Emerging storage and HPC technologies to accelerate big data analytics

Jerome Gaysse
JG Consulting
Introduction

- Big Data Analytics needs:
  - Low latency data access
  - Fast computing
  - Power efficiency
- Latest and emerging technologies
  - Memories
  - Interfaces
  - Controllers
  - New generation of SSDs
Big Data Analytics

- Standard approach
  - Data stored in HDD
  - Data transferred to DRAM memory
  - Processed by the server CPUs

- Drawbacks
  - HDD is slow
  - DRAM is non-volatile
  - CPU is not power efficient
Memories

- New memory technologies
  - NandFlash
  - Magnetic
  - Resistive
Memories - Nandflash

- Already used in data centers (SSD)
- Various interfaces (SAS, SATA, PCIe)
- Benefits:
  - Faster than HDD (few GB/s on a PCIe SSD)
  - Low latency 20-50µs
- Higher $/GB, but need less infrastructure: a 1u all-flash array can deliver the same performances a 42U rack
Memories - Magnetic

- MRAM Memory
  - MRAM memory chips in production, but low density (256Mbit chips)
  - Available as chip and DIMM form factor
- Benefits
  - Non volatile
  - Fast memory: DDR-like interface
Memories - Resistive

- RRAM
  - In development
  - Benefits
    - Non-volatile
    - High density: roadmap to 1 TB/chip
    - Faster than Nandflash, slower than MRAM
- Phase-Change Memory (PCM)
  - Technology demonstrator existing
  - 1μs latency range on a PCIe SSD
Interfaces & subsystem

A fast CPU with a fast memory technologies are not useful with a slow interface!

- Interfaces improvement
  - PCIe & NVMe
  - Memory bus & DIMM
  - CAPI
  - HMC
Interfaces & subsystem – PCIe SSD

- **NVMe**
  - NVM Express is an optimized, high performance, scalable host controller interface with a streamlined register interface and command set designed for Enterprise and Client systems that use PCI Express* SSDs. NVM Express was developed to reduce latency and provide faster performance with support for security and end-to-end data protection.

- **PCIe faster than SAS and SATA**
  - SATA 3: 12Gb/s
  - PCIe Gen 3 x8: 64Gb/s
Interfaces & subsystem – PCIe SSD

- Architecture example

![Diagram of SSD controller with interfaces and subsystems: PCIe, NVMe, NandFlash, DDR3]
Interfaces & subsystem - DIMM

- Adding NandFlash on the memory bus
  - DDR-like interface with non-volatile feature?
  - Or SSD with DDR-like interface?
- Both!
  - NV-DIMM (up to 8GB)
    - DRAM with NandFlash as a storage backup in case of power failure
  - UltraDIMM (Sandisk) up to 400GB
    - Full SSD on the DIMM bus, <5µs write latency
Interfaces & subsystem - DIMM

- NV-DIMM
  - NV-DIMM
    - DDR3
    - Controller
    - NandFlash
  - SuperCap

- UltraDIMM
  - UltraDIMM
    - NandFlash
    - Controller
    - NandFlash
  - SuperCap
Interfaces & subsystem – PCIe and CAPI

- IBM Capi interface
  - Power8 CPU interface
  - Coherent Accelerator Processor Interface
- Protocol on top of PCIe.
- Used to connect auxiliary specialized processors such as GPU, ASIC, FPGA. Can use the same memory address space as the CPU.
Hybrid Memory Cube consortium

- 3D DRAM technology using high-speed logic process technology with a stack of through-silicon-via (TSV) bonded memory die.
- A single HMC can provide more than 15x the performance of a DDR3 module.
- Utilizing 70% less energy per bit than DDR3 DRAM technologies.
- Using nearly 90% less space than RDIMMs.
Controllers

- X86 CPUs are commonly used for Big Data processing
- Easy to program
- Most important part of the power budget
Controllers - FPGA

- Field Programmable Gate Array
  - Allows full hardware acceleration processing
  - Field-update capability
  - 5 to 10x better performance/power vs a software solution
  - OpenCL programmable
Controllers - FPGA

- **Examples**
  - Microsoft is using FPGA board for Bing processing acceleration
  - Intel to come with FPGA and Xeon in a single package
  - IBM CAPI interface for FPGA
Controllers – RISC CPU

- Products available
- Software ecosystem in development
- Lower performance vs x86…
- …but very lower power
- Need multiple chips to reach the same performance, at a reduced power budget
New generation of SSDs

- With “in-situ processing capabilities
- Local big data analytics

SSD controller

- PCIe
- NVMe
- DDR3
- NandFlash

Multicore CPU

NandFlash

DDR3
New generation of SSDs

- New distributed programming models

```
Main CPU
```

```
“big data” SSD
```

Same analytics in each SSD

```
Main CPU
```

```
“big data” SSD
```

Analytics splitted on all SSDs
Future of big data analytics architecture example

- Memory: HMC, DDR, MRAM, RRAM
- DIMM
- Generic Processing: RISC, RISC, RISC
- PCIe
- IPA
- Specific Processing: FPGA
- PCIe SSD
- RRAM
- Nandflash
What’s next?
Next decade – Silicon Photonics

- **Today**

![Diagram showing network, controller, fiber channel, memory, DIMM, CPU, PCIe, and storage connections.](image-url)
Next decade – Silicon Photonics

- 2020-2025
Conclusion

- Stay tuned, technologies are evolving rapidly
  - New memories
    - Fast like DRAM
    - Density and nonvolatile like NandFlash
  - PCIe bus to connect SSD and FPGA
  - FPGA for power efficient dedicated processing
  - RISC CPU for lower power consumption
Emerging storage and HPC technologies to accelerate big data analytics

Thanks!

Jerome Gaysse
JG Consulting
jerome.gaysse@embedsyst.com