The IBM-ASTRON 64bit µServer demonstrator for SKA

Ronald P. Luijten – Data Motion Architect
lui@zurich.ibm.com
IBM Research - Zurich
14 March 2014

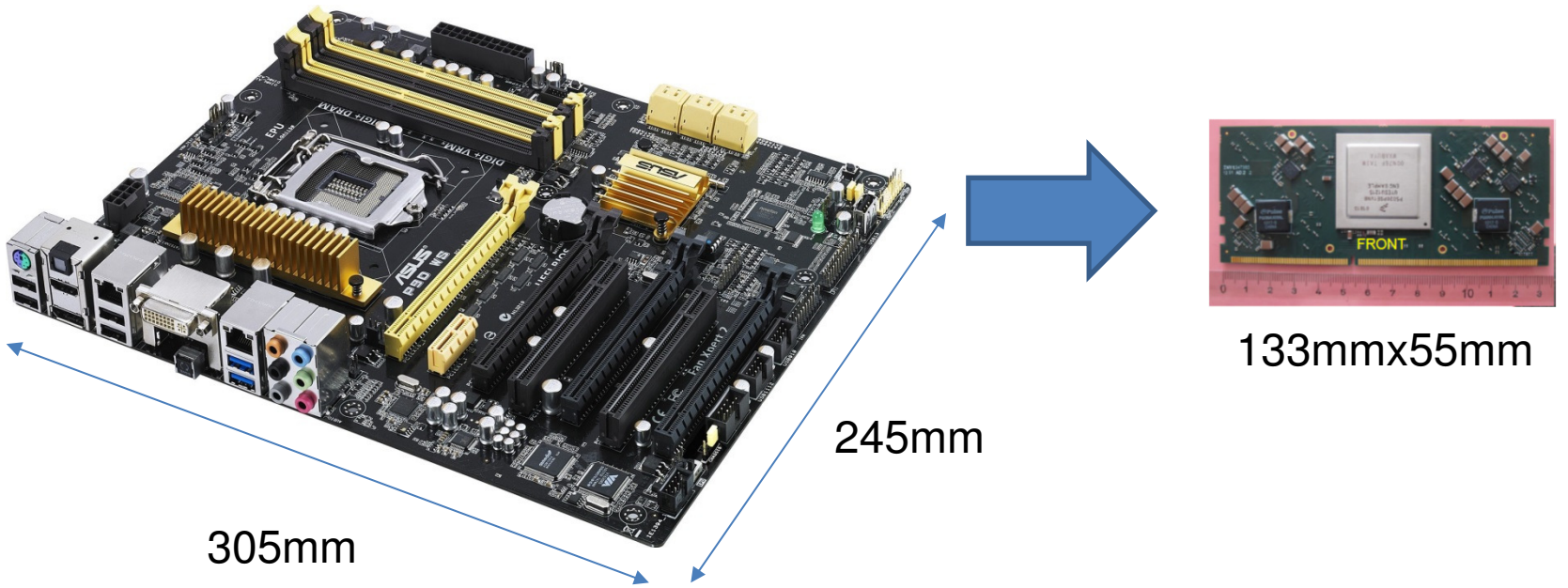
DISCLAIMER: This presentation is entirely Ronald's view and not necessarily that of IBM.

Rules

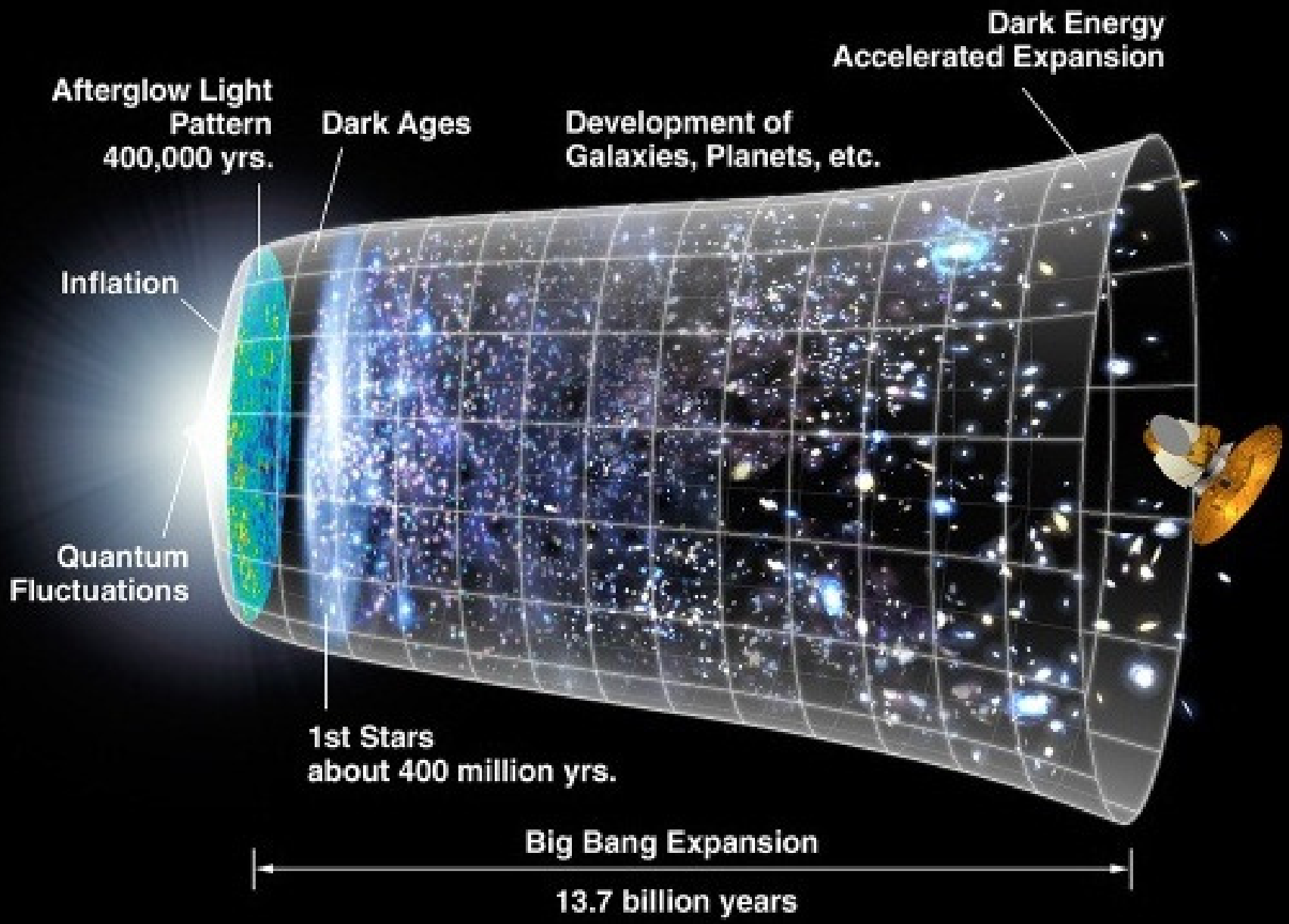
Definition

µServer:

The integration of the entire server node motherboard (no graphics) into a single microchip except DRAM, Nor-boot flash and power conversion logic.

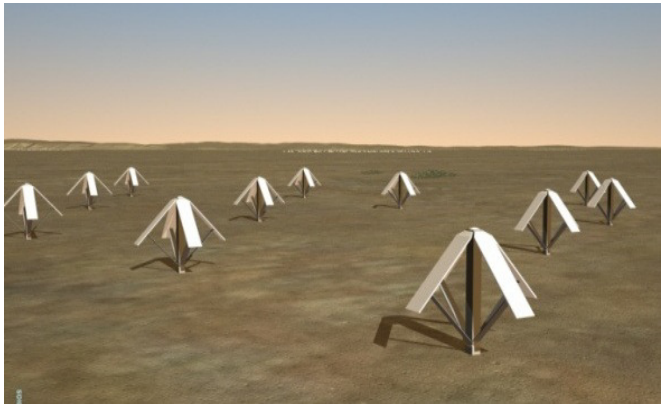


SKA (Square Kilometer Array) to measure Big Bang



SKA: Largest Radio-astronomy antenna

Big data on Steroids



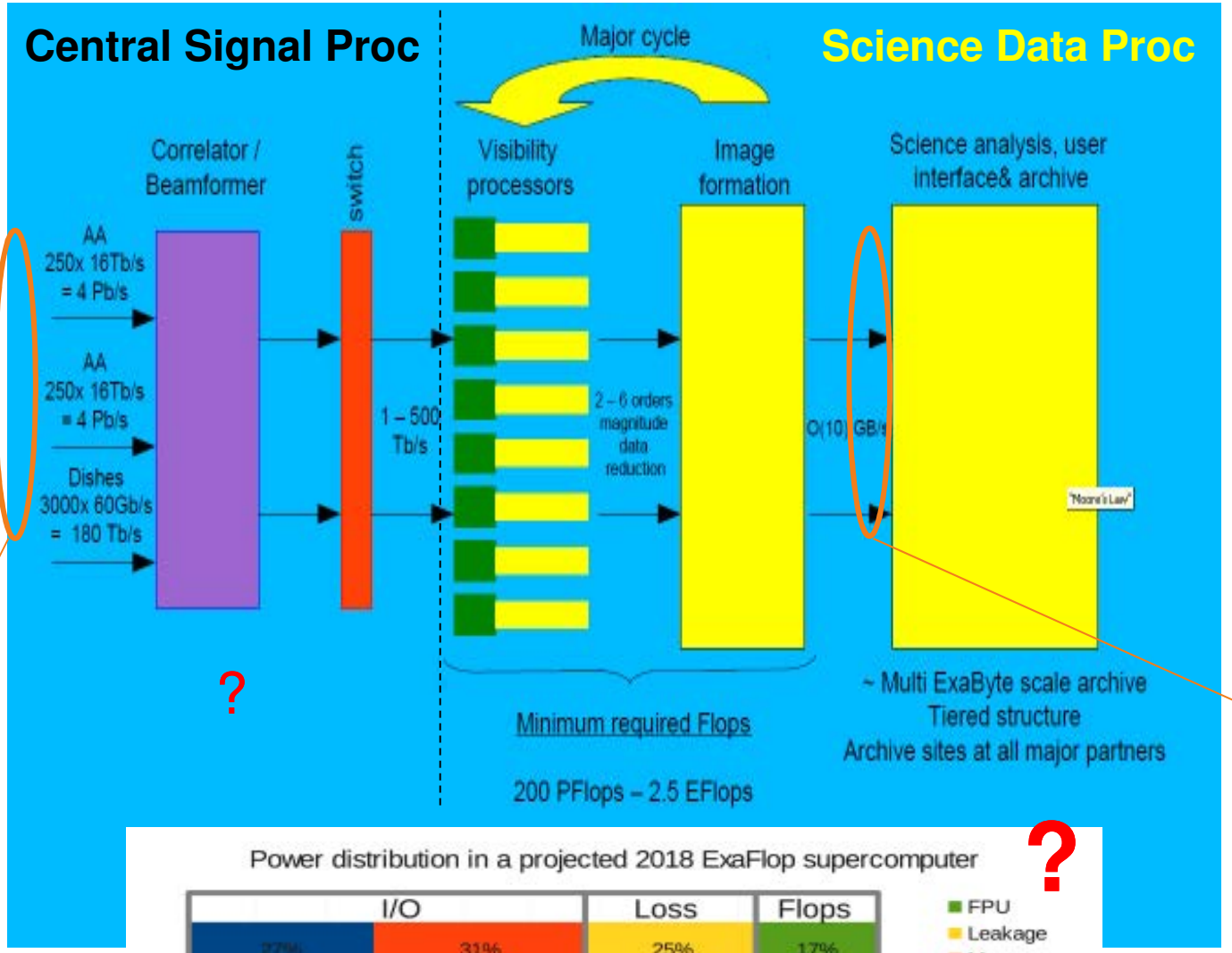
Up to 2 Million+ Antenna's
What does this mean?

Prelim. Spec. SKA, R.T. Schilizzi et al. 2007 / Chr. Broekema

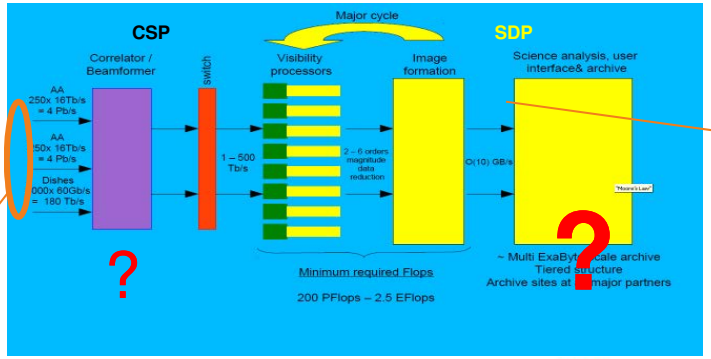
~ 10 Pb/s

86'400 sec/day

10..14 ExaByte/day



~ 1 PB/Day.



~ 10 Pb/s

86'400 sec/day

10..14 ExaByte/day

~ 1 PB/Day.

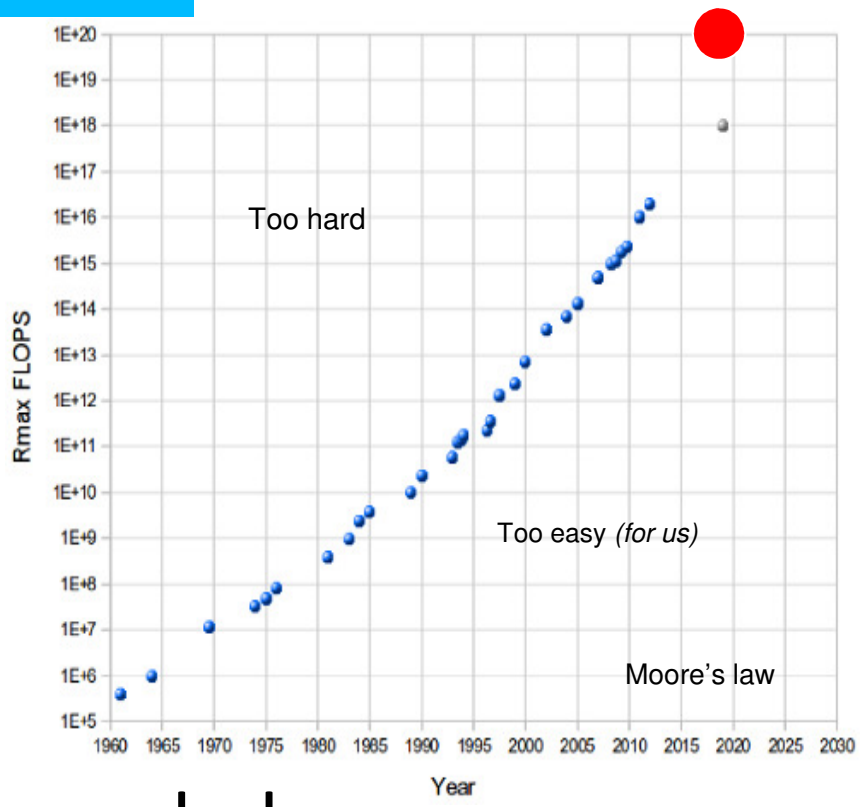
330 disks/day

120'000 disks/yr

Top-500 Supercomputing(11/2013)…. 0.3Watt/Gflop/s
→Today's industry focus is 1 Eflop @ 20MW. (2018)
➔(0.02 Gflop/s)

➔Most recent data from SKA:
- ➔CSP….max. power 7.5MW
- ➔SDP….max. power 1 MW
- ➔Latest need for SKA – 4 Exaflop (SKA1 - Mid)
- ➔ 1.2GW…80MW

Factor 80-1200



→multiple breakthroughs needed

IBM / ASTRON DOME project
Technology roadmap development

•Sustainable (Green) Computing

•Nanophotonics

•Data & Streaming

•User Platform

•System Analysis

•Algorithms & Machines

-Student projects
-Events
-Research Collaboration

•Computing

-Microservers
-Accelerators

•Transport

-Nanophotonics
-Real Time Communications
-Compressive Sampling

•Storage

-Access Patterns

IBM DOME µServer Motivation & Objectives

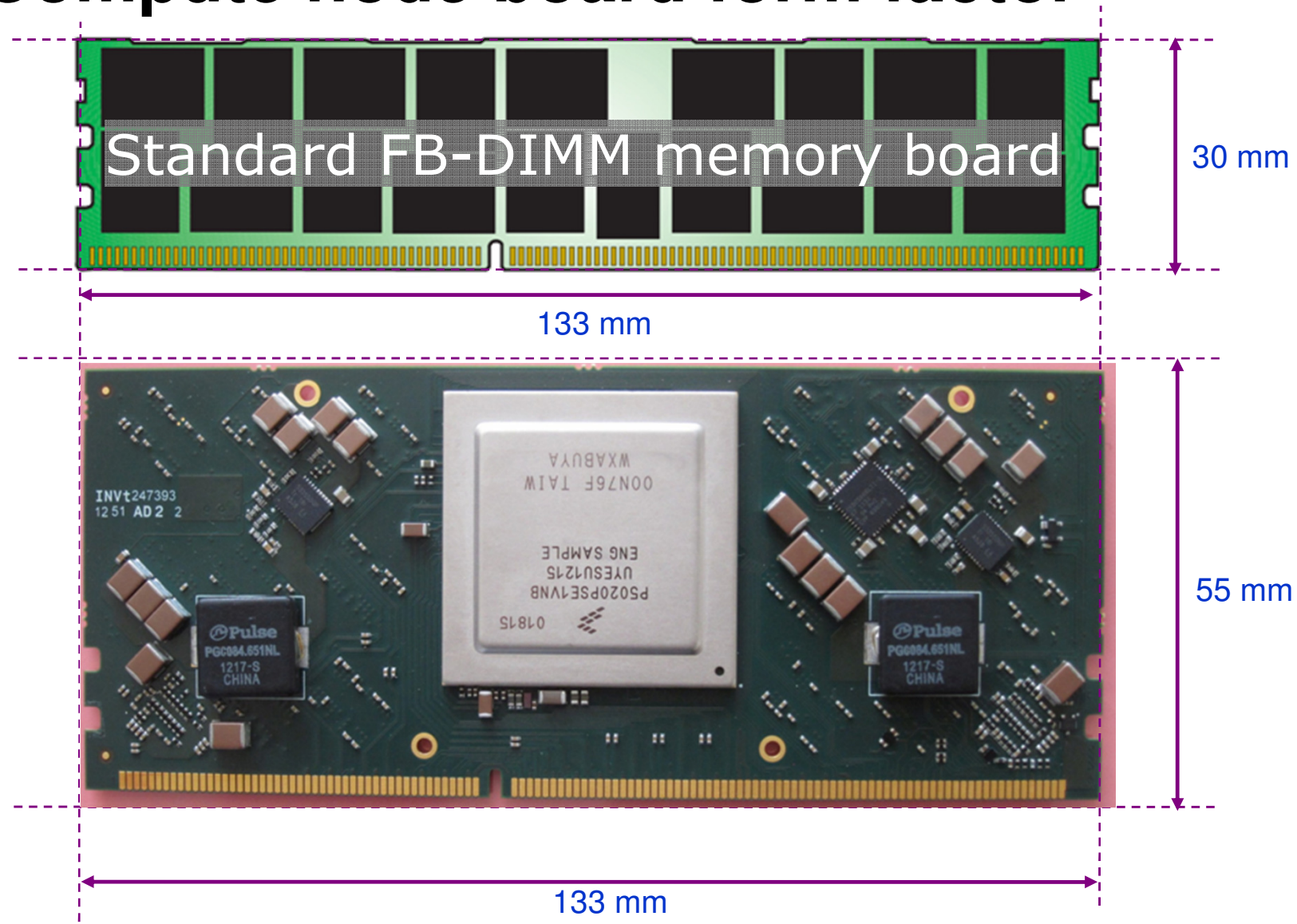
- **Create *the* worlds highest density 64 bit µ-server drawer**
  - Useful for both SKA radio-astronomy and IBM future business
    - Platform for Business Analytics appliance pre-product research
    - "Datacenter in-a-box"
  - Very high energy efficiency / very low cost (radioastroners…)
  - Use commodity components only, HW + SW standards
  - Leverage 'free computing' paradigm
  - Enhance with 'Value Add': packaging, system integration, …
  - speed of light

- **Most efficient cooling using IBM technology (ref: SuperMUC TOP500 machine)**

- **Must be true 64 bit to enable business applications**
- Must run server class OS (SLES11 or RHEL6, or equivalent)
  - Currently precludes ARM (64-bit Silicon not yet available)
  - PPC64 is available in SoC from FSL now
  - (I am poor – no $$$ for my own SoC…)

- **This is a research project – capability demonstrator only**

Compute node board form factor



Compute node processor options

FSL SoC parts	P5040	T4240
CPU GHz	2.2	1.8
CPUs	4 cores, 1 thread per core	12 cores, 2 threads per core
Primary cache	32 KB I + 32 KB D per core	32 KB I + 32 KB D per core
Secondary cache	512 KB I+D	2 MB per 4 CPUs
L3 cache	1 MB on chip	1.5 MB on chip
Memory	2 x 2 GB, DDR3/L3, ECC	3 x 2 GB, DDR3/L3, ECC
core	e5500, ppc64	e6500, ppc64
	1 DP FP unit per core	1 DP FP unit per core 128 bit SP altivec unit per core
node	45nm	28nm
TDP	55W	60W

T4240 DIMM connector:

- 2 times SATA
- 4 times 10 Gigabit ethernet
- SD card interface
- USB interface
- Some power supplies

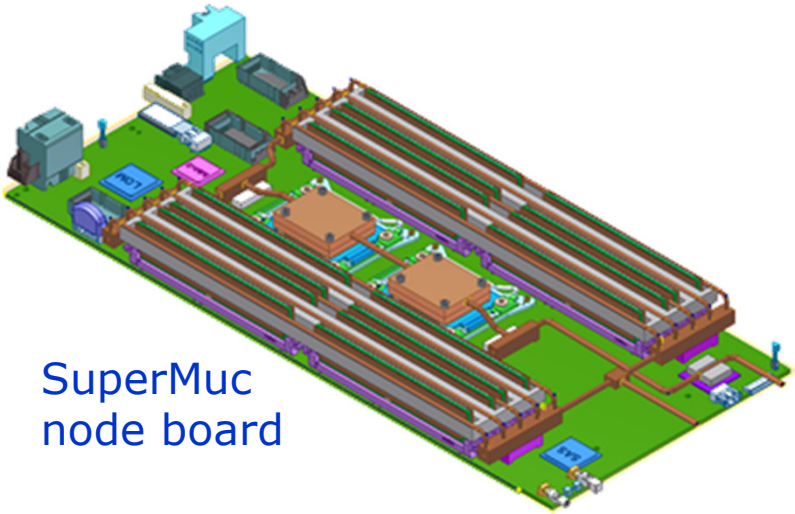
Hot Water Cooling

Most Energy Efficient solution:

- Low PUE possible (<=1.1) – Green IT
- 40% less energy consumption compared to air-cooled systems
- 90% of waste heat can be reused (CO_2 neutral according Kyoto protocol)
- Allows very high density
- Less thermal cycling - improved reliability
- Lower T_j reduces leakage current – further saving energy

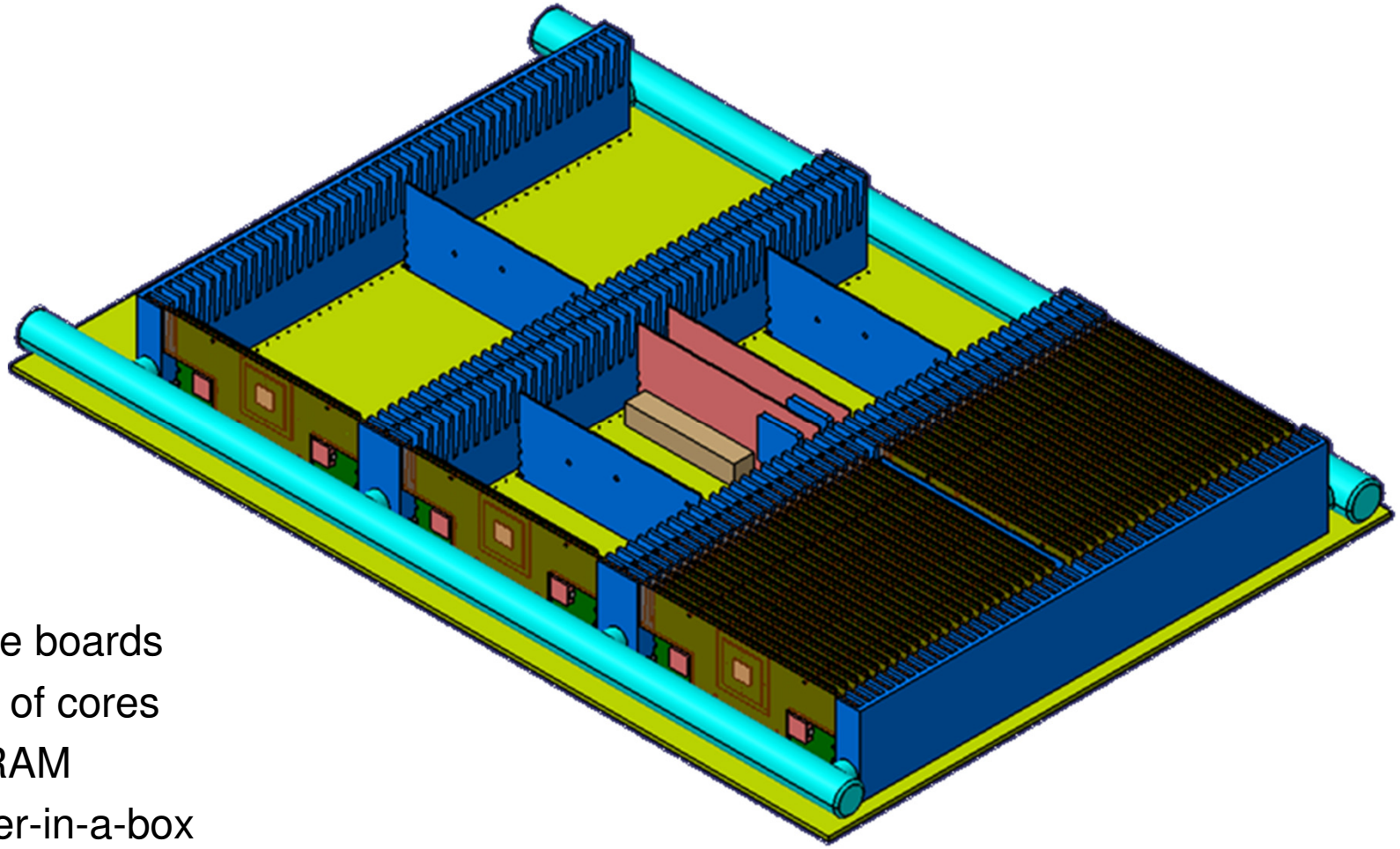
SuperMUC HPC machine at LRZ in Germany demonstrates ZRL hot water cooling

- No 4 on June 2012 TOP500 HPC list



SuperMuc
node board

19” 2U Chassis with Combined Cooling and Power

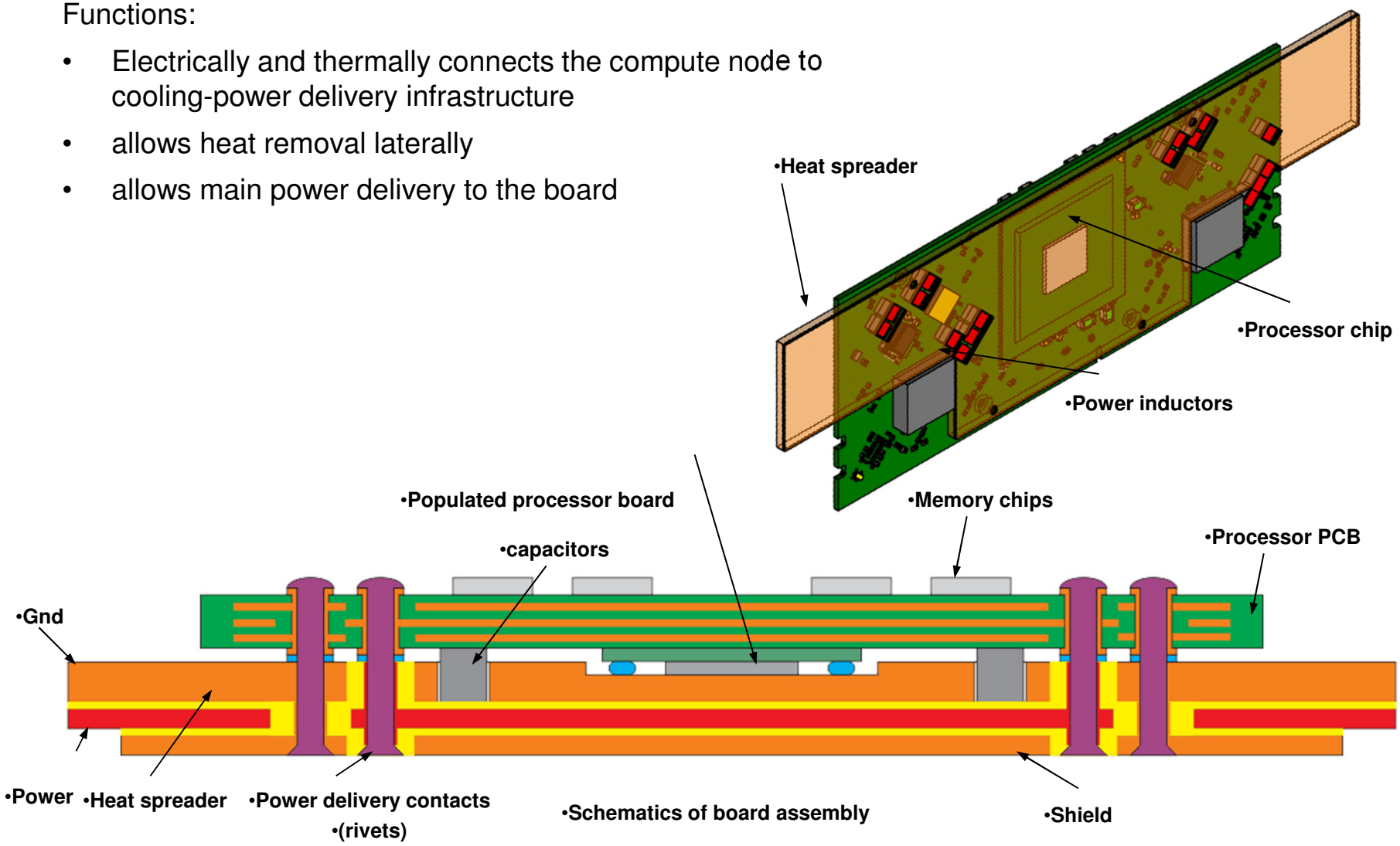


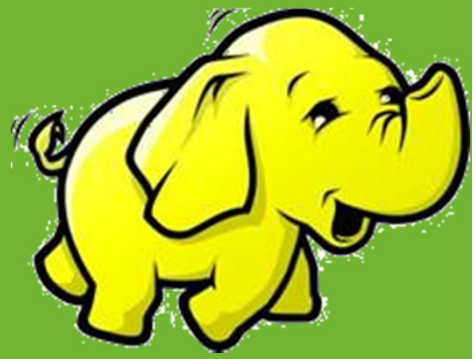
~100 node boards
hundreds of cores
~2 TB DRAM
Datacenter-in-a-box

Compute node heat spreader

Functions:

- Electrically and thermally connects the compute node to cooling-power delivery infrastructure
- allows heat removal laterally
- allows main power delivery to the board



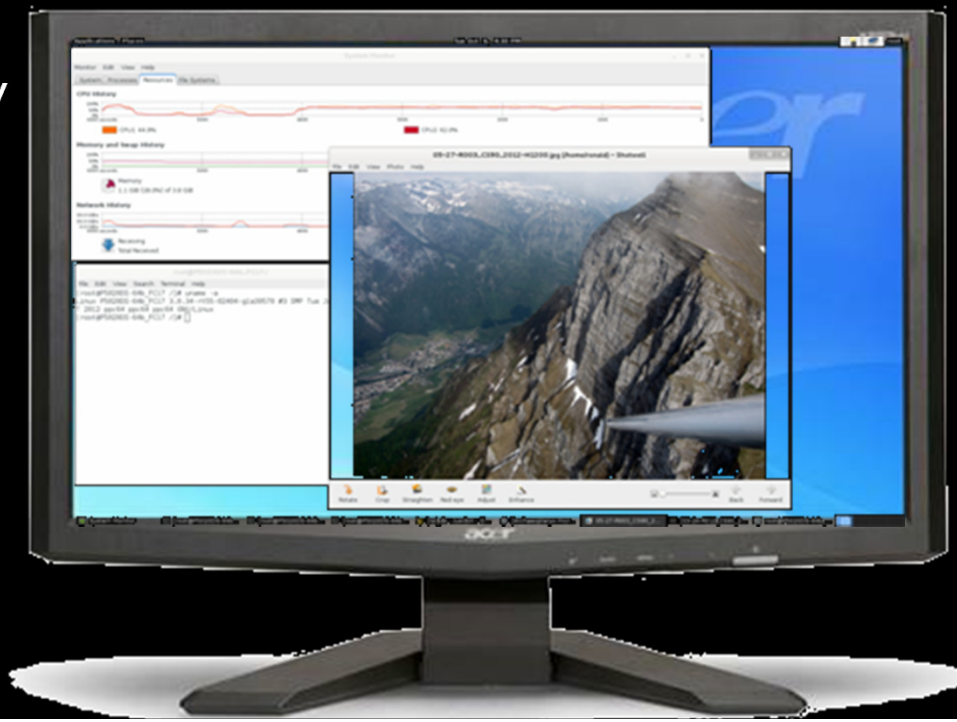


And now the Software story…

64 bit Fedora 17 on P5020DS

- Freescale took kernel version 3.0.34 from kernel.org
- Configured and compiled it for P5020
- Took Fedora user space root FS (thru another PPC platform)
- Runs 100% OK - YUM, Gnome desktop, networking, apache, etc…
  - System up and running > 40 days
  - Java, Python, …

- This effort took approximately ONE day



IBM DB2 installation on P5020

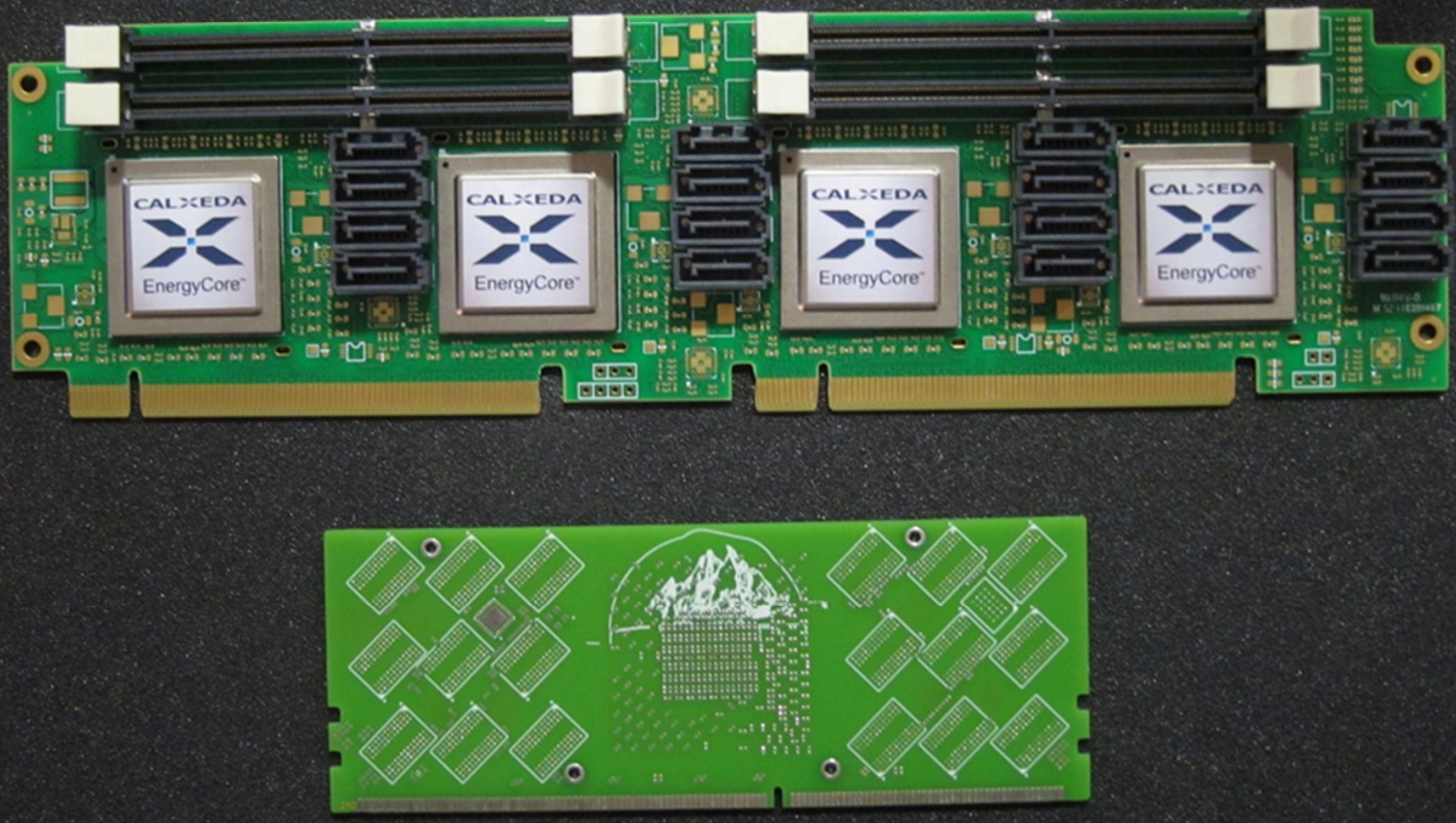
- Simple install of IBM XL C/C++ runtime (XLC compiler runtime)
- Install libaio
- Simple install of IBM DB2 (express-C, v10.1)
- Some minor configuration adjustments required
- Entire process only took a few hours -- *no compilation was needed*
- Demo available
  - Technology explorer (runs php in browser)
  - WMD Workload Multi-User Driver (Java based)
  - DB2 data base engine

- Runs stable – able to exercise without any issues

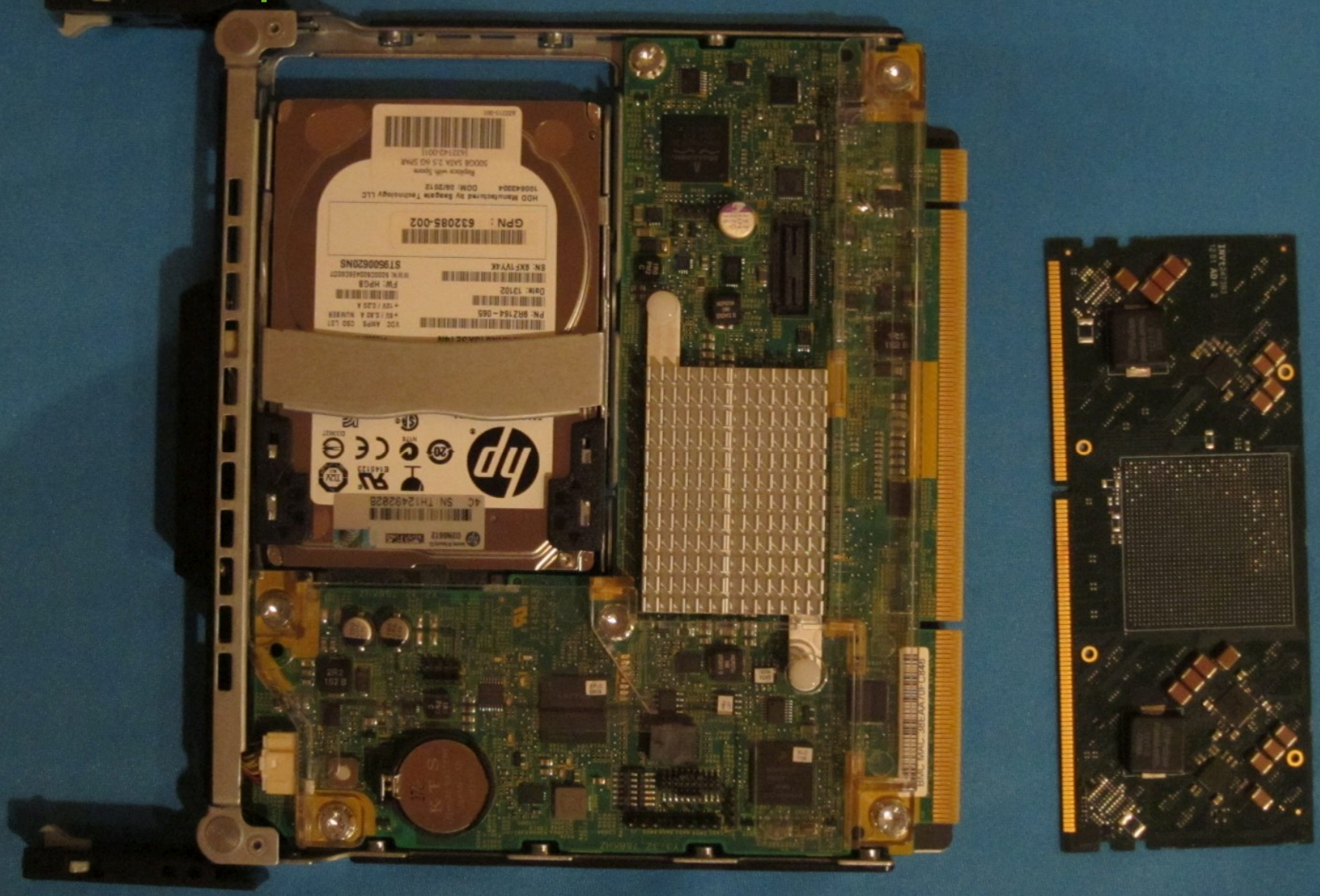
Hadoop install on P5020

- Simple install (version 1.0.3 for ppc64)
- Minor configuration effort required
- Works for single node and pseudo-distributed mode
- No compilation necessary
- Demo available

Comparison to Calxeda node board



Comparison to Moonshot node board




SKA IN AFRICA
The Department of Science and Technology welcomes you to
CARNARVON, home of the Square Kilometre Array
www.dst.gov.za
www.nrf.ac.za
www.ska.ac.za
www.skatelescope.org
science & technology
Department: Science and Technology
REPUBLIC OF SOUTH AFRICA
NRF
National Research Foundation
SKA AFRICA
Northern Cape Government
Pixley Ka Seme District Municipality
Kareeberg Municipality
SKA
www.skatelescope.org


Z A
CA 26
4071
MELVILL & MOON

Acknowledgements

This work is the results of many *people*

- Peter v. Ackeren, FSL
- Ed Swarthout, FSL Austin
- Dac Pham, FSL Austin
- Yvonne Chan, IBM Toronto
- Andreas Doering, IBM ZRL
- Tom Wilson, IBM Armonk
- Alessandro Curioni, IBM ZRL
- Stephan Paredes, IBM ZRL
- James Nigel, FSL
- Gary Streber, FSL
- Patricia Sagmeister, IBM ZRL
- Boris Bialek, IBM Toronto
- Marco de Vos, Astron NL
- Hillery Hunter, IBM WRL
- Vipin Patel, IBM Fishkill
- And many more remain unnamed….

Companies: FSL Austin, Belgium & Germany; IBM worldwide; Transfer - NL

Questions???

PS. I like lightweight things
µServer website: www.swissdutch.ch