

ARM in the server space

2nd CERN Advanced Performance Tuning workshop

22 November 2013

Robert Richter <robert.richter@calxeda.com>

CALXEDA[®]
Power Your Tomorrow

Content

- ARM servers - outside and inside
- SoC
- IO
- Fabric
- Management
- Software

ARM Devices

Mobile Embedded Home



Raspberry Pi



ARM Servers



HP Moonshot hyperscale servers



Foxconn's 4U Storage Server with 60 drives



Gigabyte's new Web Server



Boston Viridis clusters (96 nodes)



Aeon's new Indus Cloud Storage Server

ARM Servers stand for

- Less power consumption
- SoC: higher integration - smaller footprint
- Many, many servers in one rack
- Less compute power than x86
- High I/O performance
- Fabric
- Manageability
- Reliability

ARM Servers - Looking inside

4 servers (nodes) on a card

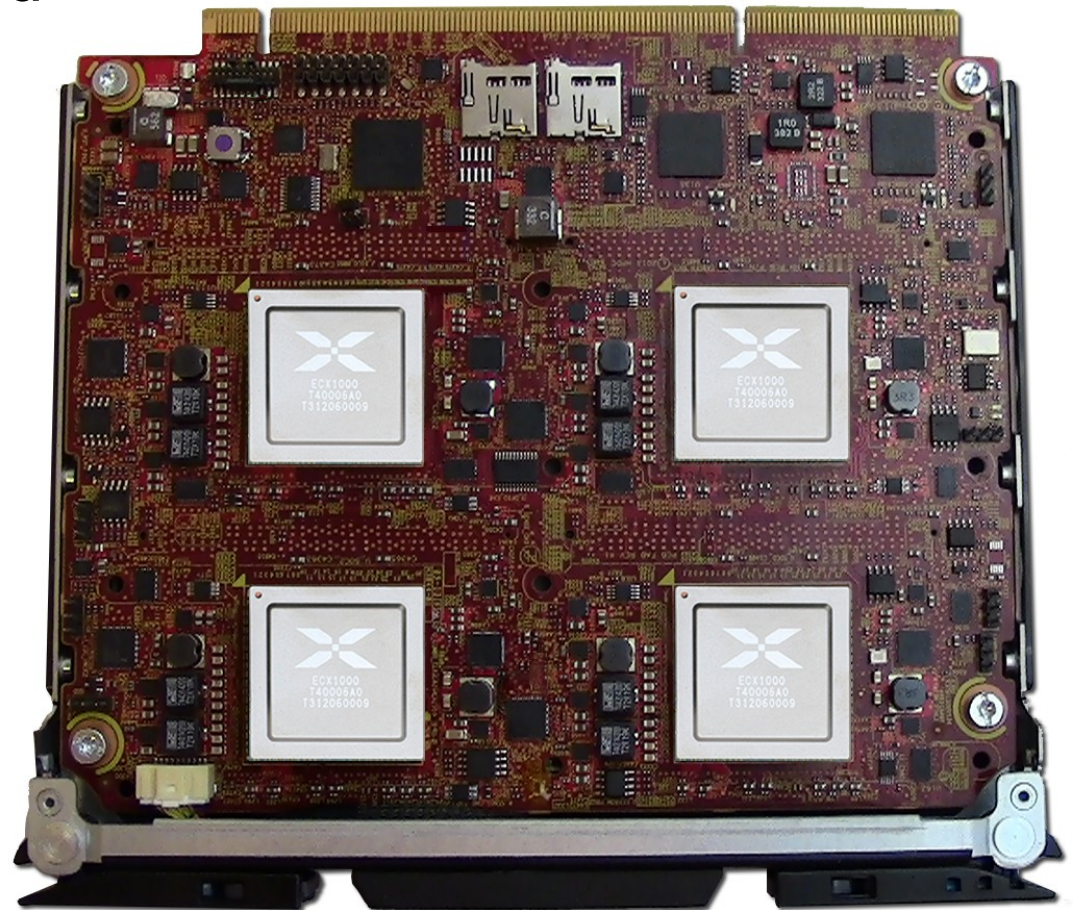
Usually 6 to 24 cards per chassis

Each node:

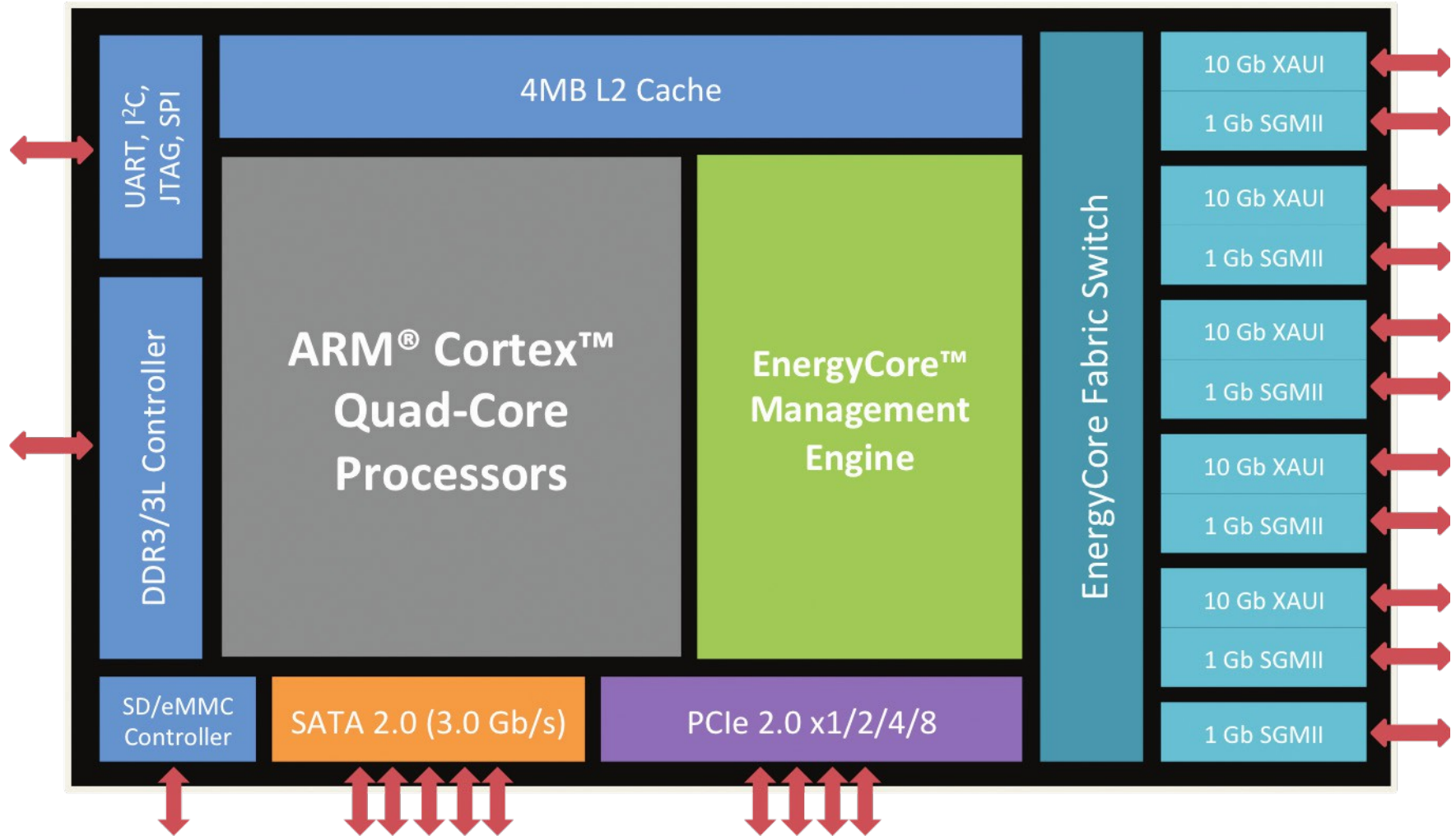
- 4 cpus
- up to 16 GB ECC
- up to 5 SATA, 5 10Gb Ethernet, 4 PCIe (x4)

20 Watts all nodes

No cables



System-on-Chip (SOC)



ECX-2000 SoC

- 4 ARM Cortex-A15: up to 1.8 GHz, hw virtualization, LPAE, secure boot (ARM TrustZone)
- 32 KB L1 icache/core, 32 KB L1 dcache/core, 4 MB shared L2 cache/soc with ECC
- ECC DDR3/1600 memory (40 bit phys addressable)
- up to 5x SATA
- up to 5x 10Gb XAUI or 6x 1GB SGMII Ethernet
 - node-to-node latency under 200 nanoseconds
 - 3 checksum offload engines)
- up to 4x PCIe G2 (x1, x2, x4) or 2x PCIe G2 x8
- 12 power domains

Microservers

- Up to 4096 nodes configurable
- Less cooling and thermal efficient
- Means same power envelope, but more computers
- Many, many small servers in one case
- Works well for highly parallelizable loads
- High speed interconnects between the nodes „fabric“
- Management becomes a challenge

I/O Driven Workloads

- Distributed, stateless server workloads: web server, caching servers and content distribution
 - Distributed application level storage, clustered file systems (Ceph, GlusterFS)
 - Processing large data sets: MapReduce (Apache Hadoop), distributed NoSQL data stores (Apache Cassandra)
 - Data intensive applications ([<http://www.graph500.org/> Graph 500 benchmarks])
 - efficient web servers (lack of multiple network ports, balanced
 - CPU performance and network I/O)
- Remote direct memory access (RDMA)
- Hyperscale Workloads

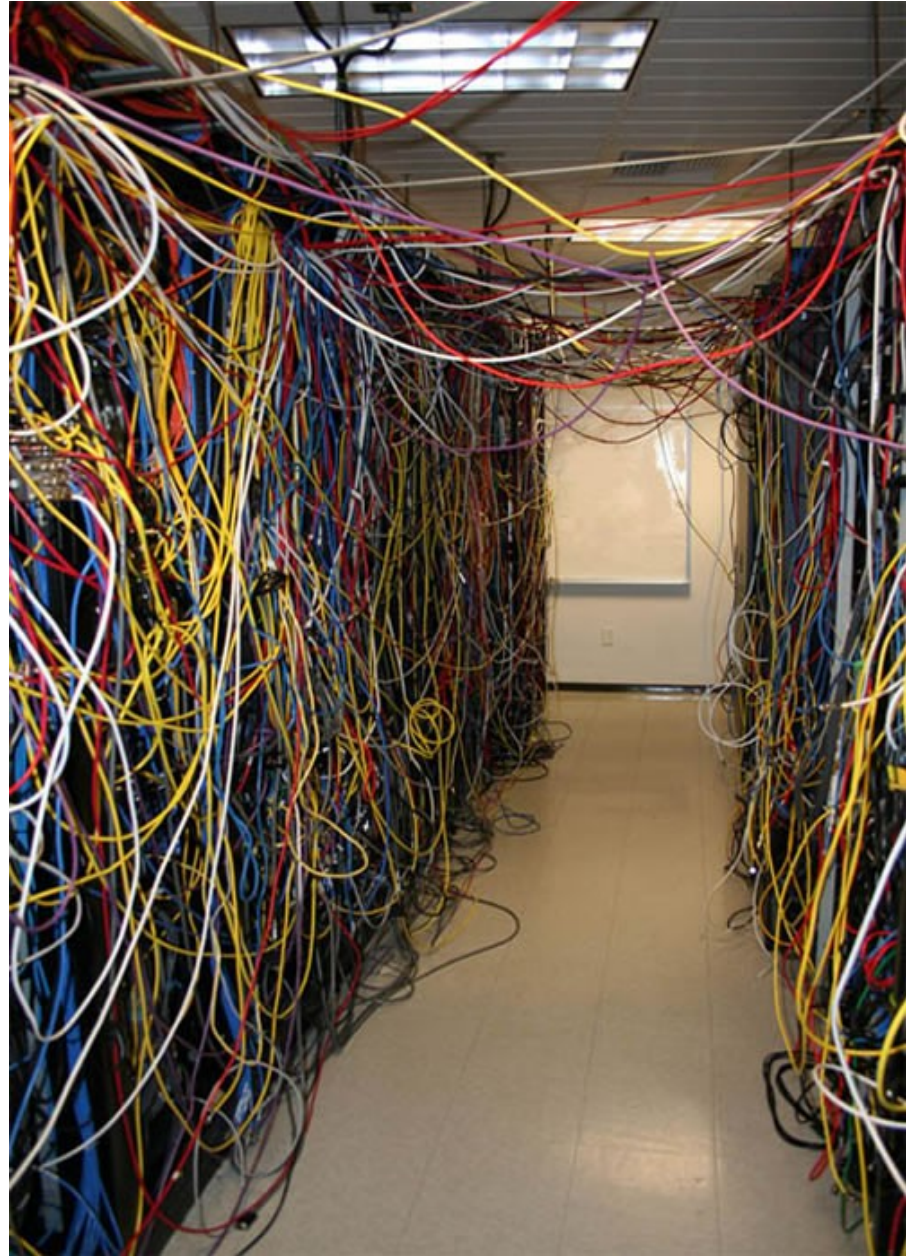
Why Fabric?

12 cards/chassis (2U)

4 nodes/card

20 chassis/rack

= 960 network ports



What is Fabric?

Transparent, reconfigurable interconnect between many SoCs

- No physical network ports required between systems
- Redundant paths and routing optimizations built-in
- Cluster level, server level and/or policy level controls
- Ethernet transparent in-the-rack workloads

Enhanced IO and networking to all nodes

- Less switches and cables
- Low latency & high bandwidth
- Lower power and higher density

Management Controller

- Dual-core Cortex-A7 for systems management
- On the SoC
- Integrating this on the chip (reduced power consumption and costs)
- For both single-node and cluster-wide management
- Runs firmware:
 - SoC Management: power management
 - Fabric Management: cluster-aware, routing, power
 - Remote Systems Management (IPMI, DCMI)
 - Custom extensions
- Out-of-band remote access, dedicated ethernet MAC

IPMI

- Say good bye to display and keyboard
- Serial-over-LAN (SoL)
- Power on/off nodes
- Read sensor data
- System logging

Software on ARM servers

- Ubuntu 13.10, openSUSE 12.3, Fedora 18
- OpenStack cloud controller (Havanna)
- Clustered file systems (Ceph, GlusterFS)
- distributed databases (Apache Cassandra)
- Virtualization (XEN, KVM)
- Linaro

Linaro



- not-for-profit engineering organization consolidating and optimizing
- open source Linux software and tools for the ARM architecture
- 200+ engineers
- Working on:
 - kernel: consolidation, ARMv8
 - power management
 - tool chain (gcc, etc.)
 - Enterprise Group (LEG): kernel, virtualization, boot loader,
 - UEFI/ACPI, RAS, profiling
 - etc.

Linaro kernel statistic

Linux Kernel Upstreaming

- Linux 3.12 RC6
- Total patch sets in this kernel release: 10480
- 212 known companies contributed to this kernel release

Linux 3.10 Contributors by company		
1	Intel	1028 (9.8%)
2	None	964 (9.2%)
3	Linaro	732 (7.0%)
4	Red Hat	707 (6.7%)
5	Unknown	492 (4.7%)
6	Samsung	256 (2.4%)
7	IBM	249 (2.4%)
8	Freescale	245 (2.3%)
9	Renesas	225 (2.1%)
10	Texas Instruments	350 (2.6%)



Linaro is #2
company
contributor

Source: <http://lwn.net/Articles/570483/>

Comming soon

ARM64

ARM in the server space

Questions?