

Total-Power Usage Effectiveness on SuperMUC

A feasibility study...



Herbert Huber, Axel Auweter
Leibniz Supercomputing Centre



Bayerische
Akademie der Wissenschaften

SuperMUC



TOP 500[®]
SUPERCOMPUTER SITES

2897 TFlop/s HPL Performance

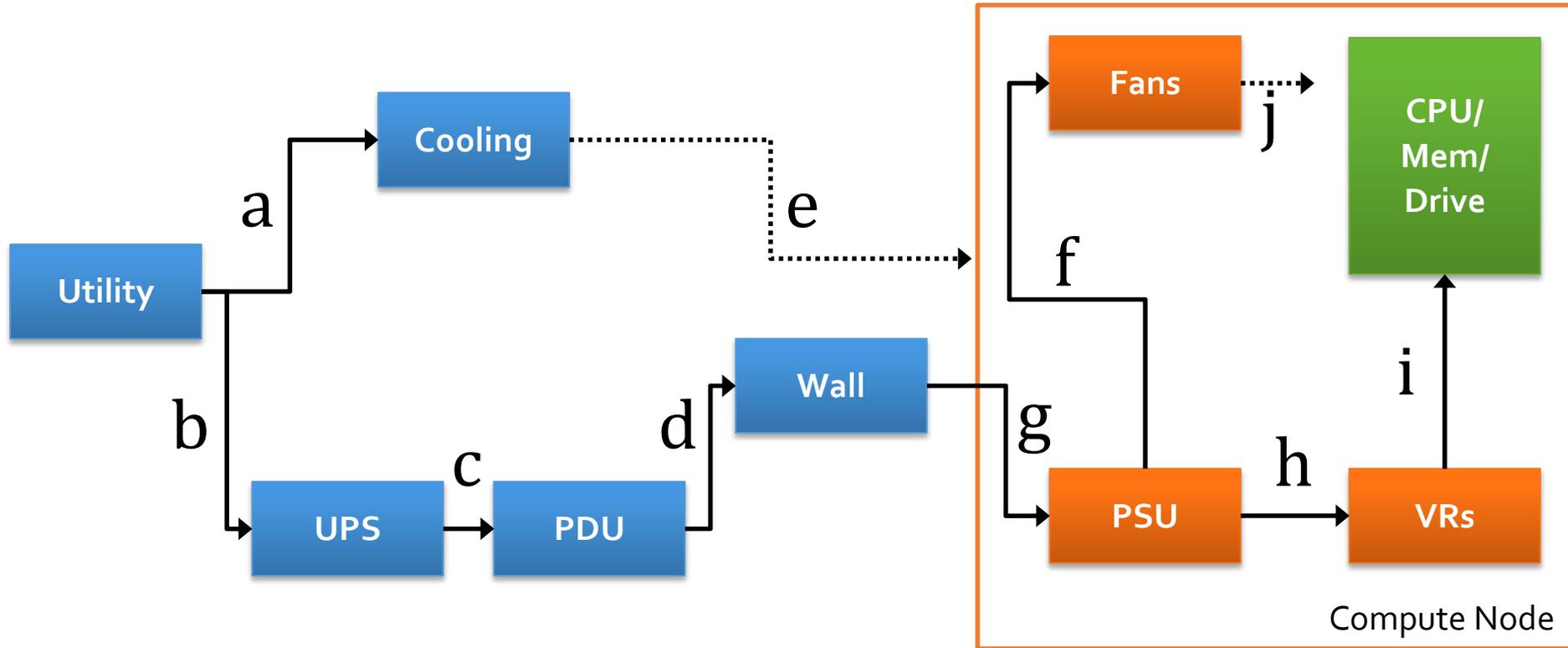
Rank 4 (June 2012)

Rank 6 (November 2012)

Rank 9 (June 2013)

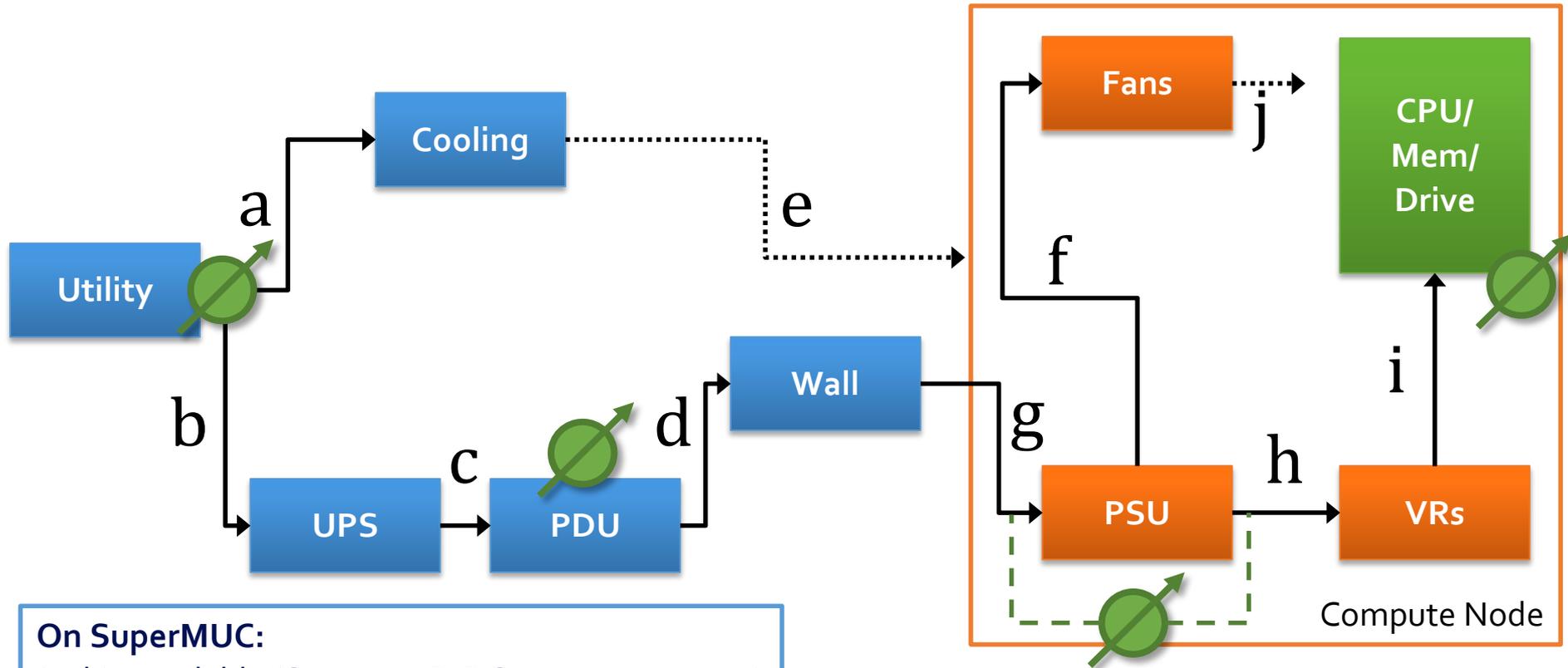
Rank 10 (November 2013)

TUE Power Measurements



$$TUE = \frac{a + b}{i}$$

SuperMUC Power Measurements



On SuperMUC:

(a+b): Available (Socomec DIRIS A40/A41 meters)

(c/d): Available (IBM PDU 46M4004)

(g): Available (IBM Paddle Card AC Side)

(h): Available (IBM Paddle Card DC Side)

(i): Estimation Available (Intel® RAPL)

$$TUE = \frac{a + b}{i}$$

SuperMUC Power Measurements

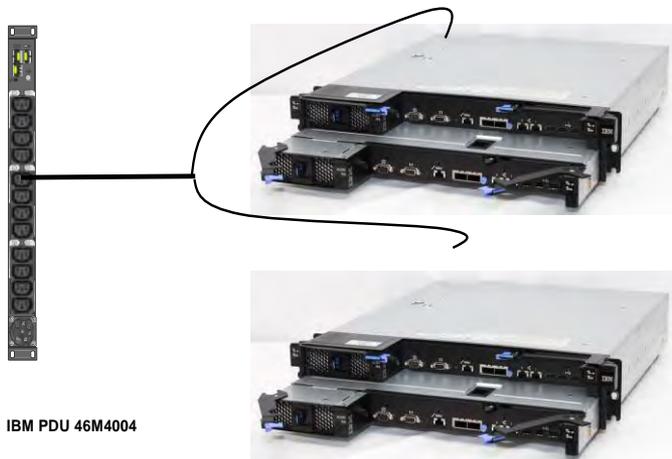


Air cooled iDataPlex nodes



IBM PDU 46M4004

Warm water cooled iDataPlex nodes



IBM PDU 46M4004

Stream Benchmark

	Node 1	Node 2	Node 3
RAPL PKG	155,7	157,4	167,6
RAPL DRAM Power	46,2	42,9	86,3
Paddle Card DC Power	368	363	403

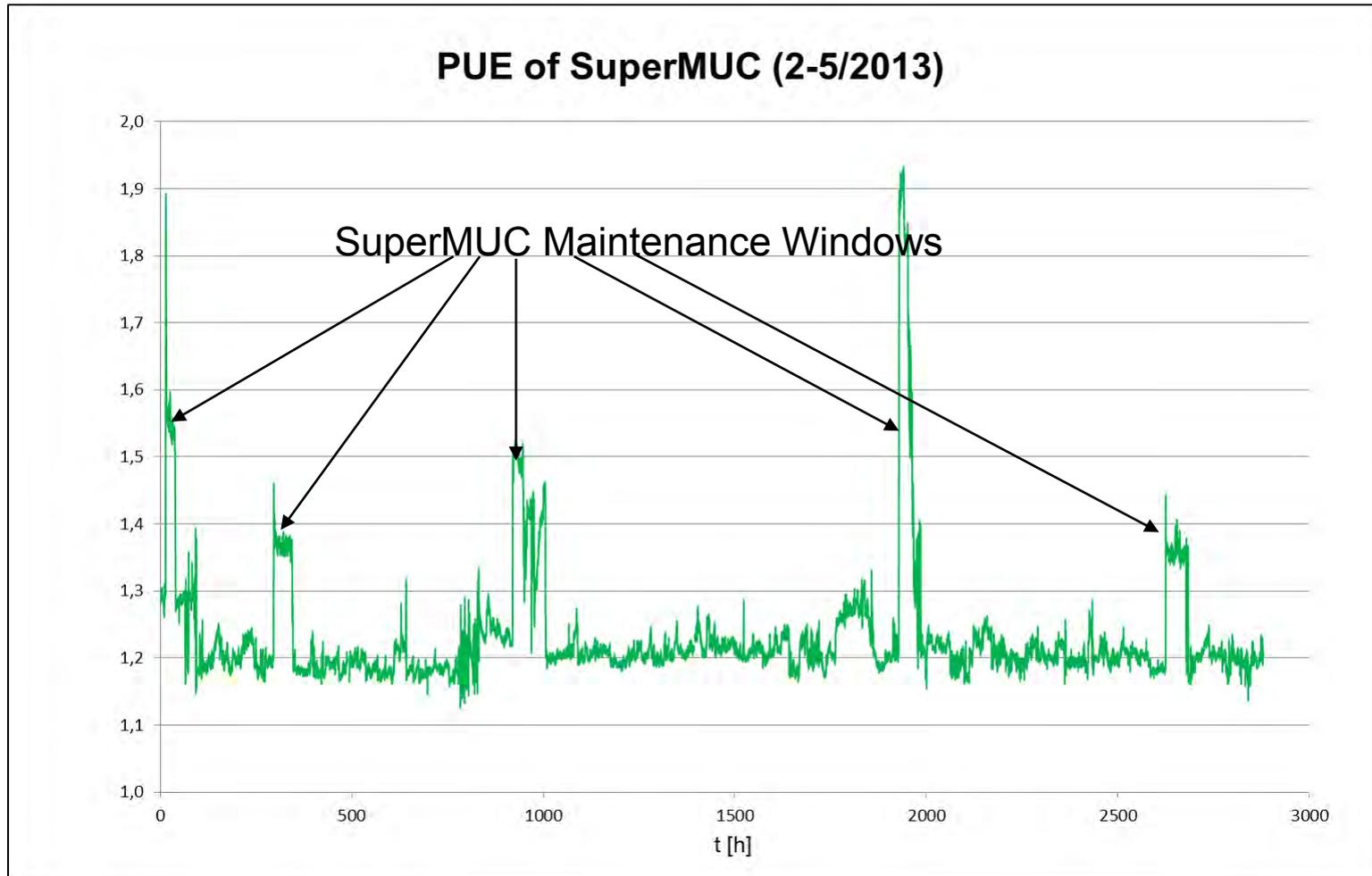
- Node 1 and Node 2 are equipped with **8 16 GB DIMMs**
- Node 3 is equipped with **16 8 GB DIMMs**

Stream Benchmark

	Cnode1	Cnode2	Cnode3	Cnode4
RAPL PKG	165,0	164,6	163,6	131,5
Paddle Card DC Power	253	255	254	225

- Cnodes are diskless and have less memory installed (**8 4 GB DIMMs**)

Measured PUE of SuperMUC



TUE Considerations @ LRZ



- SuperMUC is (warm-)water cooled
 - Infrastructure overhead is almost constant (15 kW) due to pumps instead of temperature controlled fans

 - Weaknesses of PUE indeed affect SuperMUC
 - Energy-Aware Scheduling lowers (IT-)power consumption while infrastructure power consumption remains constant
 - The compute nodes suffer from only little infrastructure overhead (no CPU fans, high-efficiency power supplies, etc.)
- PUE is not an all-encompassing metric for Energy Efficiency!



Summary



- TUE metric addresses intra-node overheads – very good!
- However, TUE is unlikely to give us additional insight on SuperMUC:
 - One fan per node (in the power supply) – **overhead is known**
 - Power supply conversion loss was part of client specs during procurement and has been verified – **overhead is known**
 - VR loss is within the scope of IBM – **overhead is unknown**
- Measuring TUE on SuperMUC is feasible
- Actual TUE measurements on SuperMUC are not so easy:
 - Lots of un-calibrated measurement data to deal with (Paddle cards, RAPL counters)
 - Paddle cards and RAPL counters only deliver actual power values