Computing : Present and Future

- Mandar Gurav

Present

- Single cores
  - Performance improvments due to
    - Clock frequency scaling
    - Single thread performance
      - pipelining
      - out of order executions
      - Caches
      - ...
- Multicores
  - Moore's law is alive!

Present – Sandy Bridge

- Intel Advanced Vector Extensions (Intel AVX)
- Enhanced front-end and execution engine
- Cache hierarchy improvements for wider data path
  - two symmetric ports for memory operation, increased buffers, Improved prefetching
- System-on-a-chip support
  - Integrated graphics and media engine, PCIE controller, memory controller
- Next generation Intel Turbo Boost Technology

Courtesy: Intel

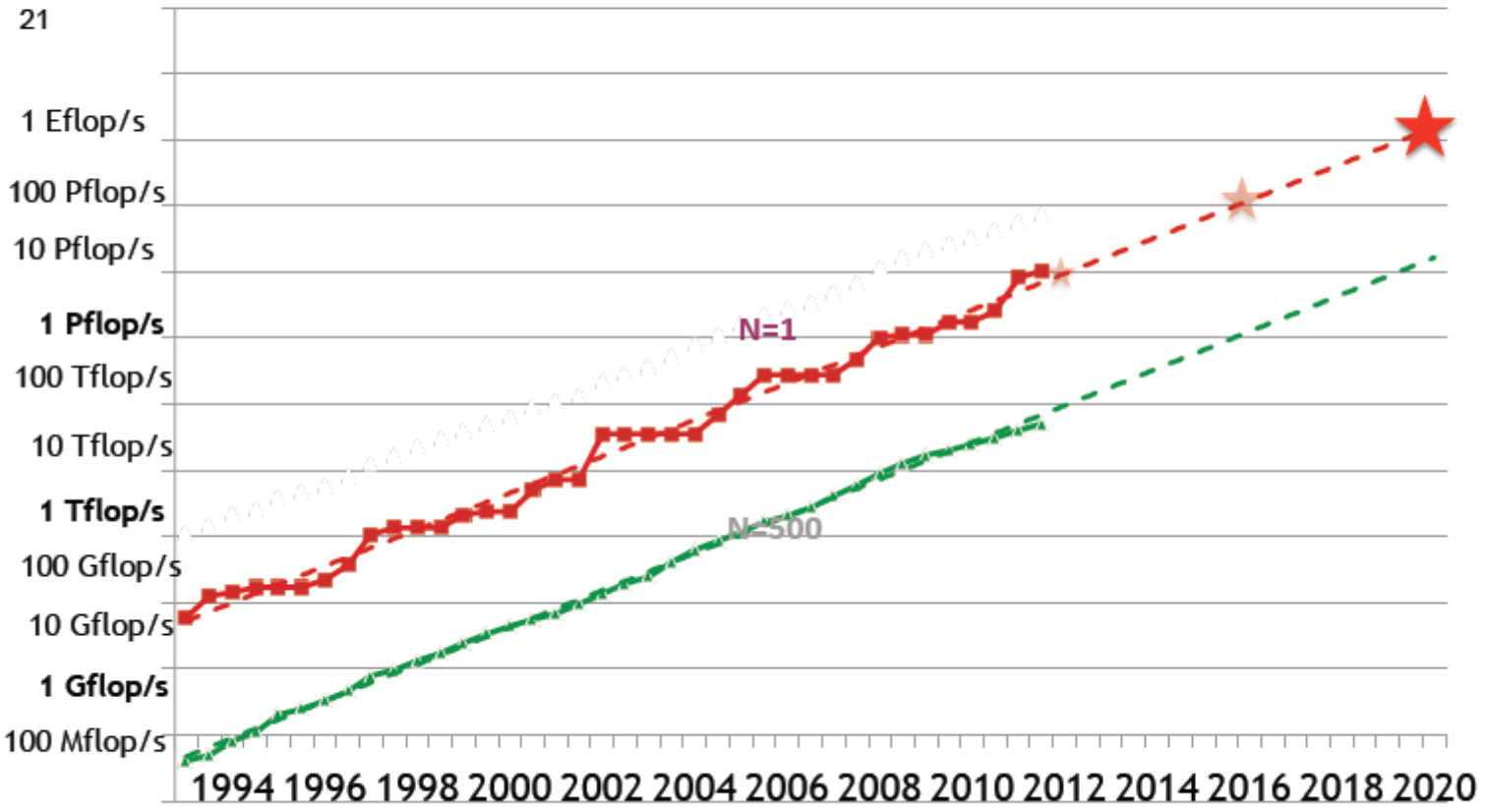
Accelerators

- GPGPU's
  - Nvidia
    - Stake in top 10 systems
  - AMD/ ATI
- Other accelerators
  - IBM Cell Broadband Engine
    - Roadrunner – First petaflop system
  - FPGA

Future

- Driving forces
  - Desktop users
    - Graphics /User friendlyness
    - Office
    - ...
  - HPC users
    - Performance matters
    - Code portability
    - Stability
    - ...

Performance Development in Top500



Credits : Prof. Jack Dongarra

Exascale

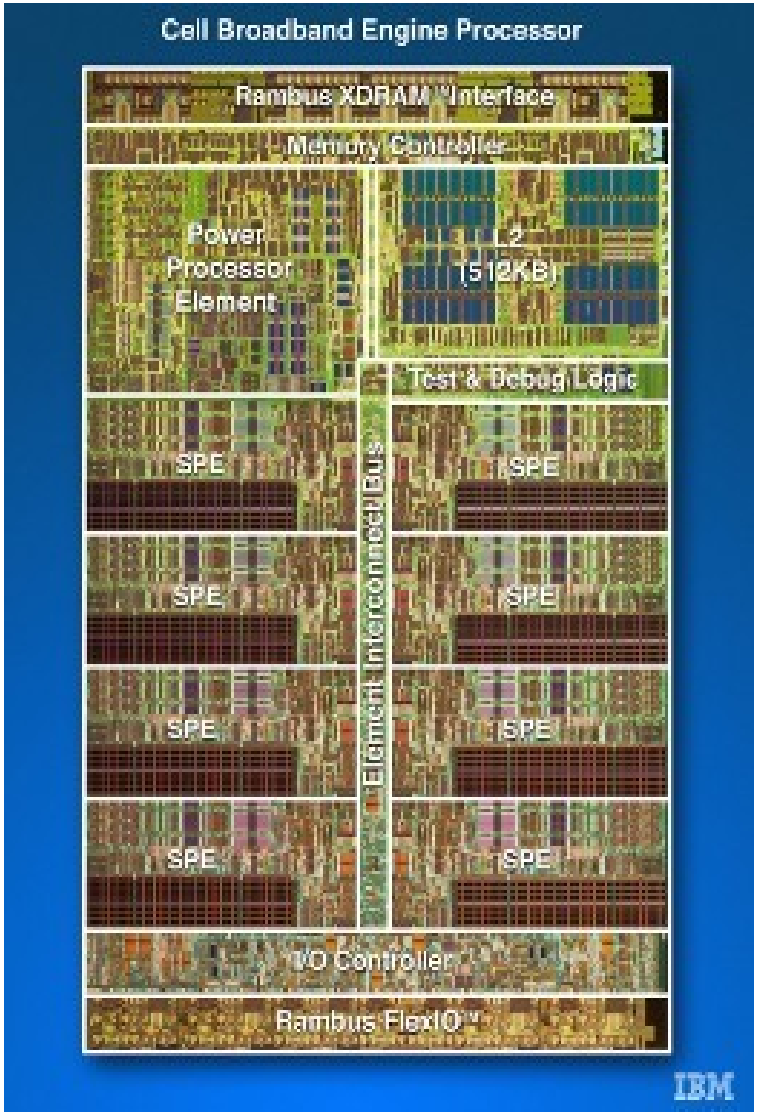
- 10^18 flops
- Challenges
  - Power
  - Fault tolerance
  - Concurrency
  - Others
    - Storage capacity
    - Programmability
    - Efficiency

Credits : International Exascale Software Project

How we are going to attack this?

IBM Cell Broadband Engine

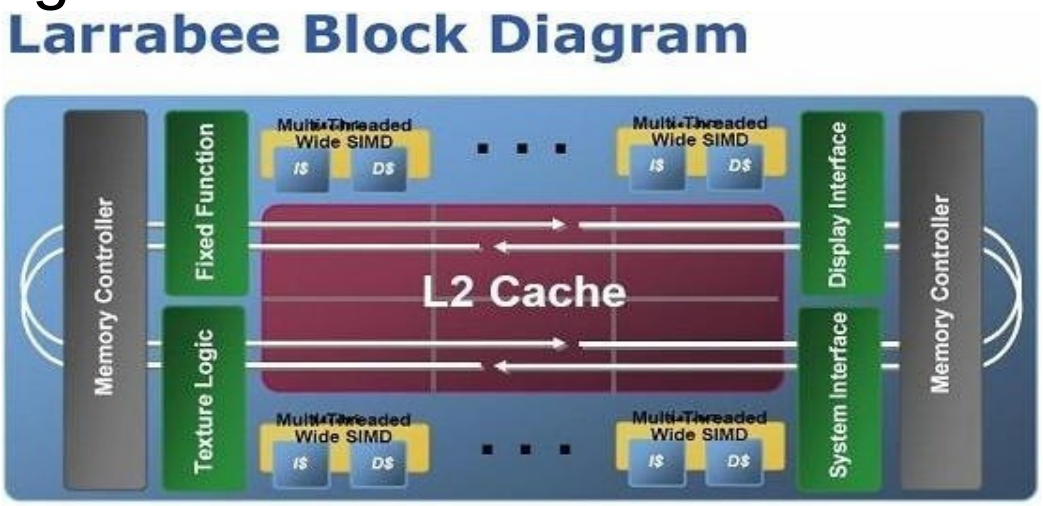
- Power Processor Element(PPE)
  - 64-bit PowerPC RISC core (can run OS)
- Synergistic Processing Elements (SPE)
  - a 128x128 register file
  - a floating-point unit
  - two fixed-point units
  - VMX vector arithmetic unit
  - Local Store
  - DMA controller



Courtesy: IBM

Larrabee

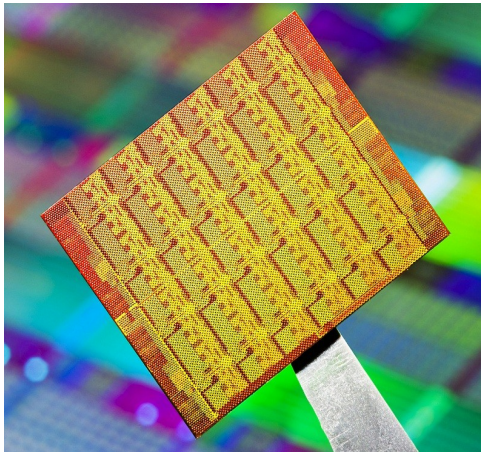
- Originally a GPGPU chip
- 32 general purpose x86 cores
- Shared global memory
- Interconnection network: a ring
- Shared L2
- Cache control instructions
- 4 way MultiThreading
- VPU
  - 16-wide Integer / single-precision FP
  - 8-wide double-precision FP



Courtesy: Intel

Single Chip Cloud Computer (SCC)

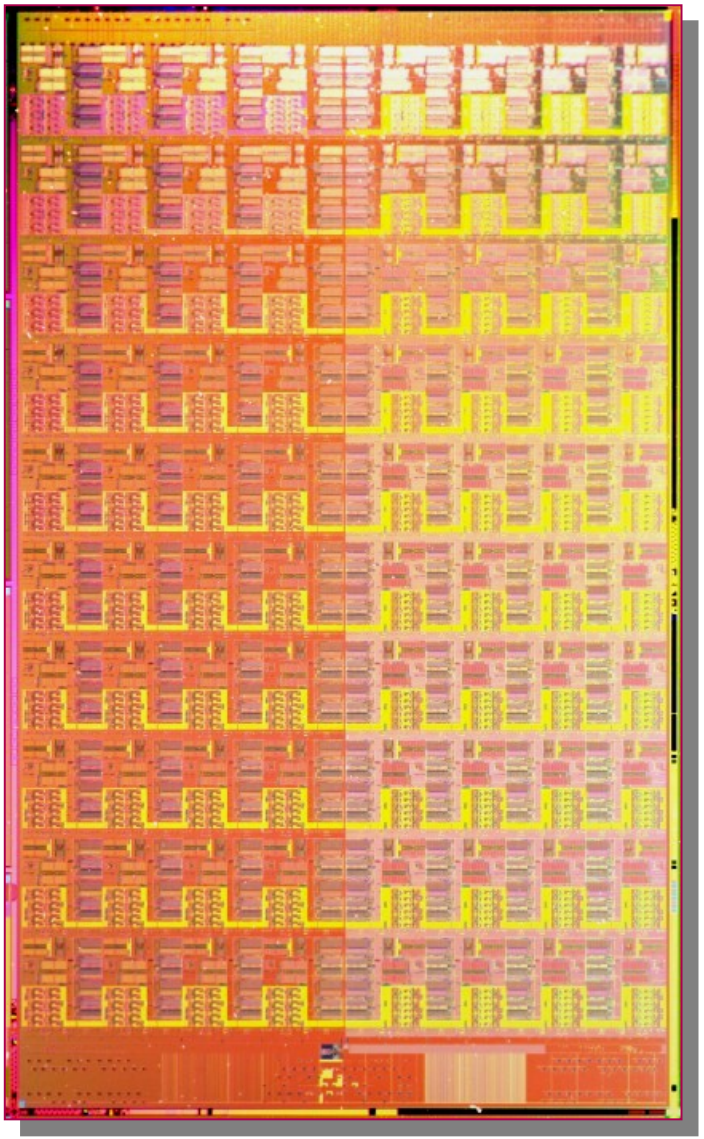
- First Si with 48 iA cores on a single die
- Power envelope 125W, Core @ 1GHz, Mesh @ 2GHz
- Message passing architecture
  - No coherent shared memory
  - Proof of concept for scalable many-core solution
- Next generation 2D mesh interconnect
  - Bisection B/W 1.5Tb/s to 2Tb/s, avg.power 6W to 12 W
- Fine grain dynamic power management



Courtesy: Intel and Karthik.V.M.

Terascale Research chip

- Deliver Tera-scale performance
  - Single precision TFLOP at desktop power
  - Frequency target 5GHz
  - Bi-section B/W order of Terabits/s
  - Link bandwidth in hundreds of GB/s
- Prototype two key technologies
  - On-die interconnect fabric
  - 3D stacked memory
- Develop a scalable design methodology
  - Tiled design approach
  - Mesochronous clocking
  - Power-aware capability

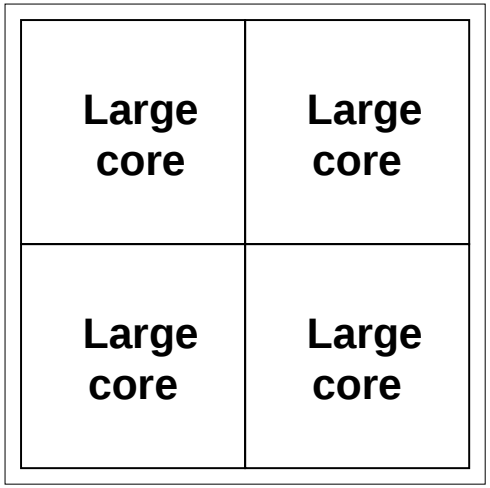


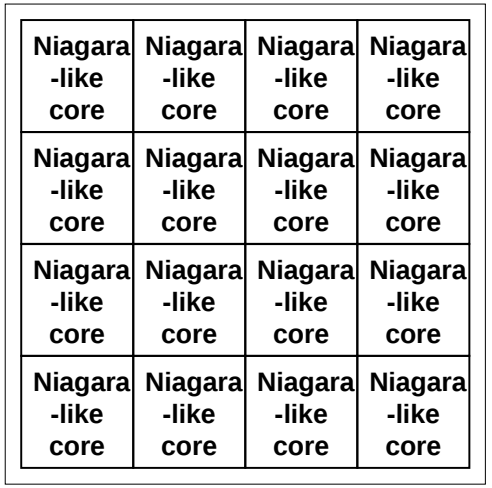
Credits : Jim Held, Intel

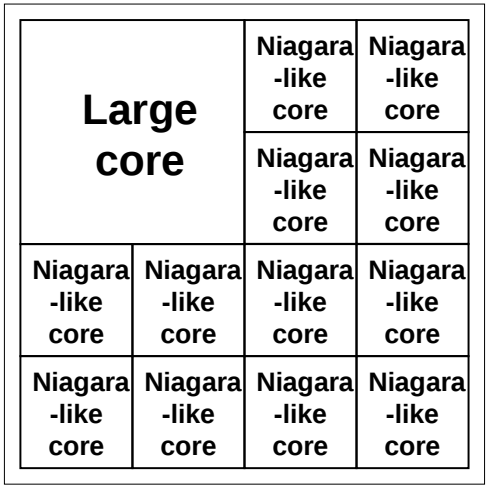
Many Integrated cores

- Co-processor – **x86 Based**
- Has own linux running
- can run program in following modes
  - Offload mode - Host as initiator
  - Native mode – Device itself executes the binary
  - Offload to Host – Device as initiator
- PCIE based
- 512 bit SIMD unit
- Cache coherent
- On-chip Network with message passing support

Heterogeneous vs Homogeneous cores







"Tile-Large" Approach

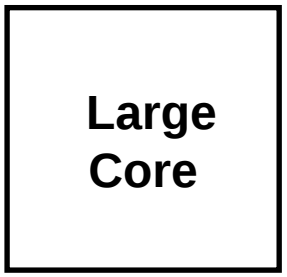
"Niagara" Approach

ACMP Approach

ACMP: Asymmetric Chip Multiprocessor

Credits : Prof. Yale Patt

Large core vs. Small Core



- ***Out-of-order***
- ***Wide fetch e.g. 4-wide***
- ***Deeper pipeline***
- ***Aggressive branch predictor (e.g. hybrid)***
- ***Many functional units***
- ***Trace cache***
- ***Memory dependence speculation***

- ***In-order***
- ***Narrow Fetch e.g. 2-wide***
- ***Shallow pipeline***
- ***Simple branch predictor (e.g. Gshare)***
- ***Few functional units***

Credits : Prof. Yale Patt

Some questions...

- Instruction Level Parallelism
- x86 or Arm or something else
- MIC or GPGPU or something else
- How many cores
- Programming models
- Role of Compilers
- ??

Thank you.