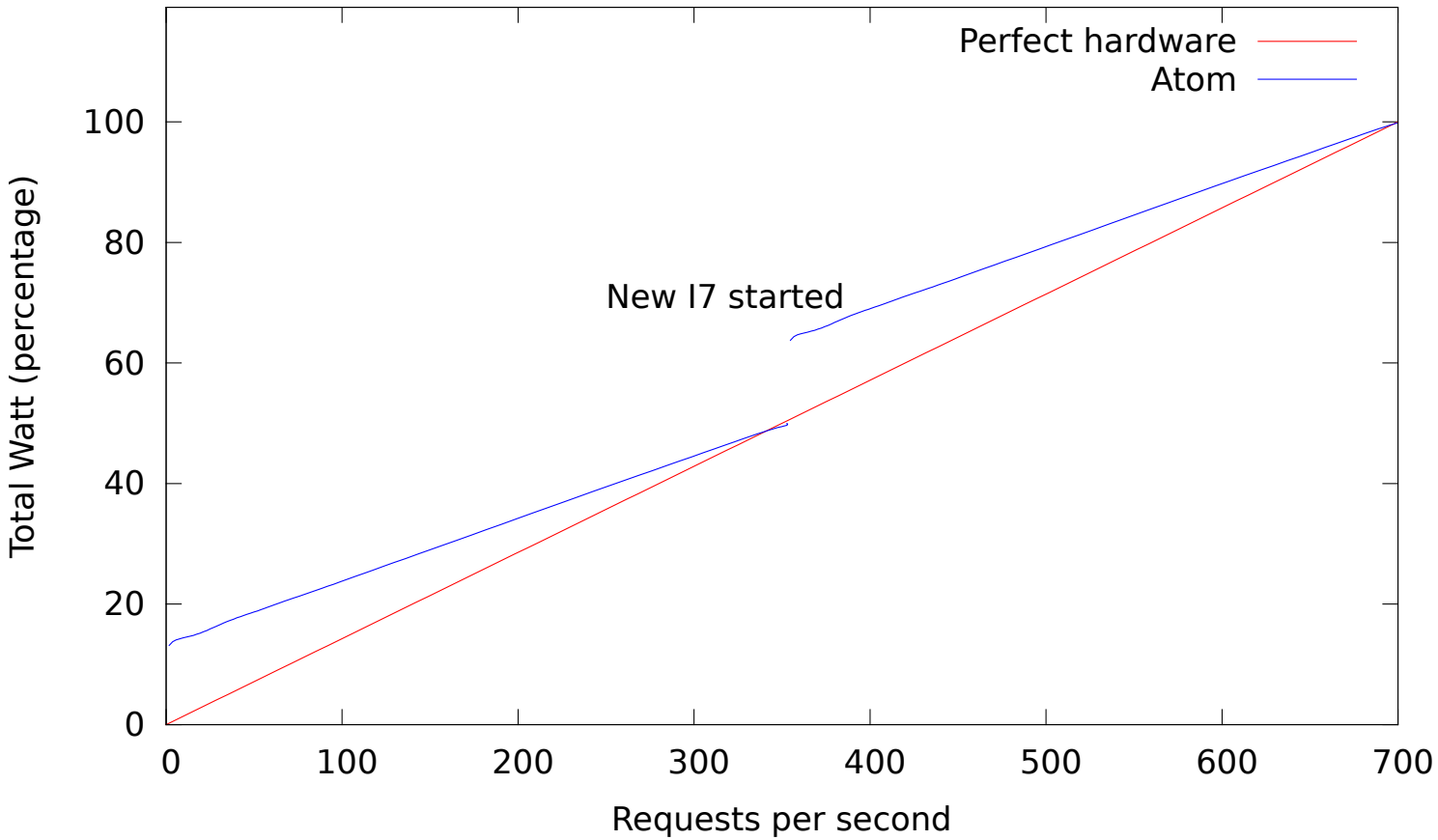
Work in progress
29/01/2013


Green Days @ Luxembourg

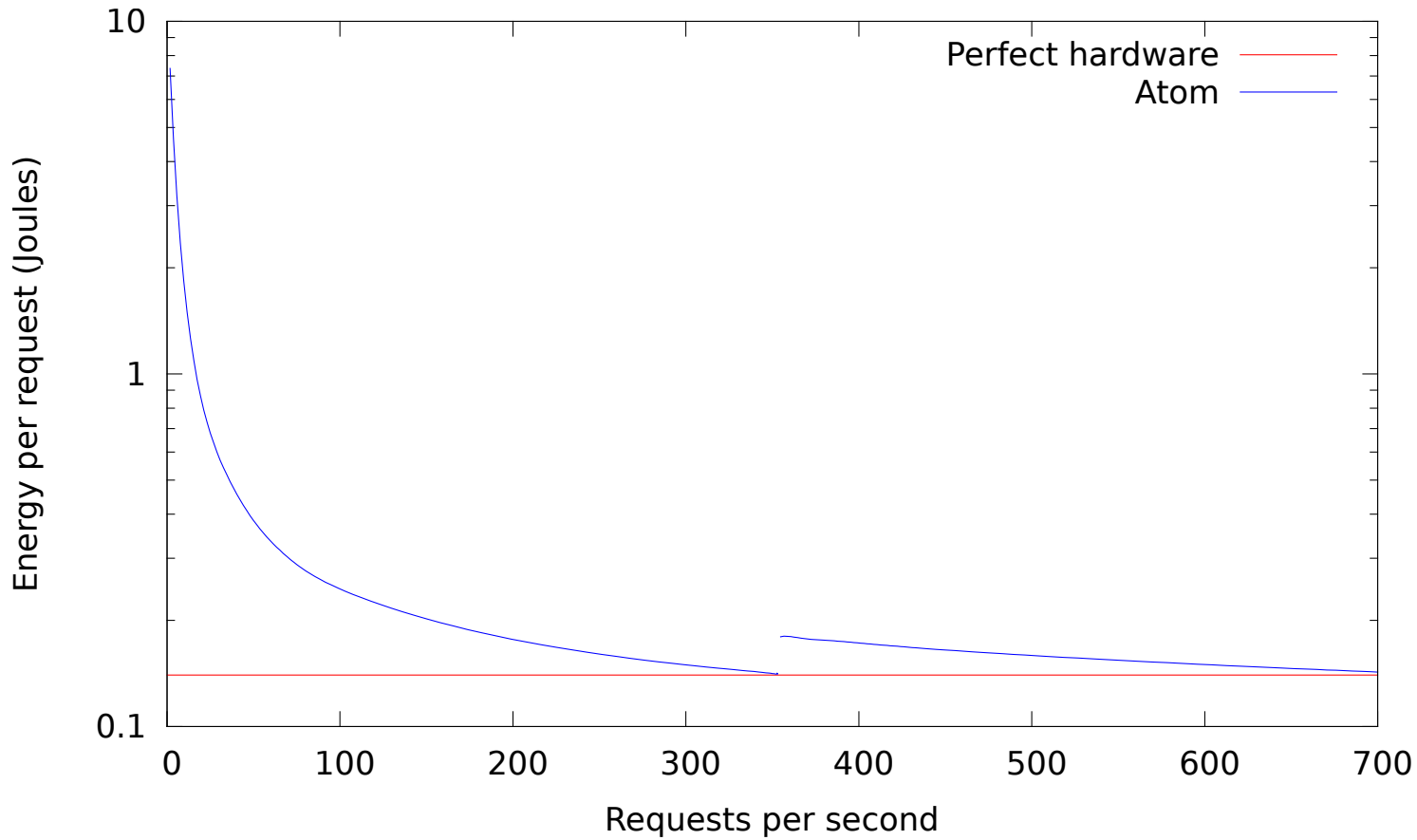

# Plan

# Small is beautiful

- Currently, lots of work is done on consolidation
- Metric are evaluated at full charge
- For all usages ?
- Not for:
  - Irregular workload
  - Low workload
- Why is it important ?
  - Because we are not in a linear world !
  - Even for really efficient systems (at full charge)




Load: Complex web service (computing power is preponderant on communication)




Problem: On low load, Watt/Request is bad

# Current methods

- Current methods:
  - On high load, consolidation.
  - On high number of requests, overhead is spread on lots of nodes
  - But wasted Watts continue to add-up
- What do we want ?
  - proportional computing
  - idle load = 0W

# Plan

# The big lie of homogeneous systems

- Most current approach : Homogeneous systems
- True in HPC centers
- But data centers are constructed step by step
  - Historical building of DC $\rightarrow$ quite different hardware
  - Position in a data center : distance to cooling infrastructure
- But also, compare:
  - Very energy efficient systems at low load, high overhead
  - Energy-inefficient system at max load, low overhead

# Our idea

Use nodes depending on the real load, not the peak load

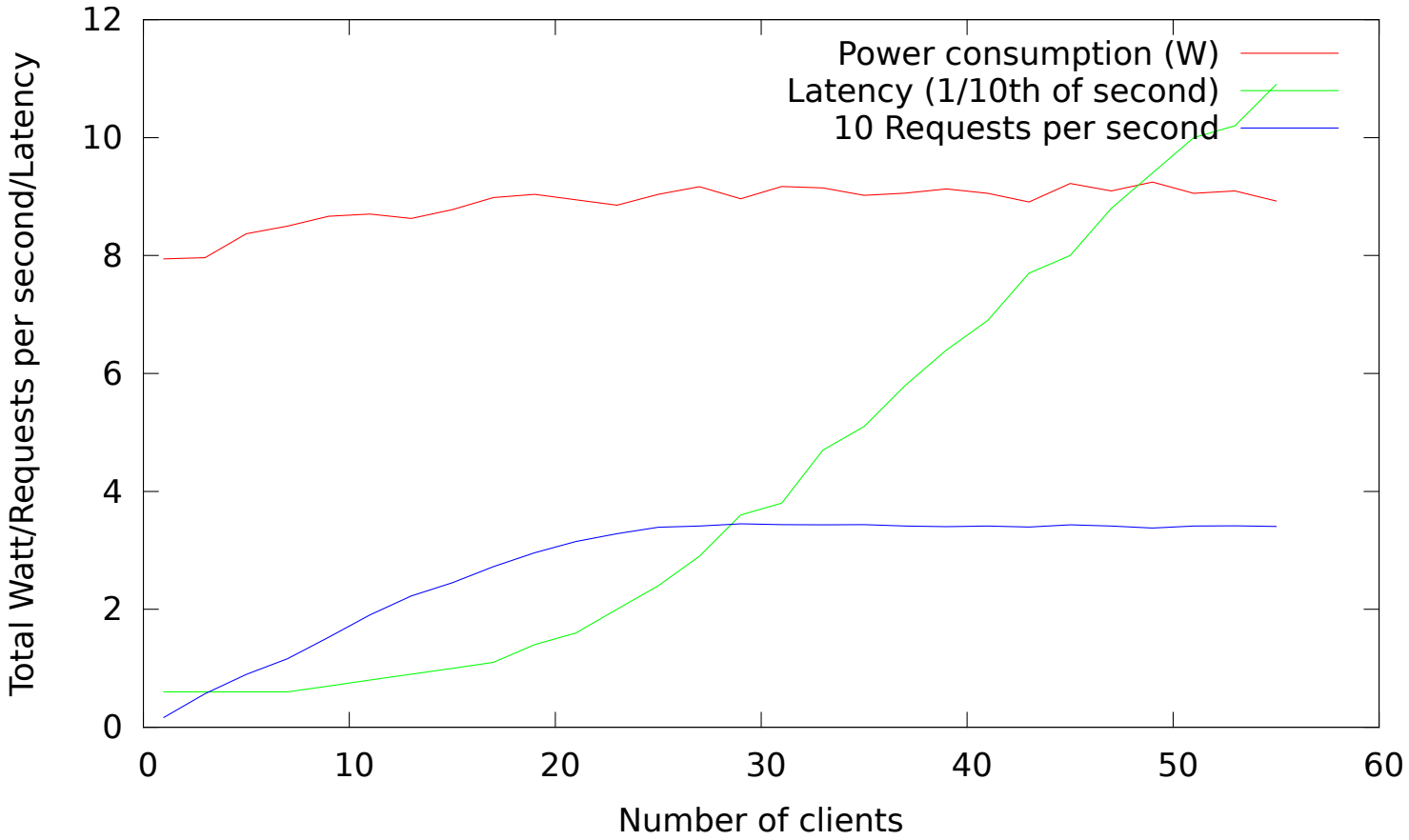
**Example:**

| Processor | Watt range | Max request/s | Efficiency (W/r) |
|---|---|---|---|
| Intel I7 | 11 - 42 | 353 | .12 |
| Intel Atom | 8 - 9 | 34 | .26 |
| Raspberry Pi | 2.56 - 2.81 | 5.6 | .50 |

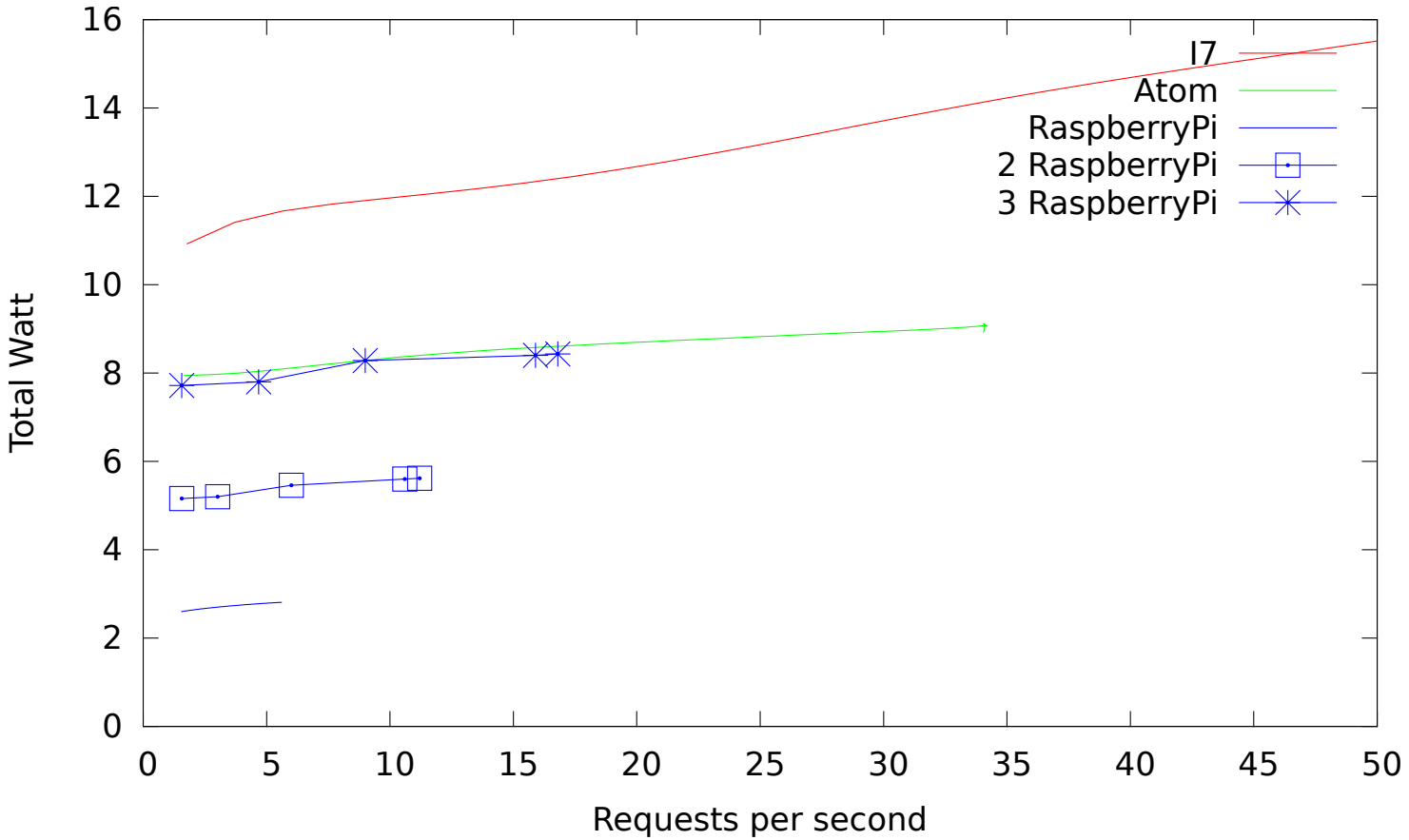
Intuition: Several small node and intermediary nodes to have a multi-scale smooth curve

# Zoom on Intel I7

# Zoom on Intel Atom



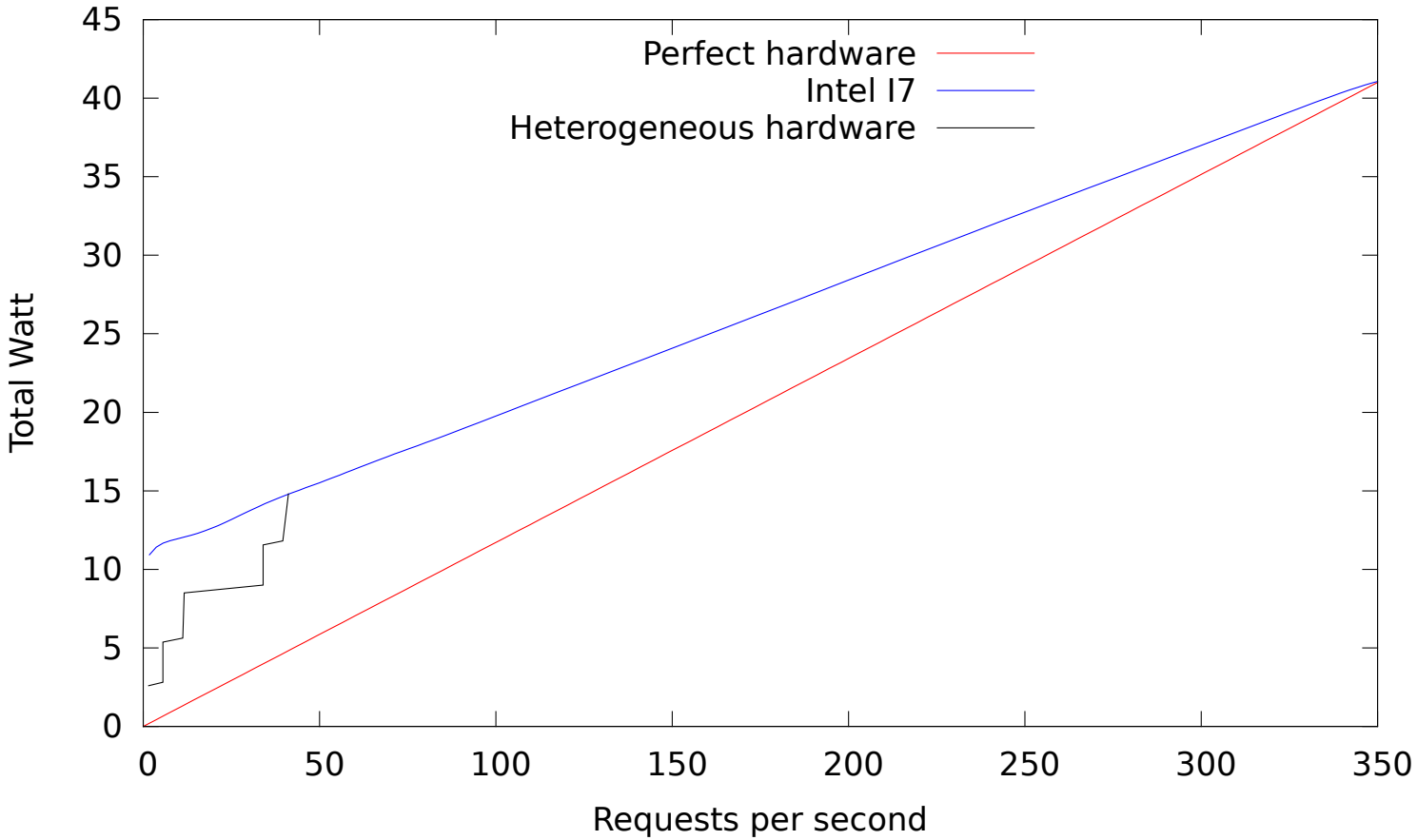
# Comparison


16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
40
45
50
Total Watt
Requests per second
I7
Atom
RaspberryPi
2 RaspberryPi
3 RaspberryPi

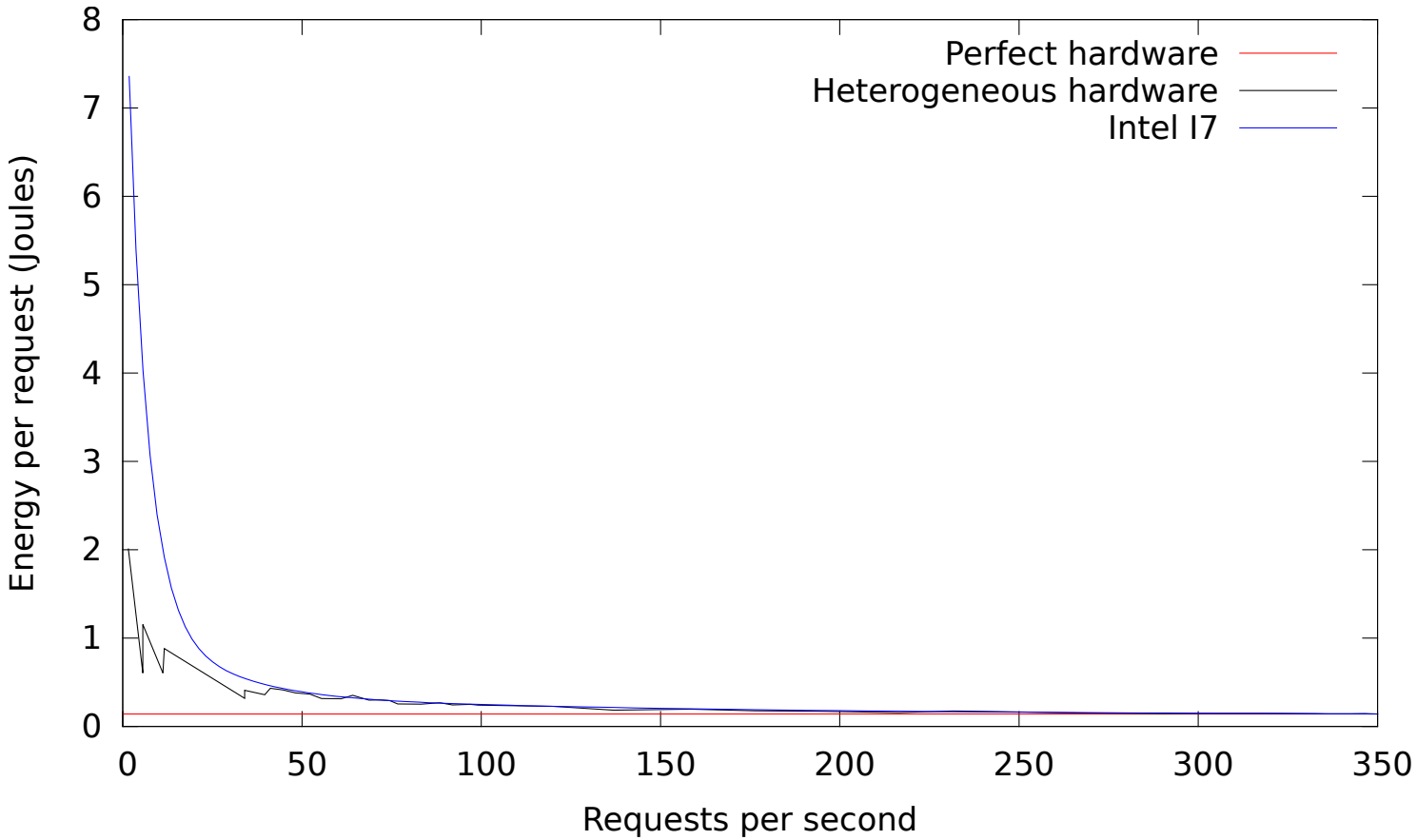
# A near-linear behavior

- 0→5 req/s 1Pi
- 5→10 req/s 2Pi
- 10→35 req/s 1Atom
- 35→40 req/s 1Atom+1pi
- 40→350 req/s 1I7

# Still far far away...



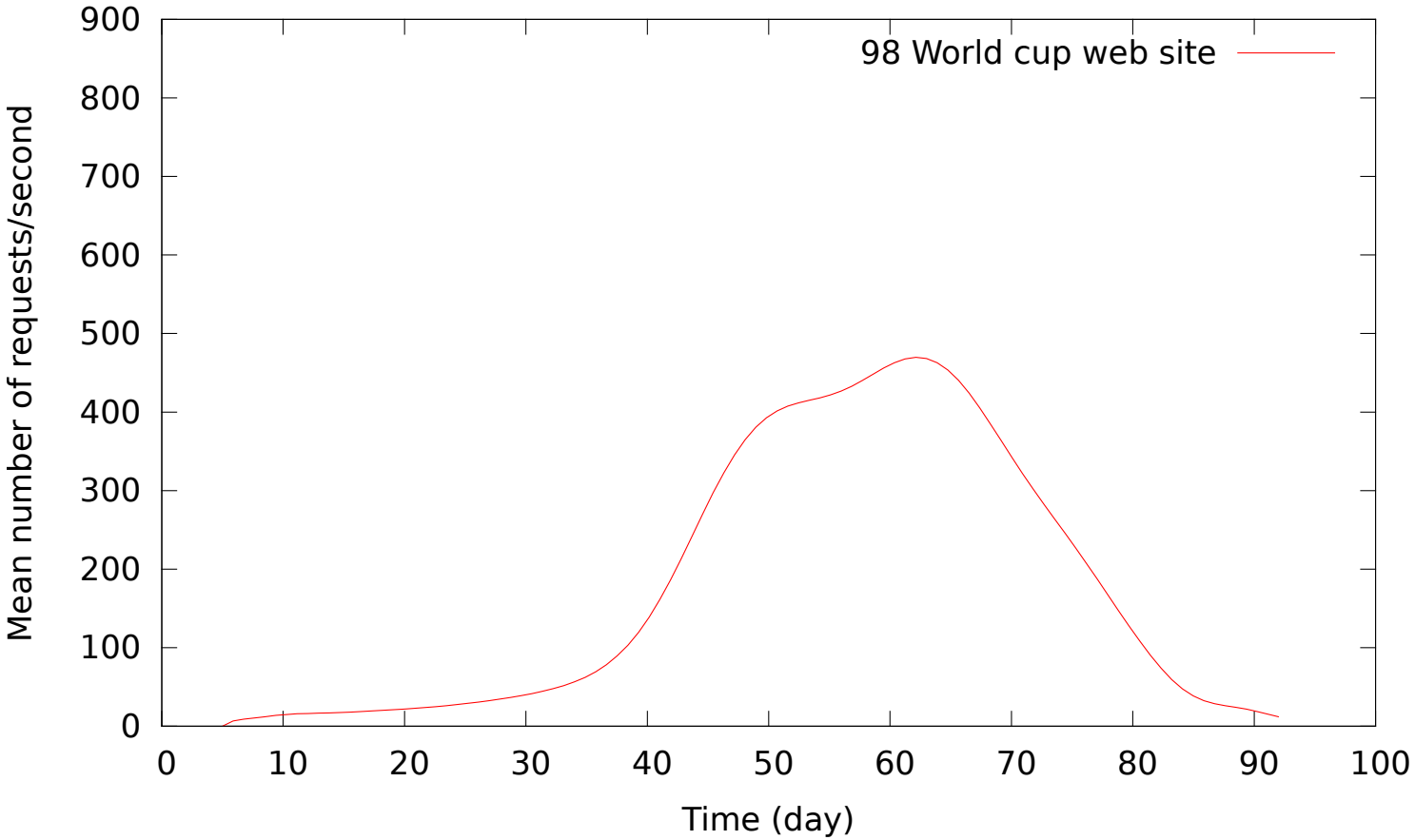
# Not so far, efficiency view !



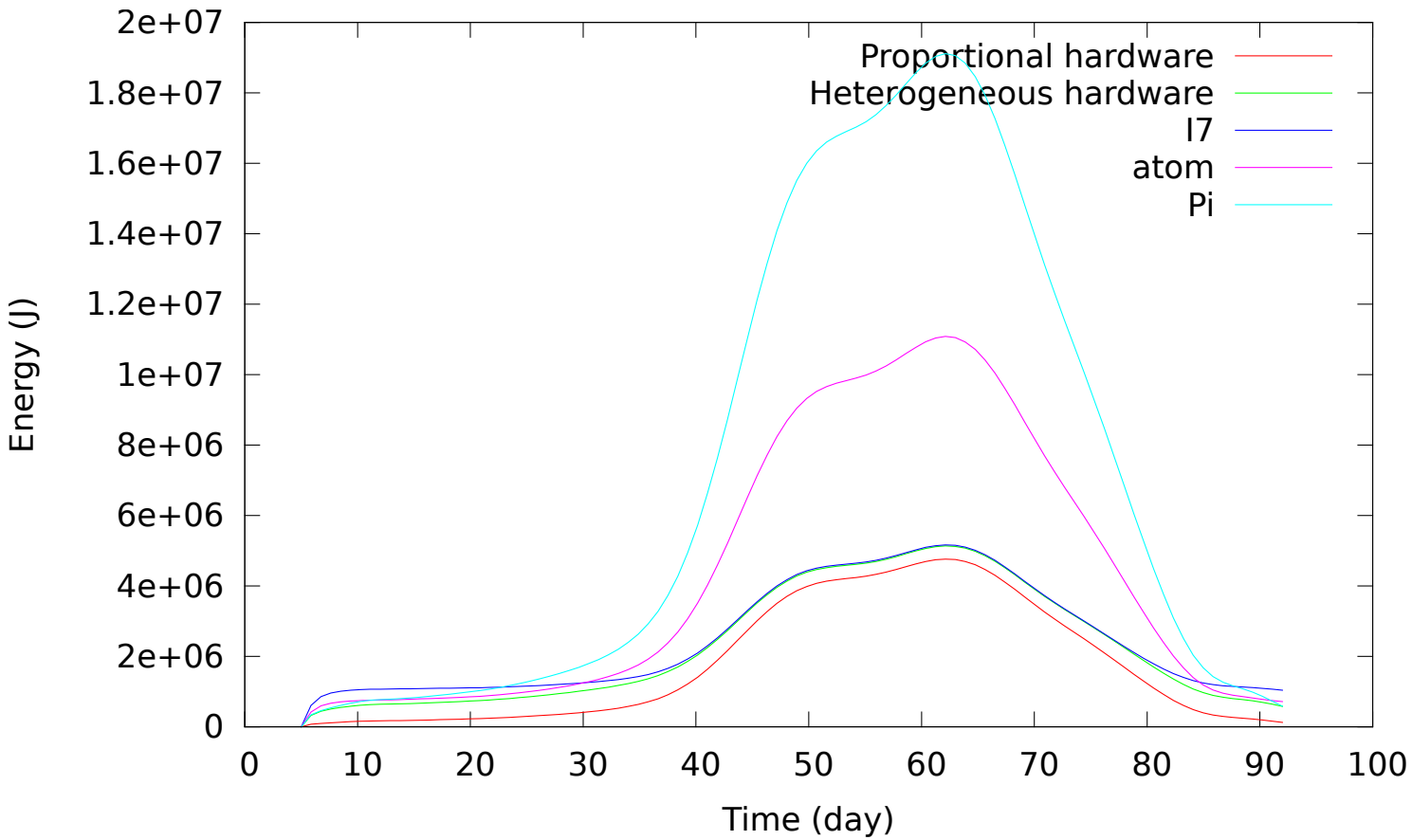
# Plan

# 98 Football World Cup

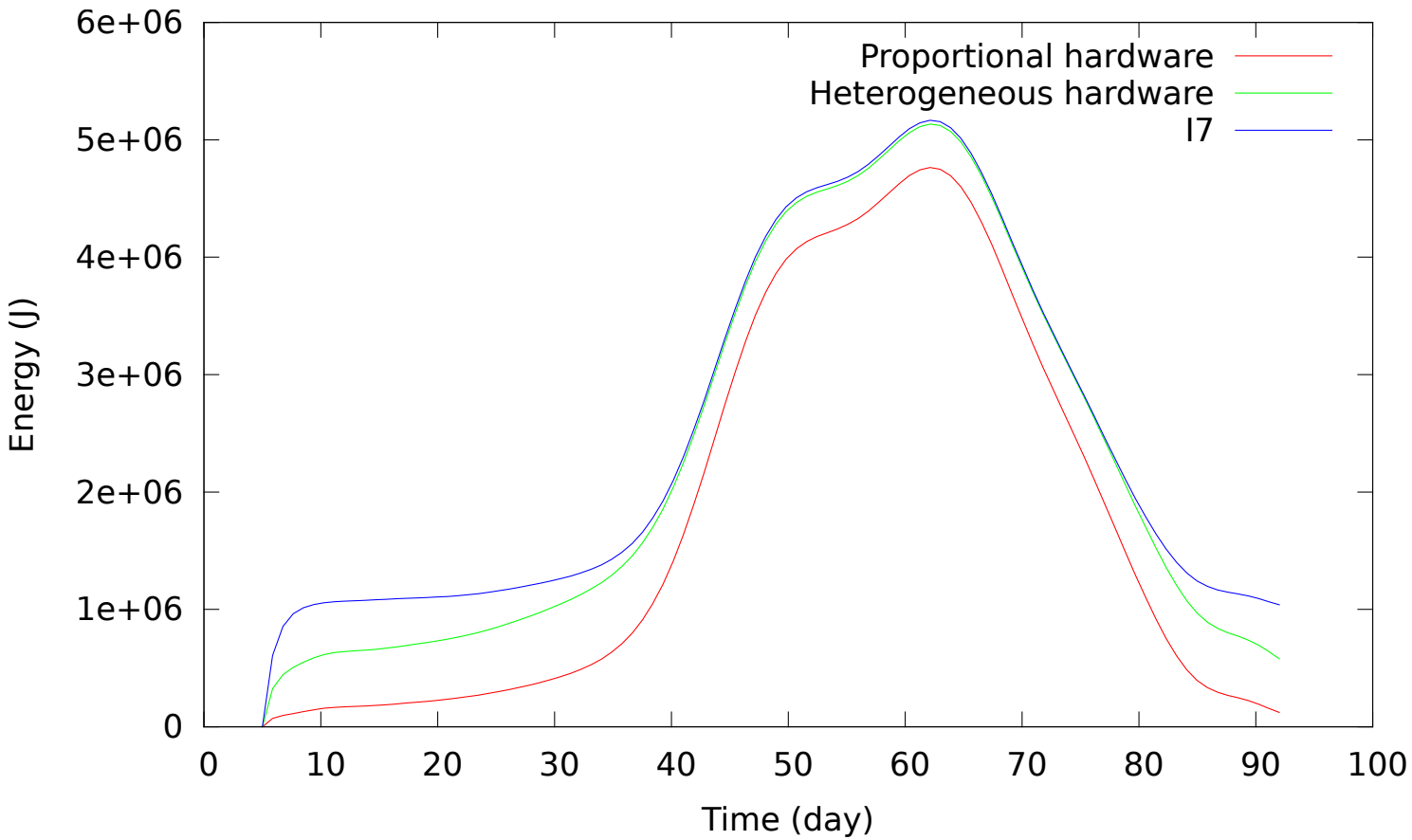
- Available data
- 92 days of web server access logs
- Workload precise at the second level
- Several phases
  - Low phases, first 40 days and last 10 days
  - High phase, during the competition


98 World cup web site
Mean number of requests/second
Time (day)
0
100
200
300
400
500
600
700
800
900
0
10
20
30
40
50
60
70
80
90
100

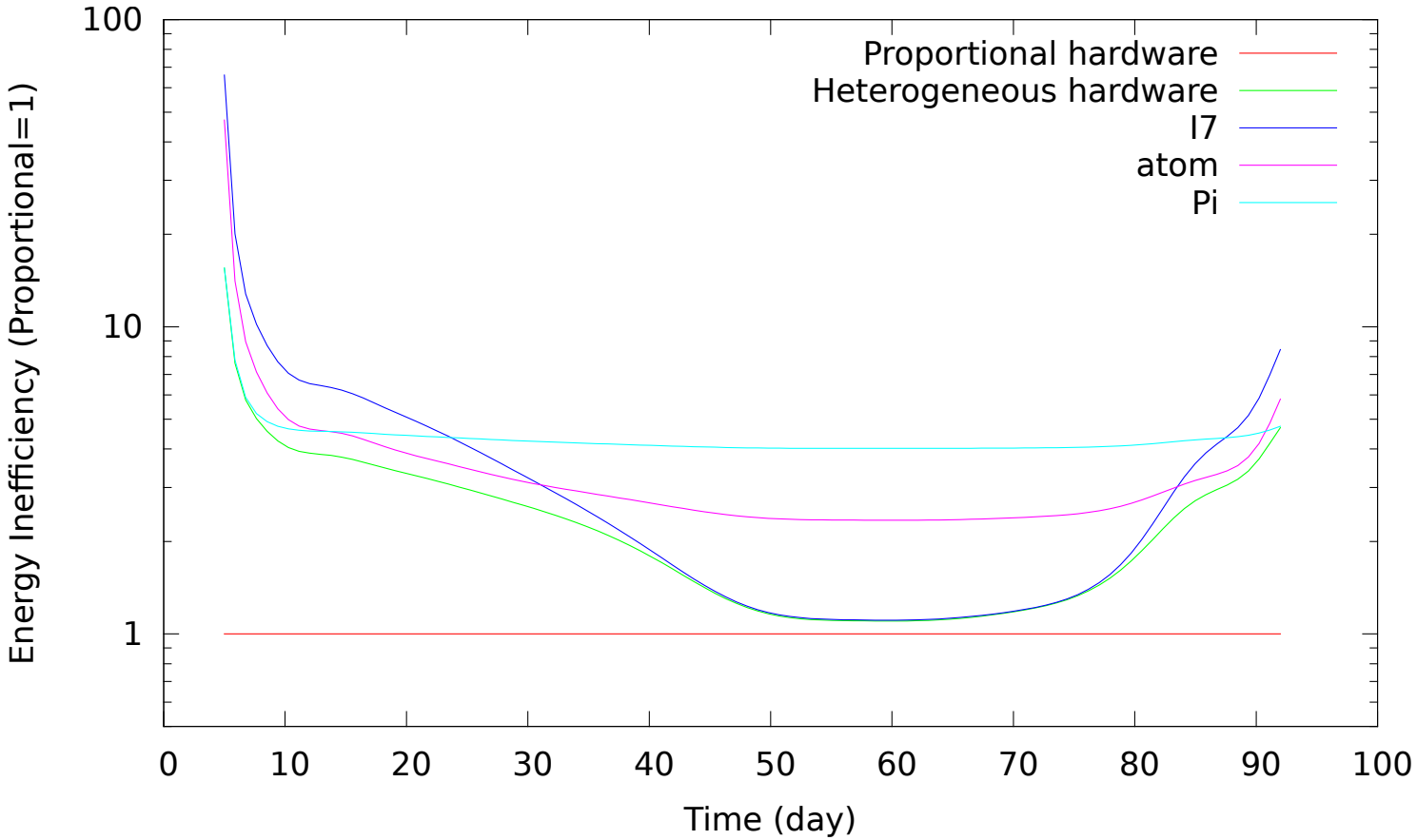
# Comparison


Energy (J)
2e+07
1.8e+07
1.6e+07
1.4e+07
1.2e+07
1e+07
8e+06
6e+06
4e+06
2e+06
0
0
10
20
30
40
50
60
70
80
90
100
Time (day)
Proportional hardware
Heterogeneous hardware
I7
atom
Pi

# Zoom on the most efficient methods



# Far reached goal: proportional computing



# Plan

# Conclusion & perspective

- Improve efficiency for low load
- Still inefficient most of the time
- Next steps
  - Increase range of efficiency with different hardware
  - Take into account frequencies
  - Save the world