IBM high-density µServer demonstration platform leveraging PPC, Linux and hot-water cooling

Ronald P. Luijten – Data Motion architect

CELEBRATING 21 YEARS OF POWER ARCHITECTURE ANNIVERSARY

ASIA
POWER
ARCHITECTURE
CONFERENCE

http://www.swissdutch.ch:6999/

SHANGHAI, CHINA OCTOBER 25, 2012

DOME – Research Phase for SKA (SKA = Square Kilometer Array)

The SKA will be the largest and most sensitive radio telescope ever built.

A single instrument with >10'000's of antennas will become operational in 2024 with frequency ranges 70MHz to 10GHz. This will generate huge amounts of data, which need to be *transported, analyzed, stored and retrieved* – at *very low* power and *very low* cost.

A true Exascale Analytics Challenge!

DOME is a research phase project before start of SKA deployment in 2017

- 5 year collaboration between ASTRON (NL) and IBM, started Feb 2012
- Co-funded by Dutch government and IBM
- Multi project program including high scale-out and scale-in micro server project

IBM DOME µServer Motivation & Objectives

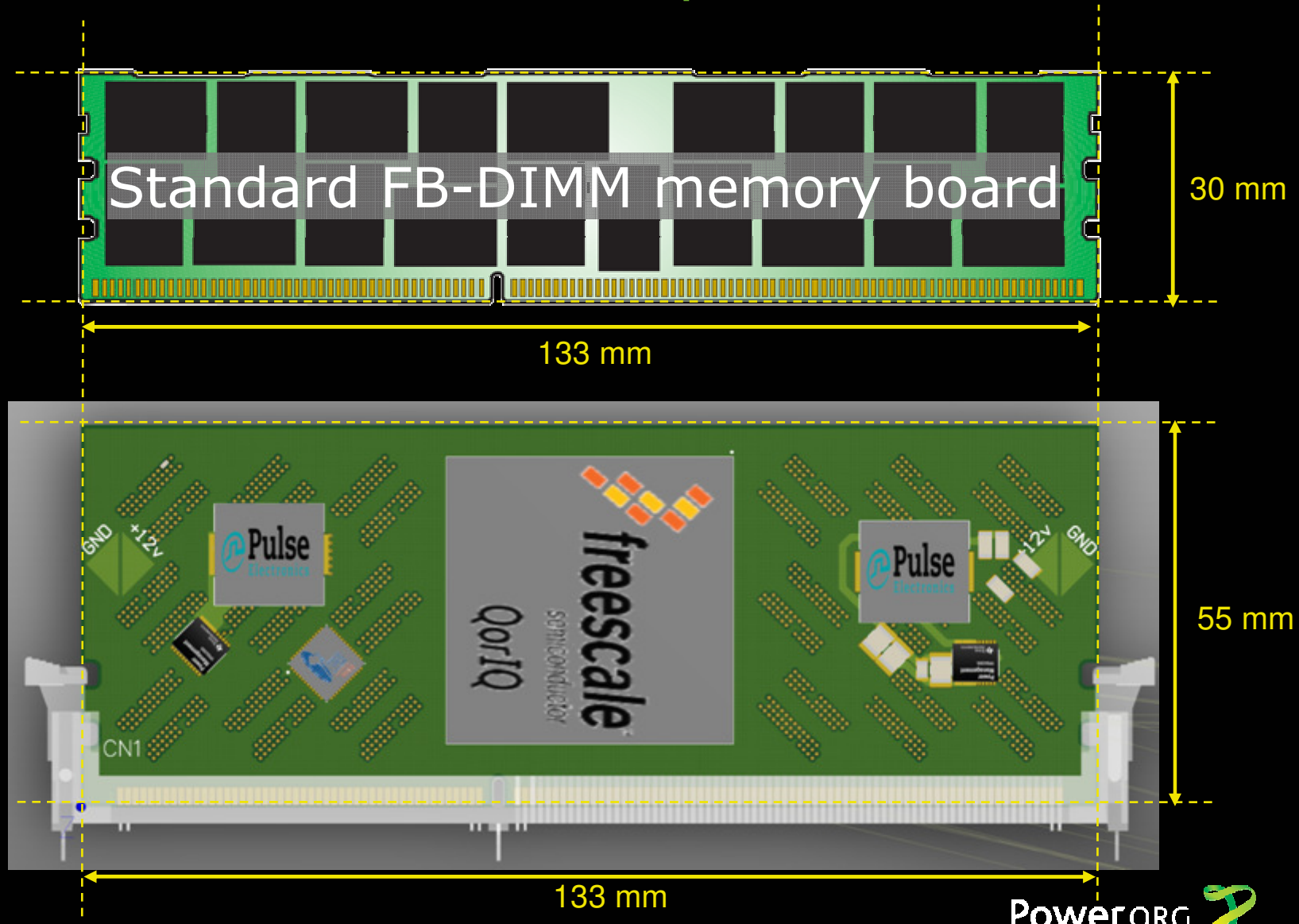
Target density in 2U 19":
-More than 100 nodes
-More than 500 cores
-Around 2TB memory
-More than 250 Gbps network BW out of shelf

- **Create *the* worlds highest density 64 bit µ-server drawer**
  - Useful for both SKA radioastronomy and IBM future business
  - Very high energy efficiency
- **Most efficient cooling using IBM technology (ref: SuperMUC TOP500 pos #4)**
- Platform for Business Analytics appliance pre-product research
- "Datacenter in-a-box"

- Must be true 64 bit to enable business applications
  - Currently precludes ARM (currently no 64-bit Silicon available)
  - PPC64 is most compelling based on ecosystem compatibility
- Must run server class OS (SLES11 or RHEL6, or equivalent)
- Must use commodity components only, HW standards, standard SW based
- Must be a true microserver (IBM ZRL definition):
  - integrates the entire compute node motherboard, except DRAM and NOR-boot flash
  - Must integrate Ethernet on 'microserver' SOC.
- **This is a research project – capability demonstrator only**

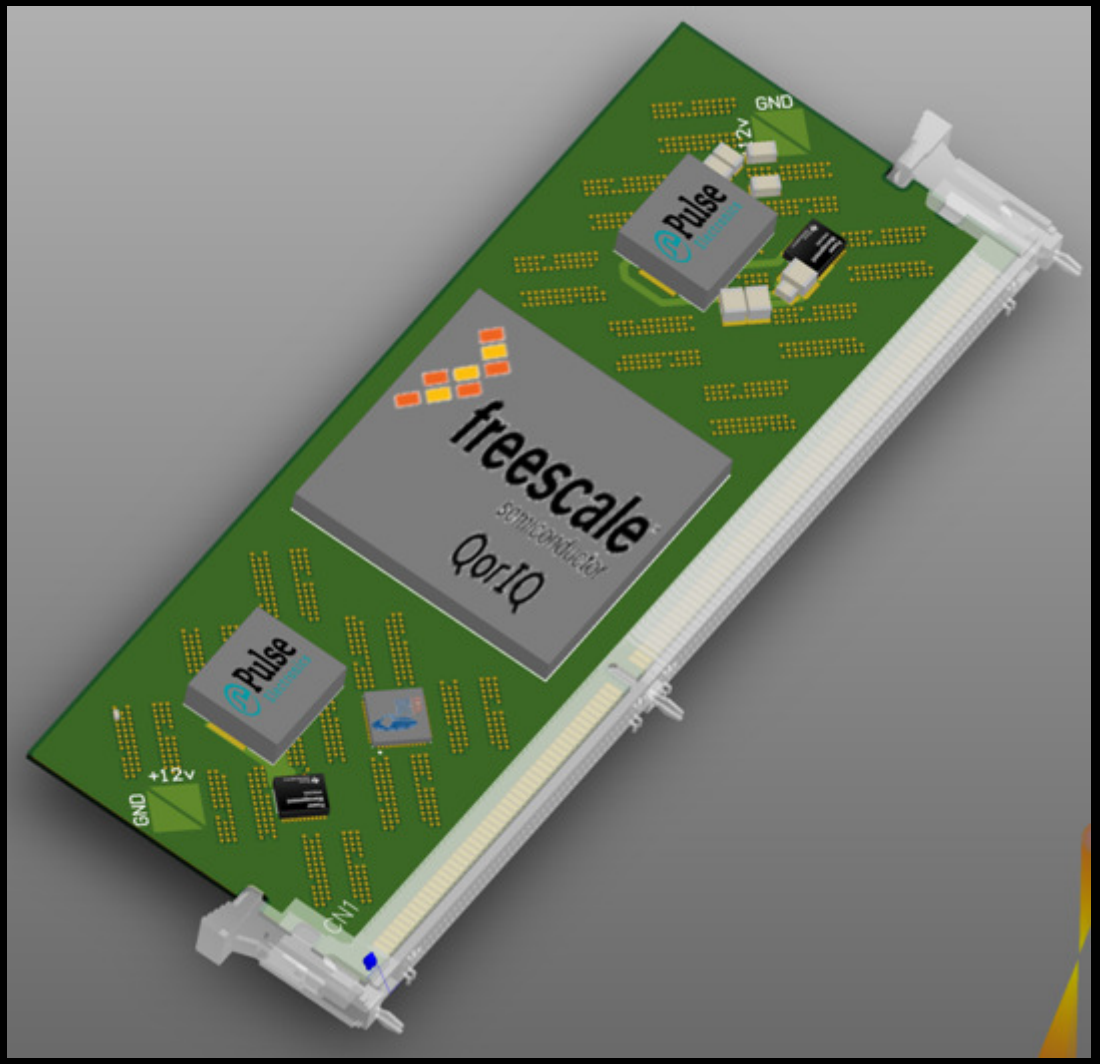
Power.ORG

DOME Demonstrator compute node board



Compute node interfaces across DIMM connector

- 1 interface SATA
- 5 interfaces Gigabit ethernet
- 2 interfaces 10 Gigabit ethernet
- SD card interface
- USB interface
- Various power supply levels



Power.ORG

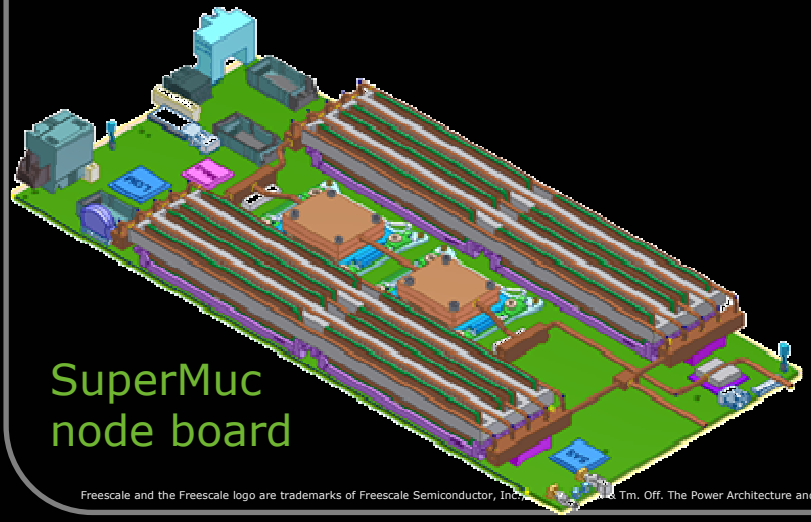
Hot Water Cooling

Most Energy Efficient solution:

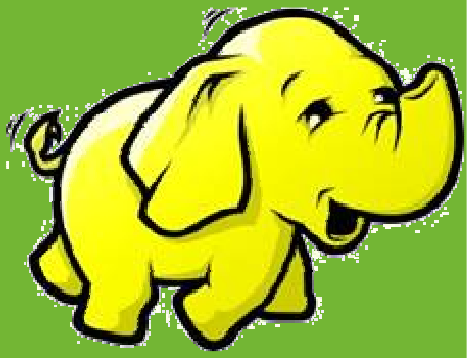
- Low PUE possible (<=1.1) – Green IT
- 40% less energy consumption compared to air-cooled systems
- 90% of waste heat can be reused (CO_2 neutral according Kyoto protocol)
- Allows very high density
- Less thermal cycling - improved reliability
- Lower T_j reduces leakage current – further saving energy

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

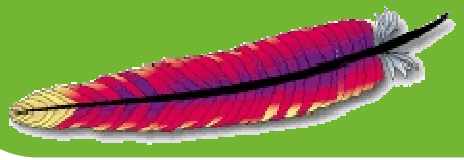
- No 4 on June 2012 TOP500 HPC list



SuperMuc
node board



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*
- Come and see our demo!
  - Technology exporer (runs php in browser)
  - WMD Workload Multi-User Driver (Java based)
  - DB2 data base engine

- Runs stable – able to exercise without any issues

Power.ORG

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
- Come and see our demo!

Power.ORG

HPC CPMD application port

HPC Carr-Parinello Molecular Dynamics package
For Ab Initio simulations - a key HPC application

- LAPACK install: compile required - 10 min job
  - Using Gfortran and GCC – no errors
- CPMD code base configured for PPC64, 2 cores
  - Natively compiled in 15 mins
  - ~100k lines of Fortran

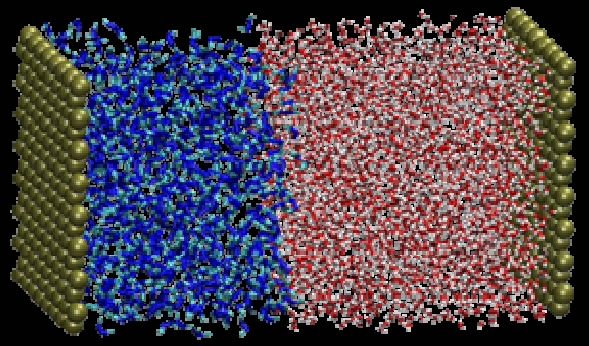


Image Courtesy Jülich Forschungszentrum

- Come and see our demo!

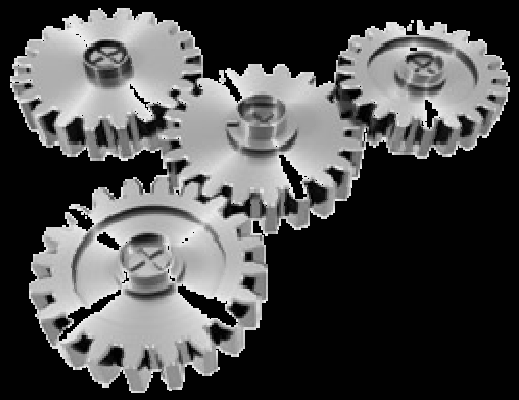
Power.ORG

Conclusion

- Server Class 64 bit OS and business applications on PowerPC commodity SOC have arrived

- IBM and Freescale demonstrated on *Embedded* PPC64 (Book E):
  - 64 bit Fedora 17
  - IBM DB2 – no compilation necessary to run
  - Hadoop – no compilation necessary to run
  - HPC CPMD application – straightforward port in a few hours

- Come and see our live demos from 16:30 – 18:00 today

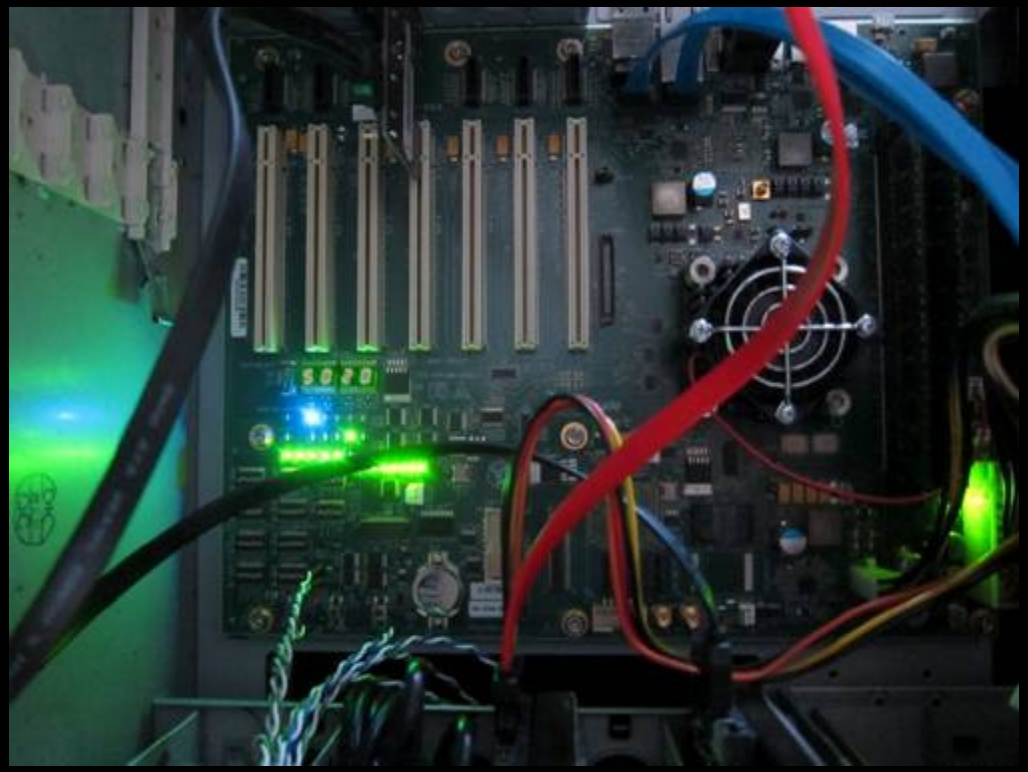


Power.ORG

You can access a P5020 with this link:

http://www.swissdutch.ch:6999/

… is a domain hosted on a P5020DS system, running Apache HTTPD server on top of Fedora 17, PPC 64 bit.

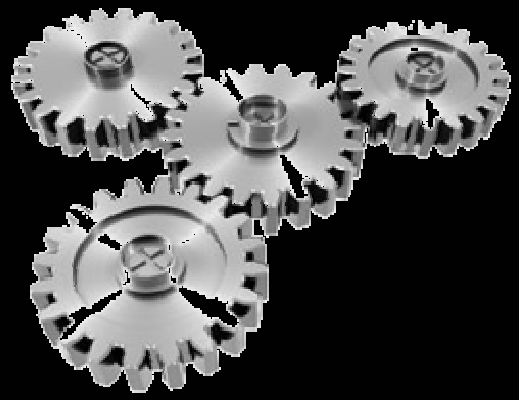


Acknowledgements

This work is the results of many *people*

- Peter v. Ackeren, FSL
- 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 & China; IBM worldwide; Dsgnworx - NL



we can make things fly