Battle of the Accelerator Stars

Xipeng Shen

The College of William and Mary
& MIT

Top500 6/2012

Rank	Site	Computer
1	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu
3	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
4	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR IBM
5	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 NUDT
6	DOE/SC/Oak Ridge National Laboratory United States	**Jaguar** - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 Cray Inc.
7	CINECA Italy	**Fermi** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
8	Forschungszentrum Juelich (FZJ) Germany	**JuQUEEN** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
9	CEA/TGCC-GENCI France	**Curie thin nodes** - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR Bull
10	National Supercomputing Centre in Shenzhen (NSCS) China	**Nebulae** - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 Dawning

source: top500.org

Xipeng Shen

Green500 (6/2012)

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	2,100.88	DOE/NNSA/LLNL	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	41.10
2	2,100.88	IBM Thomas J. Watson Research Center	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	41.10
5	2,100.86	Rensselaer Polytechnic Institute	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	82.20
6	2,100.86	University of Rochester	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	82.20
7	2,100.86	IBM Thomas J. Watson Research Center	BlueGene/Q, Power BQC 16C 1.60 GHz, Custom	82.20
8	2,099.56	University of Edinburgh	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	493.10
9	2,099.50	Science and Technology Facilities Council - Daresbury Laboratory	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	575.30
10	2,099.46	Forschungszentrum Juelich (FZJ)	BlueGene/Q, Power BQC 16C 1.60GHz, Custom	657.50

What is the role of accelerator?

source: green500.org

Q1: Role of Accelerator

It is a trend.

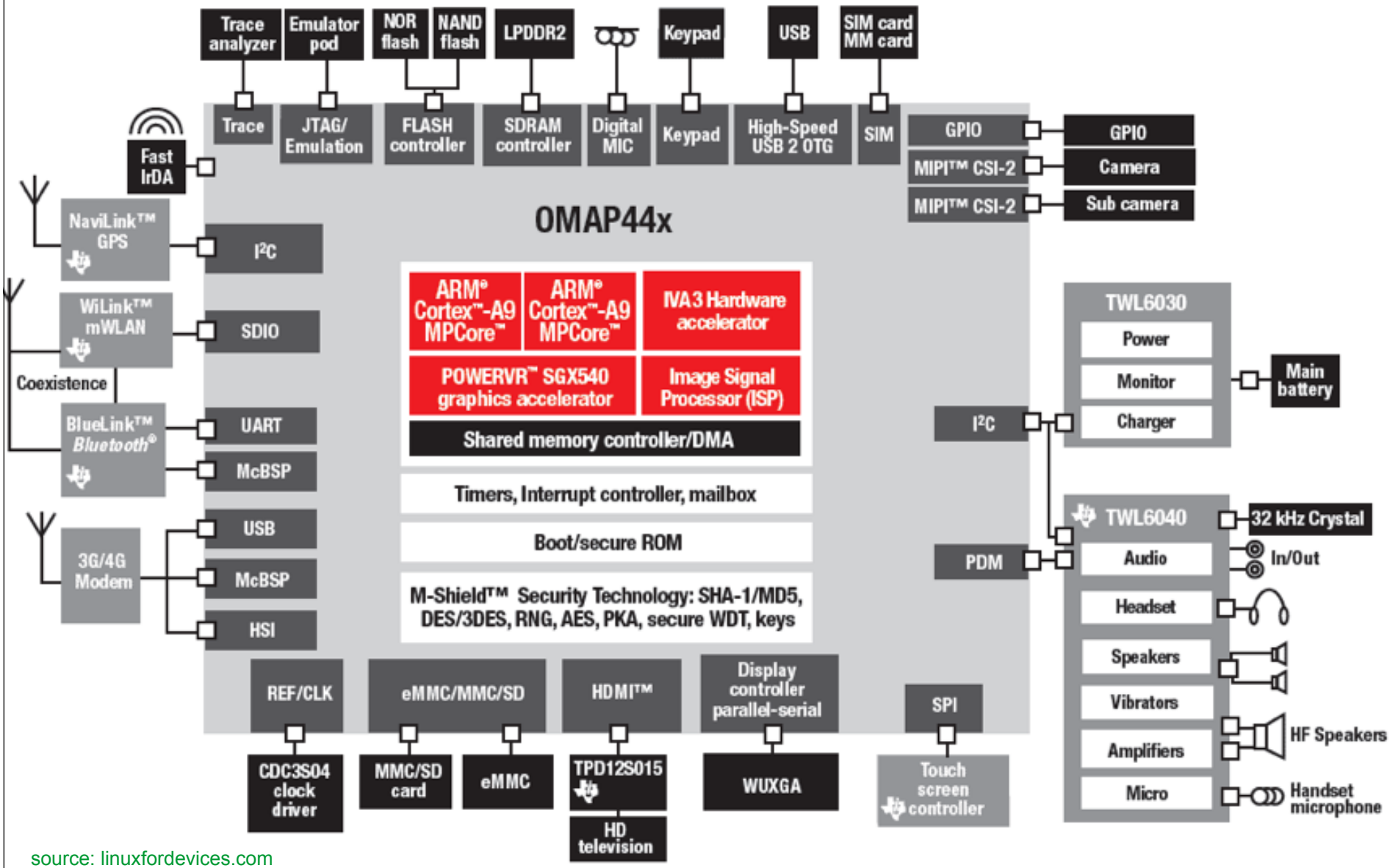
A common perception in processor industry.

- Intel: Ive Bridge, MIC
- AMD: Fusion (APU)
- NVIDIA: Kepler, Tegra, Denver Project
- IBM: (CELL), (Kilocores), specialized proc. on FPGA
- TI: OMAP
- ... ...

Q1: Role of Accelerator

- Adoption in practice
  - Embedded systems: long history.
  - Desktop, workstations, servers: GPU, FPGA.
  - HPC:
    - Titan (ORNL): 19KCPU+14KGPU, 20PF
    - Blue Waters: 49K CPU+3KGPU, 11.5PF
- Reasons
  - Moor's law continues
    - More transistors meet power wall and memory wall
  - Power efficiency
    - Specialization gives efficiency: An old receipt.

Q1: Role of Accelerator

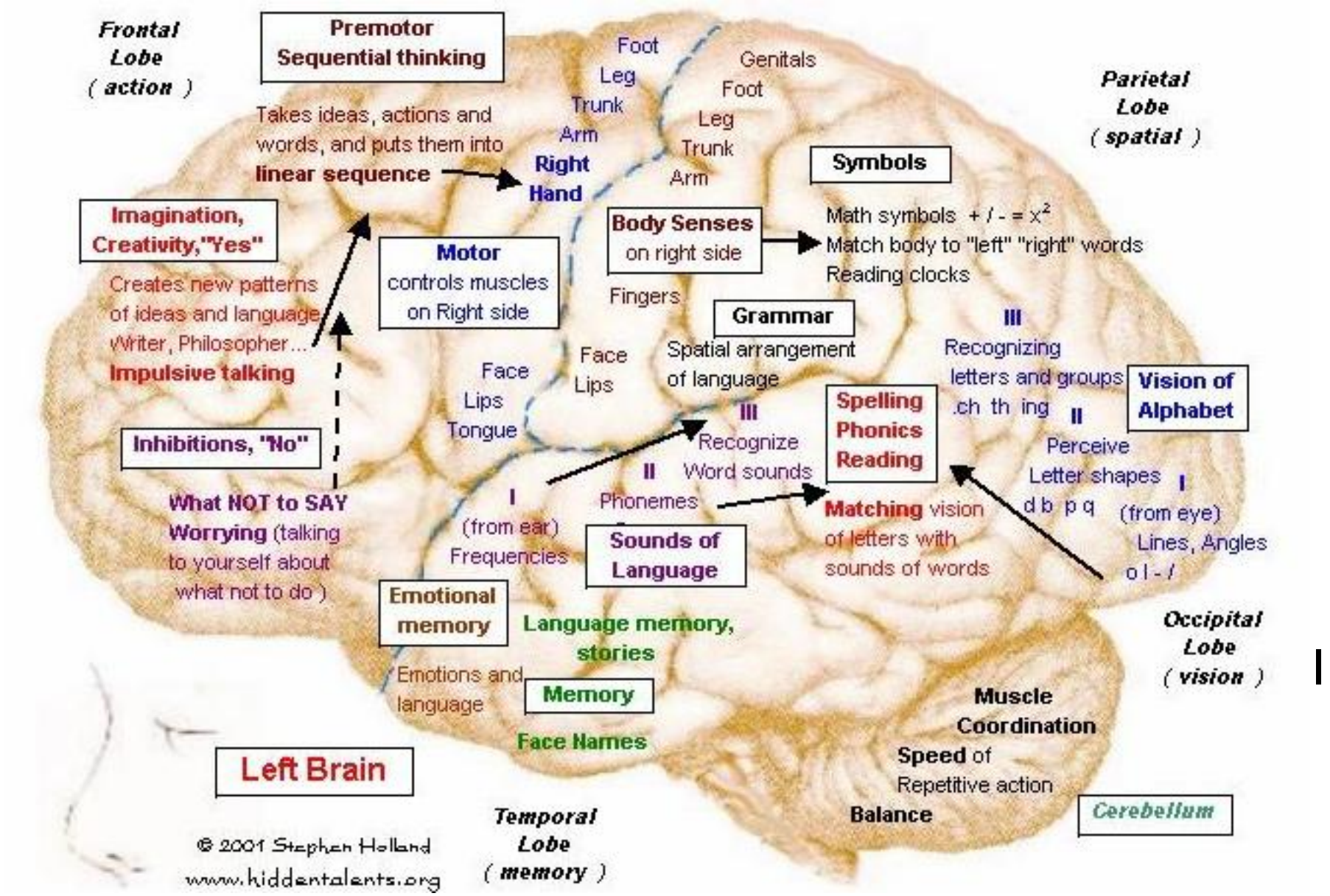


source: linuxfordevices.com

Q1: Role of Accelerator

- Adoption in practice
  - Embedded systems: long history.
  - Desktop, workstations, servers: GPU, FPGA.
  - HPC:
    - Titan (ORNL): 19KCPU+14KGPU, 20PF
    - Blue Waters: 49K CPU+3KGPU, 11.5PF
- Reasons
  - Moor's law continues
    - More transistors meet power wall and memory wall
  - Power efficiency
    - Specialization gives efficiency: An old receipt.

Q1: Role of Accelerator



Q2: SW/HW Divergence

Hardware excitement meets the cold reality of parallel programming.

- HW
  - non-uniformity, massive parallelism, variety
- Complexities are shifting to SW
  - SW trails HW more than ever
  - Multiple dimensions
    - Productivity, Efficiency, Performance, Portability, Fault tolerance

Q2: SW/HW Divergence

- Dynamically changing spectrum of SW support
  - DSL
  - CUDA
  - OpenCL
  - C++AMP
  - Cilk plus
  - Directives (e.g., OpenACC)
  - Various tools for assistance
  - ...

Q3: Which HW will win?

- NVIDIA GPU currently dominate
- Integrated CPU+GPU is promising
  - Others (e.g., FPGA, DSP) will co-exist
- Depends on SW support
- Not just accelerators, but whole processors
  - e.g., ARM for HPC?

Q4: Which SW will win?

- Current situation (for HPC)
  - CUDA has been adopted a lot
    - Pros: High performance, minor extensions from C
    - Cons: not portable yet, some programming efforts
  - OpenCL and directives have the potential
    - Pros: Portable; directives are easier to use
    - Cons: Performance
  - DSL draws increasing interest.

Q4: Which SW will win?

- In the decade
  - For HPC:
    - Key question: can the performance of OpenCL and directives catch up?
    - Some good signs, but also hard lessons (e.g., single-source compiler for cell)
  - For others:
    - OpenCL and directives are more likely
  - An analogy
    - C/C++ & Java

Q5: Challenges

- Key challenge: programming support.
  - productivity
  - performance & efficiency (locality, communication, balance)
  - fault tolerance
  - portability (exec. & performance)

- Irregular computations