Adept

addressing energy in parallel technologies

Power measurement at the exascale

Nick Johnson, James Perry & Michèle Weiland

Nick Johnson

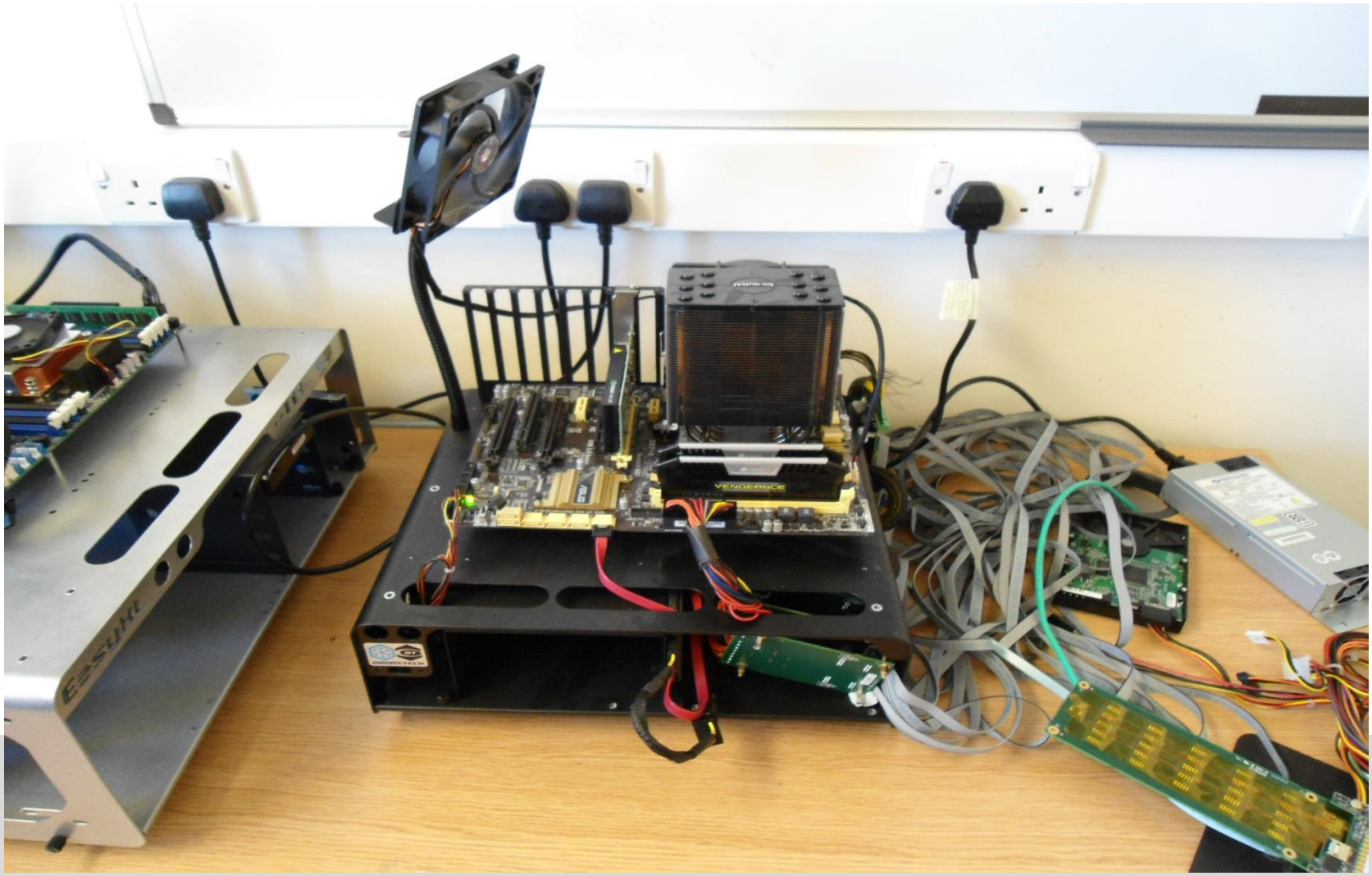
Adept Project, EPCC

nick.johnson@ed.ac.uk

Motivation

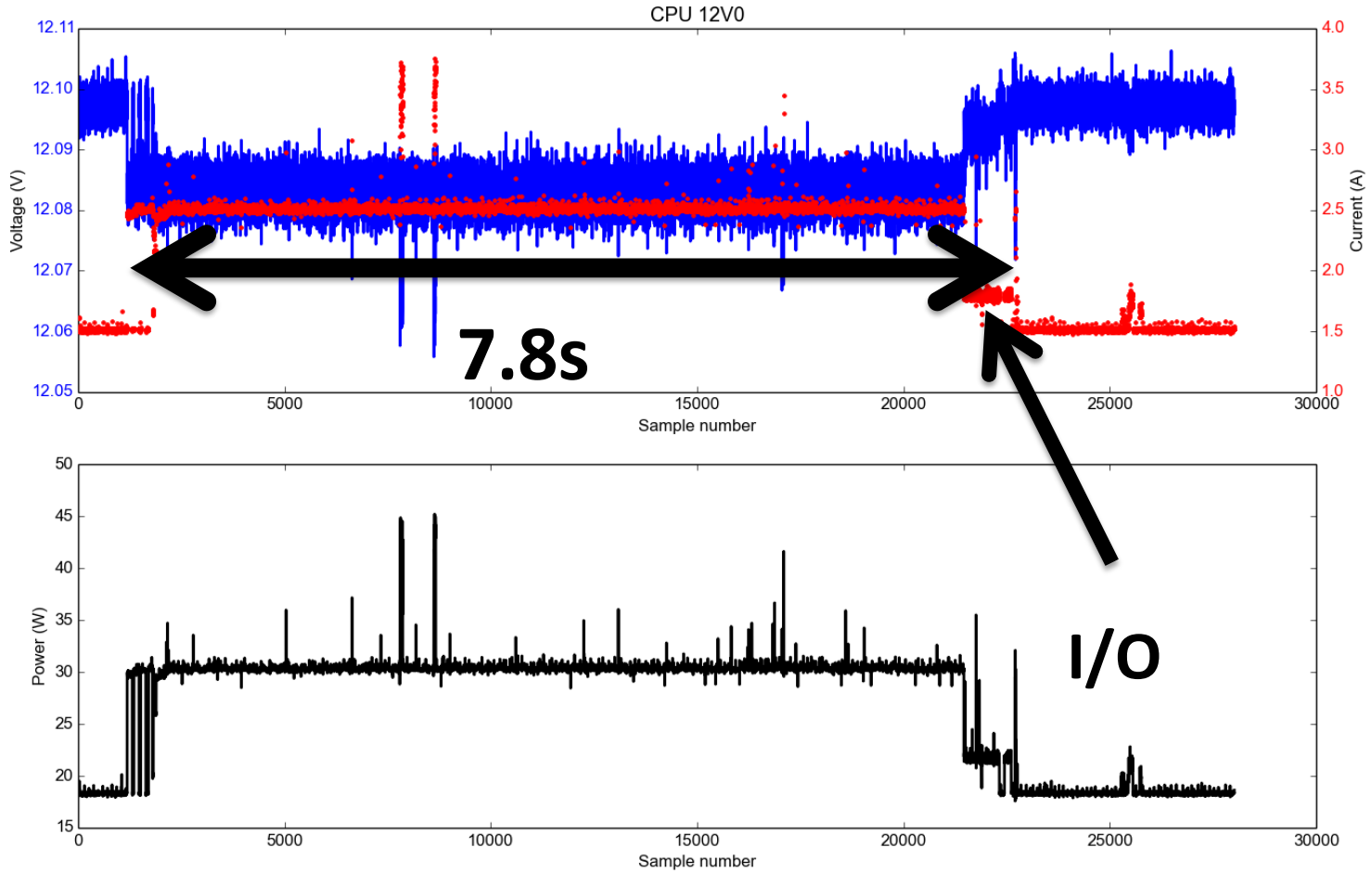
- The current exascale targets are:
  - One exaflop at a power rating of twenty to forty mega-watts (MW) by 2020.[1]
- Measuring the performance is not hard
  - Use known benchmarks; e.g. HPL
- Measuring the power is more contrived
  - Current supercomputers separate out compute, support nodes, cooling etc.
  - In a shared infrastructure, must take into account fractions of e.g. cooling infrastructure.
  - See EEHPCWG guideline for a further discussion.[2]

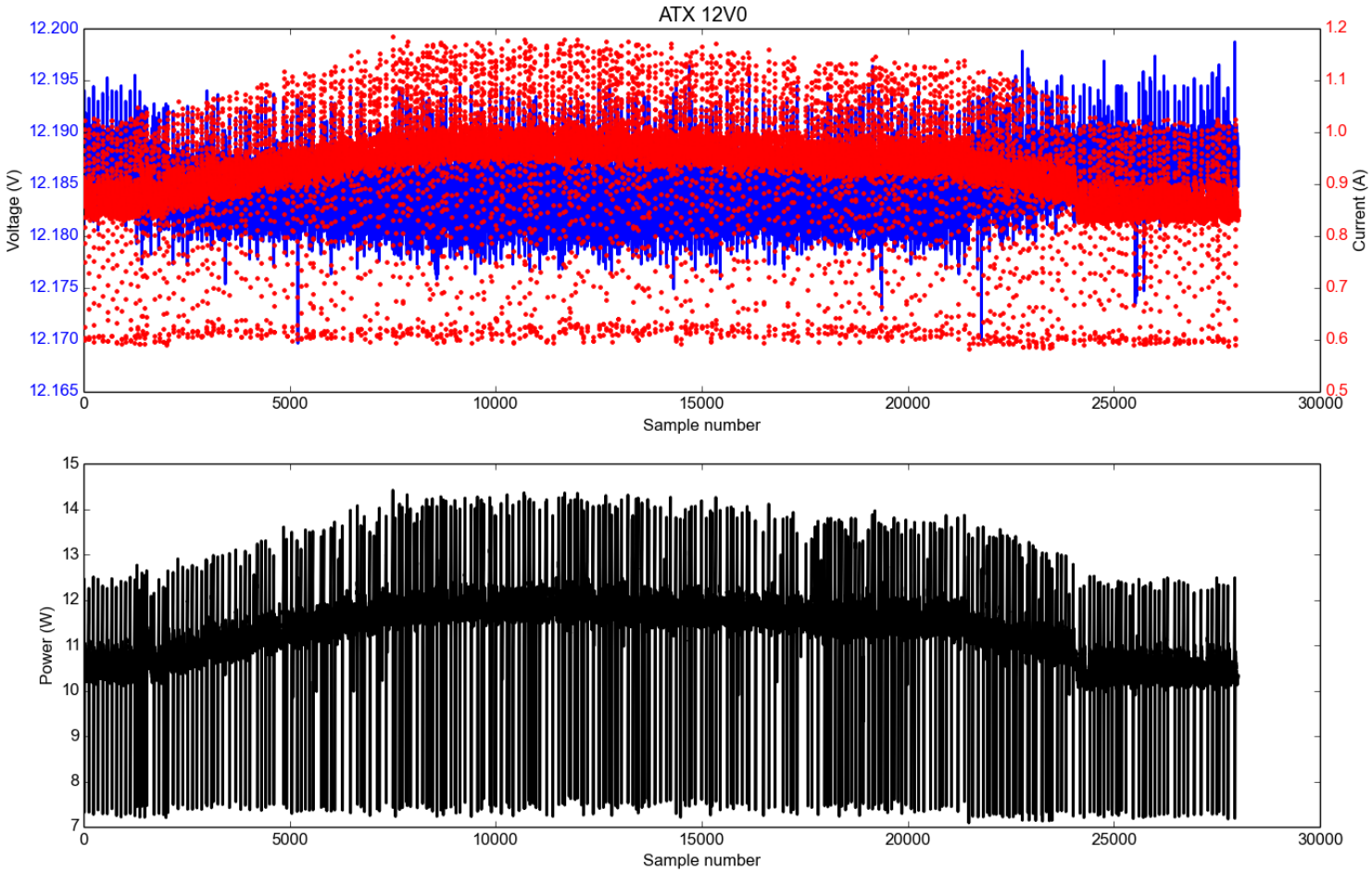
- One solution to the power problem is to consider the power profile of a code and optimize for energy to solution.
- Current measurement methods are generally:
  - In-band
  - Based on performance counters
  - Based on models
  - Almost non-existent for heterogeneous architectures.
- Using an out of band measurement system, can we measure an existing piece of hardware running a scalable code and predict exascale performance, power consumption & energy to solution.

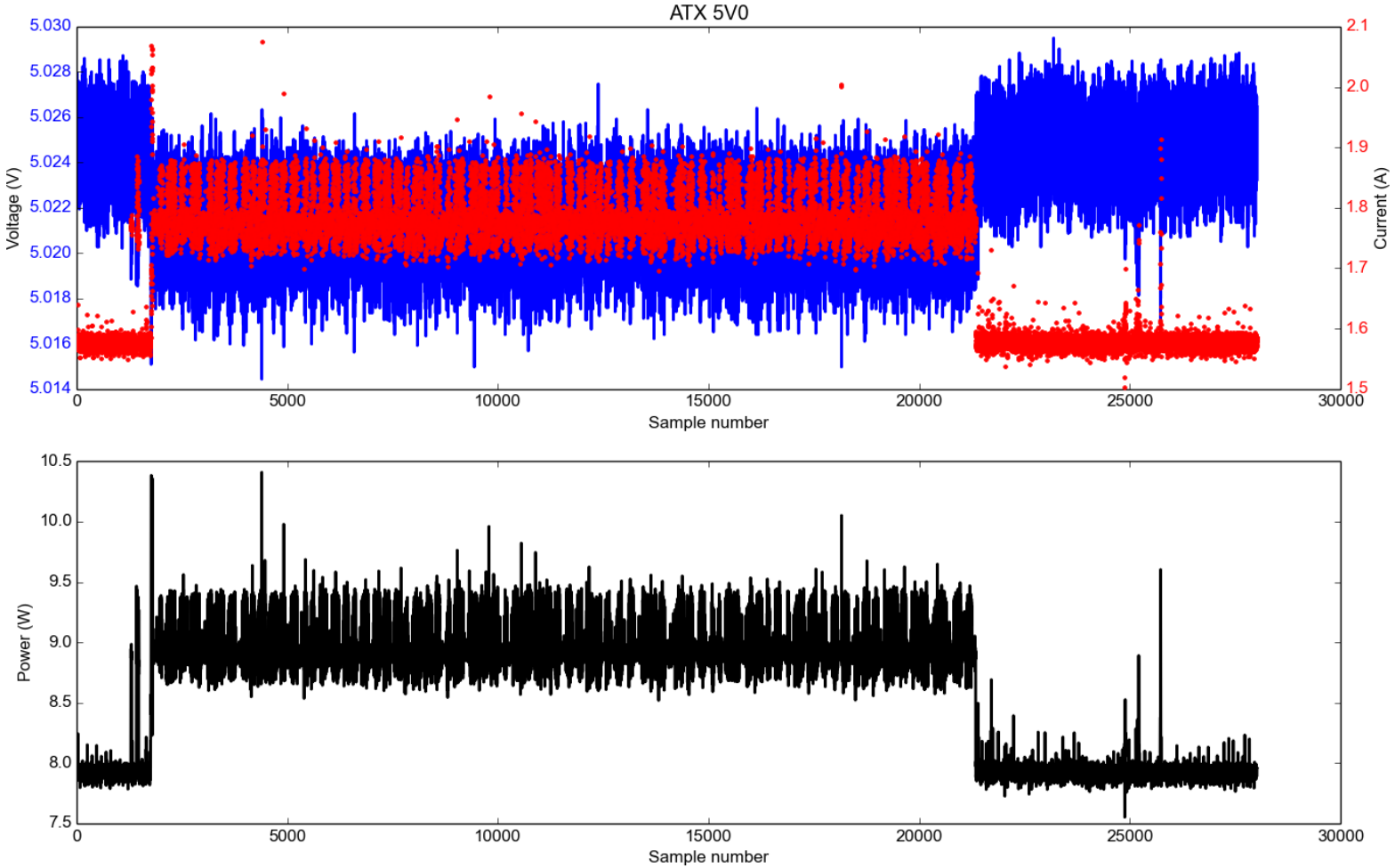

VENGEANCE

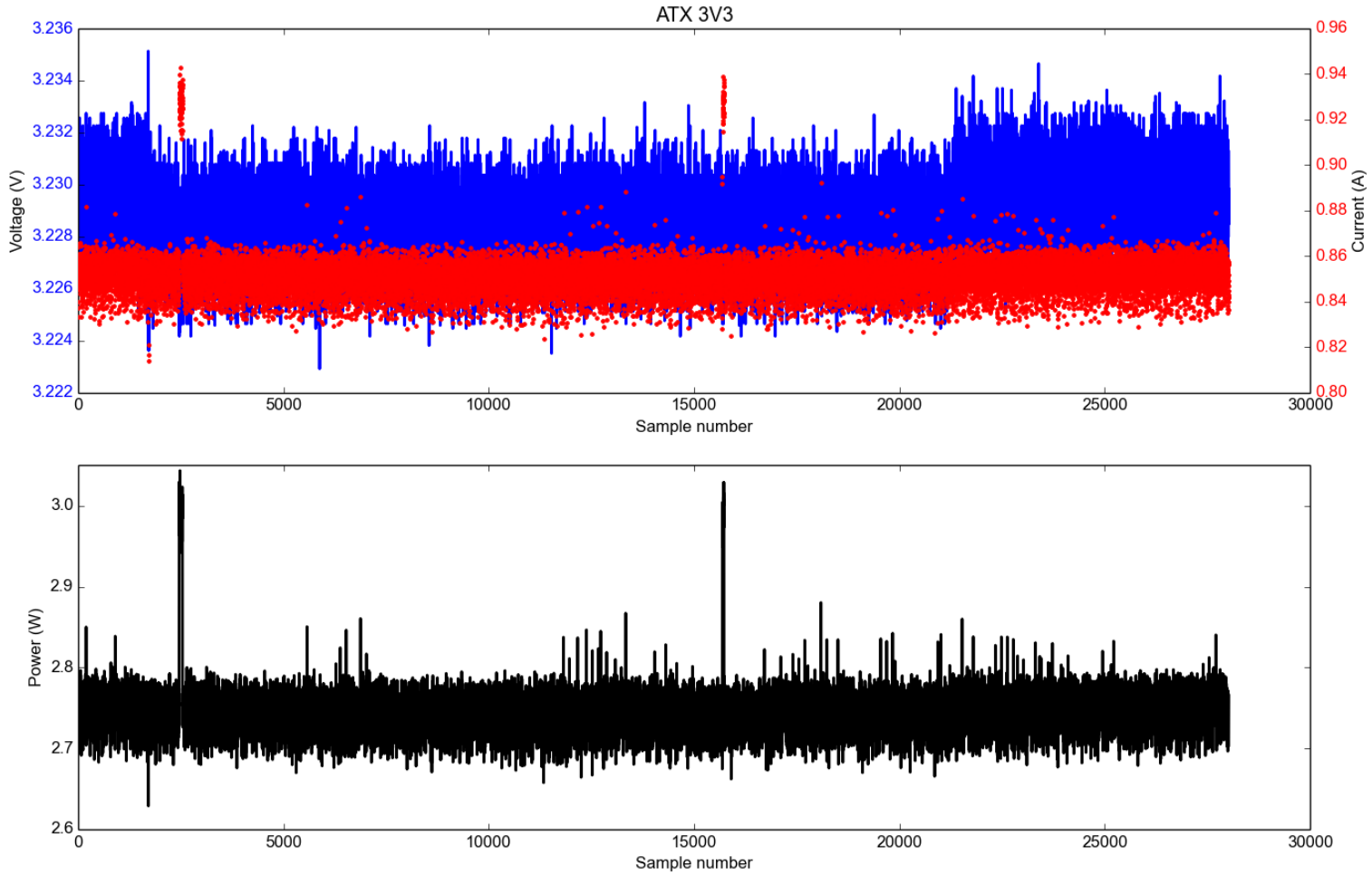
- Intel core i5 4670k
  - Consumer grade 4-core processor
  - No overclocking, modification
  - Roughly 85 GFlops
  - Stock DRAM, SSD, PSU etc.
- A little short of an exaflop!
  - Specifically, 11764705x short
  - We'll use this multiplier to project results to the exascale

- SEISMO is a real-world linear elasticity application written by Nikolay Khoklov.
  - See poster for more details/current work.
- Has been ported to many programming models, OpenCL, MPI, OpenMP, OpenACC, CUDA…
- Will be used here to give a feel for power performance with a *real* workload.


CPU 12V0
Voltage (V)
Current (A)
Sample number
7.8s
Power (W)
Sample number
I/O

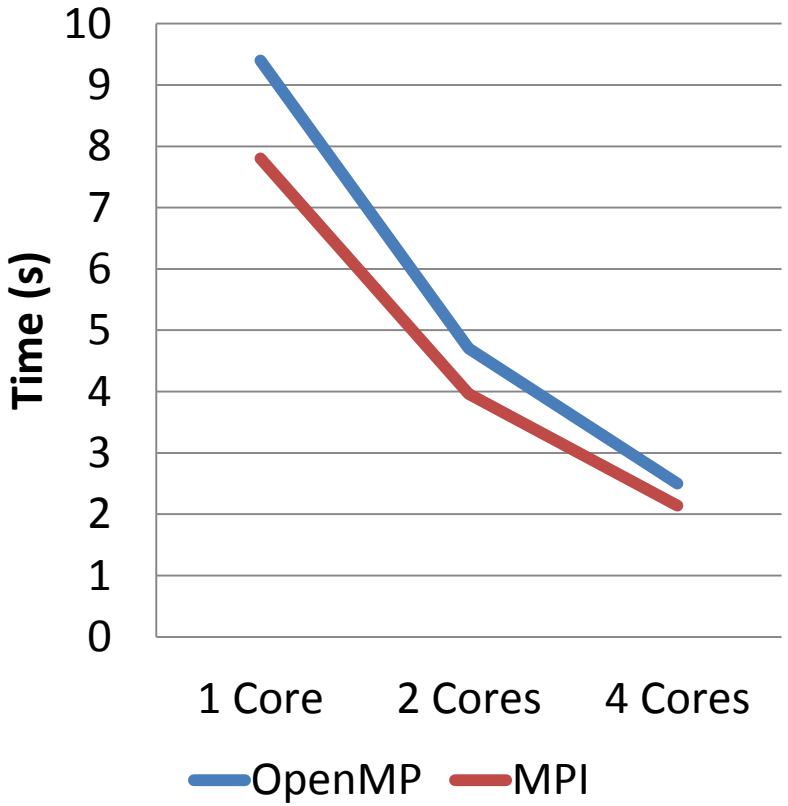

ATX 12V0
Voltage (V)
Current (A)
Sample number
12.200
12.195
12.190
12.185
12.180
12.175
12.170
12.165
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0
5000
10000
15000
20000
25000
30000
Power (W)
Sample number
15
14
13
12
11
10
9
8
7


ATX 5V0
Voltage (V)
Current (A)
Sample number
Power (W)
Sample number


ATX 3V3
Voltage (V)
Current (A)
Sample number
3.236
3.234
3.232
3.230
3.228
3.226
3.224
3.222
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0
5000
10000
15000
20000
25000
30000
Power (W)
3.0
2.9
2.8
2.7
2.6

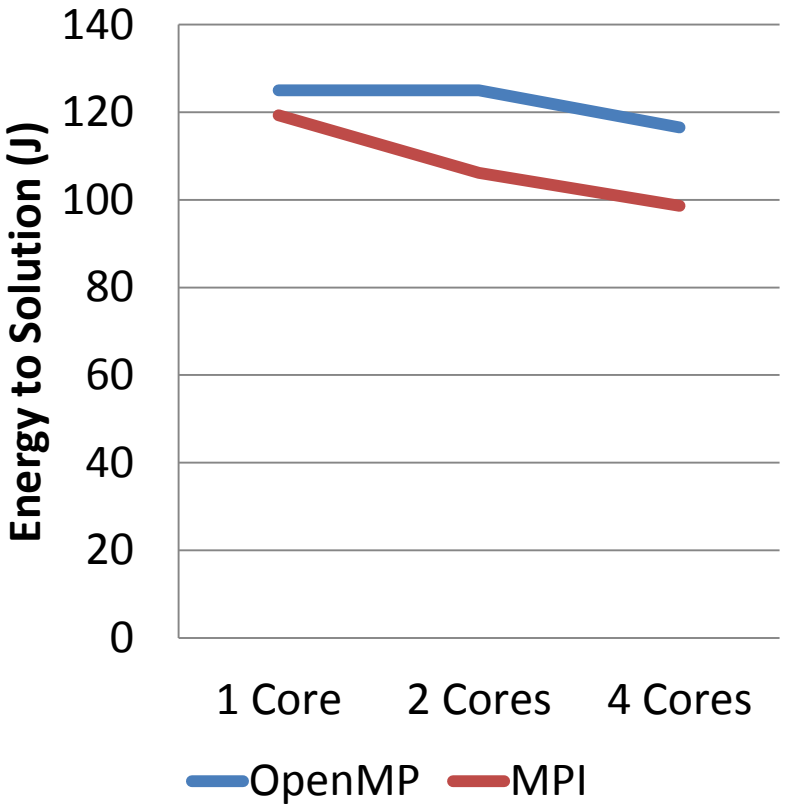
Performance

- MPI scaling:
  - 7.8, 3.96, 2.14s
  - 95% efficiency
- OpenMP scaling:
  - 9.4, 4.7, 2.5s
  - 96% efficiency



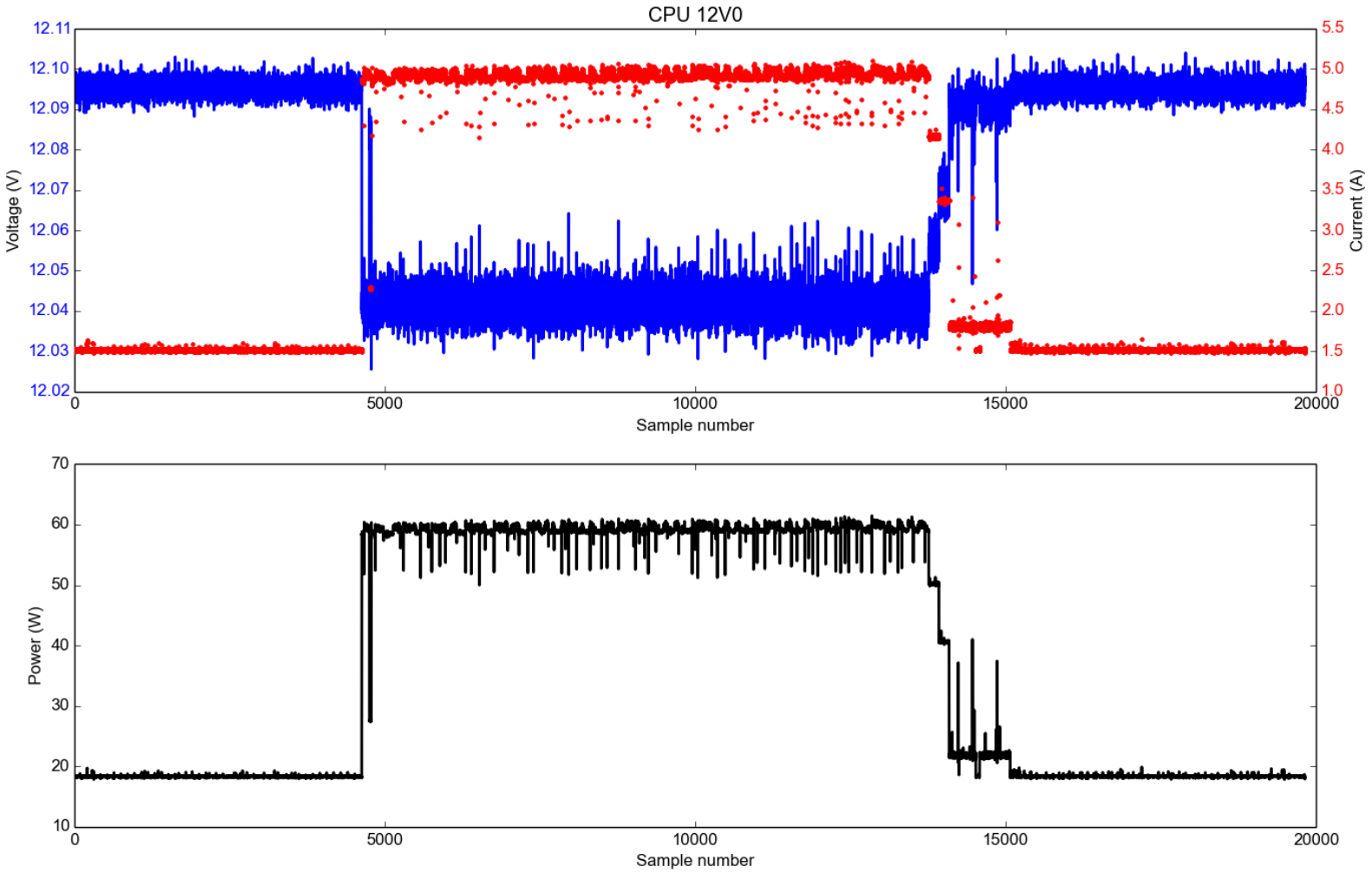
Energy to solution

- MPI scaling
  - 119.34, 106.13, 98.65J
  - Efficiency: 3.3x
  - Ratio of Pact to Pest
- OpenMP Scaling
  - 125.02, 125.02, 116.50J
  - Efficiency: 3.7x

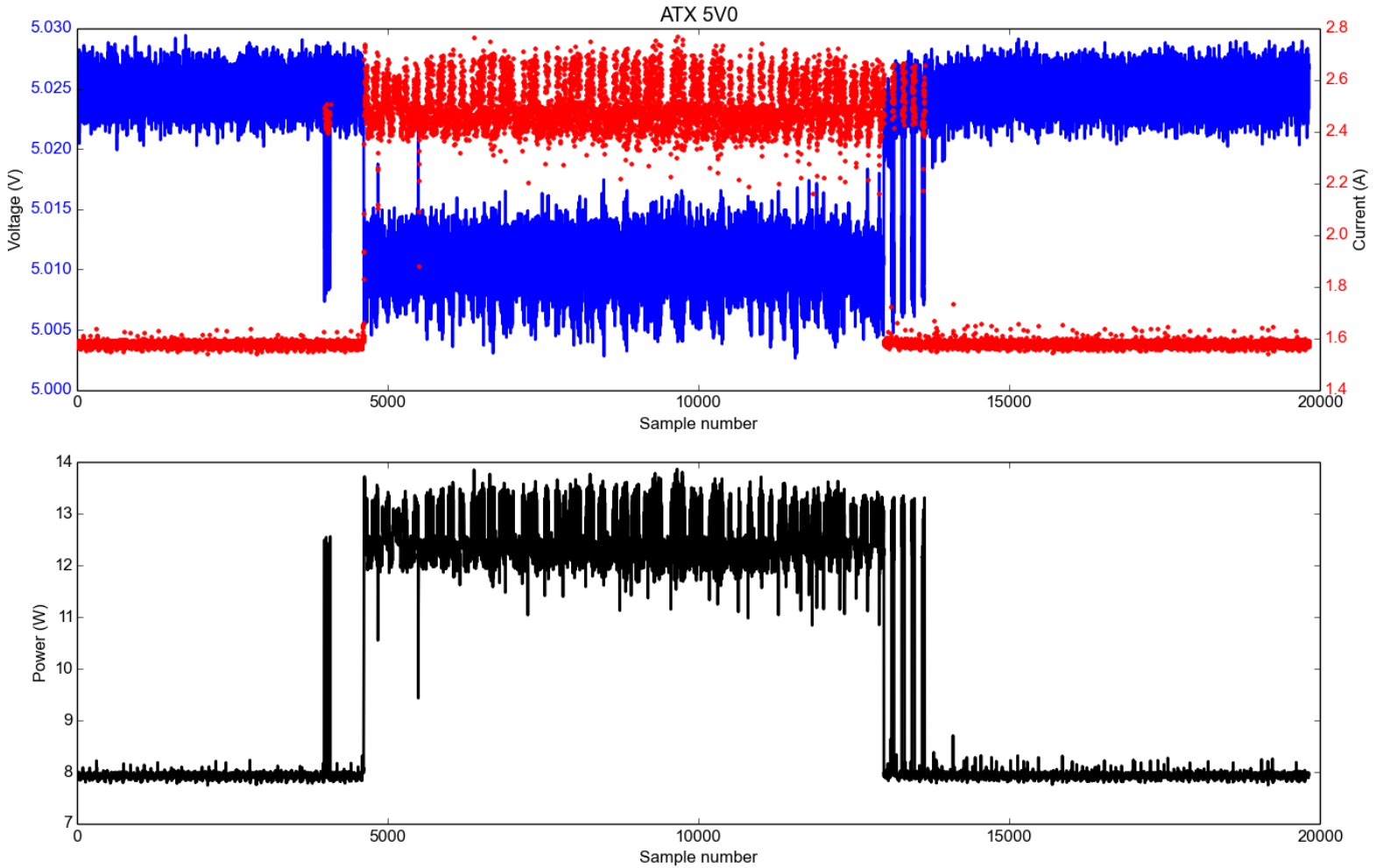


- This implies increasing the number of active cores doesn't improve our situation.
- We need, in both cases, >3x the energy to solution that we'd naively expect from the runtime scaling.
  - Optimizing for performance alone doesn't make sense if we also have to consider energy as a billing metric.
- It appears that the idle (unused) cores are drawing a lot of un-necessary power.
  - It may be other parts of the chip we cannot separate out.
  - At exascale, unused has to mean ~0 power draw.

OMP THREADS = 4



OMP THREADS = 4



- To achieve 1 exaflop on this architecture we need 11,764,706 CPUs.
  - Assuming a horizontal orientation, stacked 1 high, that's an area of 1.7km^2
- Assuming our code would scale to this number of CPUs in parallel, this would be a peak power of 941,176,480W, roughly 941MW, just over 23x the peak power of the upper bound of the target.
- This is just over the output of a CANDU type nuclear reactor.
  - And we've not touched disk, cooling etc.

Summary

- What can we gain from power measurement?
- It's clear that the power is quite distinct in phases, we can see the main computation phase, followed by a short output phase where CPU power drops.
  - High speed measurement is required to observe this behaviour.
- Power-aware scheduling will be required to exploit this.

References

1. http://science.energy.gov/ascr/research/scidac/exascale-challenges/
2. http://eehpcwg.lbl.gov/

www.adept-project.eu