Annual Update on Computational Storage

Flash Memory Summit

Chuck Sobey
Chief Scientist
ChannelScience

November 12, 2020, 8:35-10:05 AM Pacific, Session A-9

Agenda

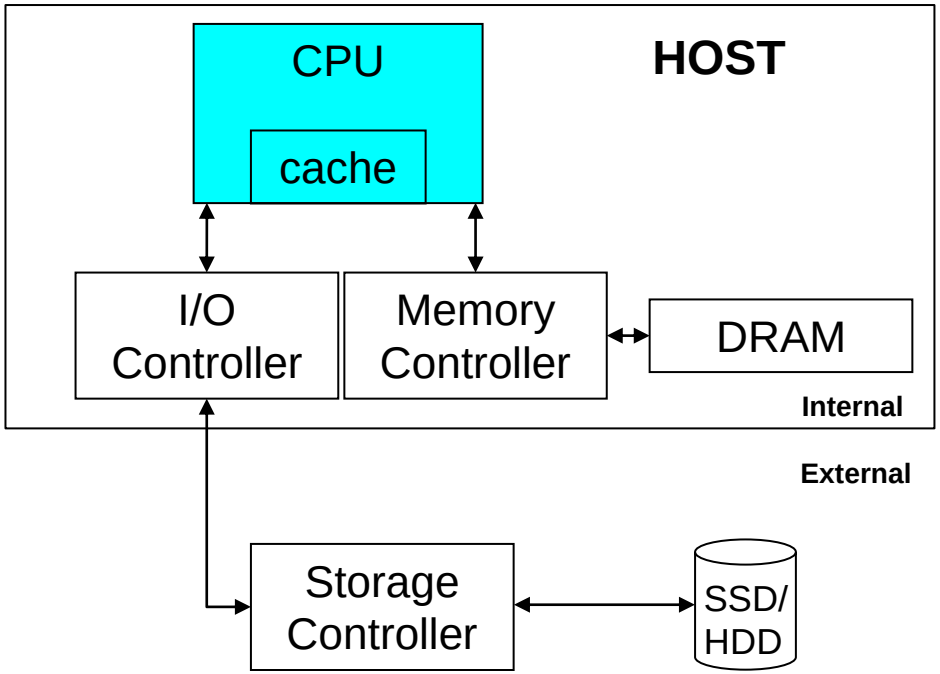
- The Problem Computational Storage Solves
- Data-centric Computing
- Terms
- Architectures
- The Industry
- Outlook

Moving Data is Costly in Time and Power

- In many applications, it takes more time and energy to get the data from memory than to process it
- It takes more time and energy to get the data from storage to the memory

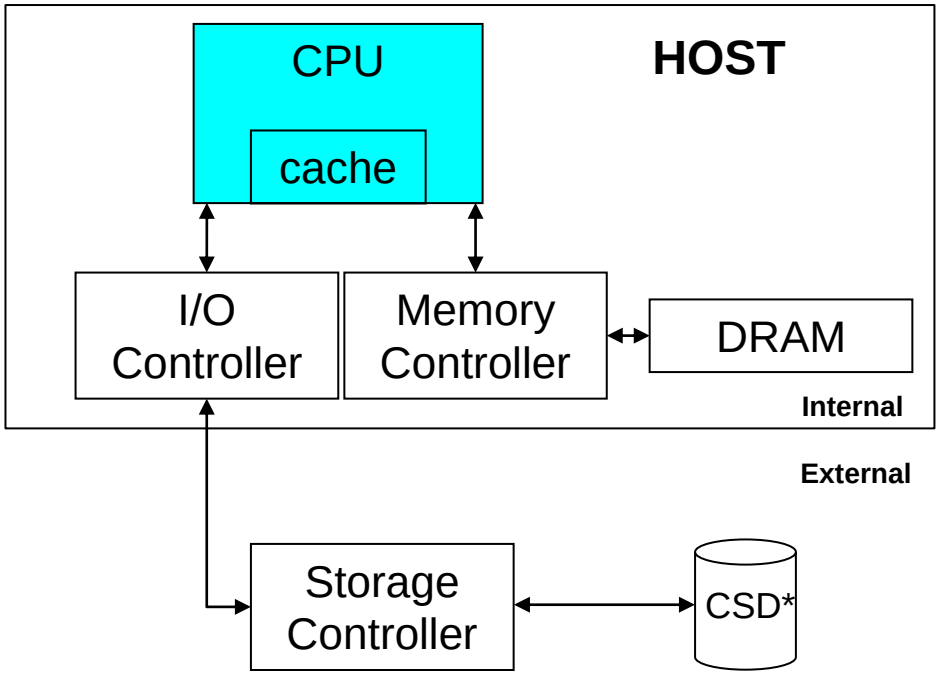
Bring compute to the data

The Typical Architecture: Database Example



- Search a database for transactions on Mondays in June
  - Load entire database into memory
  - Search for the records
- Perform the analysis on those records

Computational Storage Reduces Data Traffic and Processor Tasks



- Host commands CSD* to assemble records for transactions on Mondays in June
- CSD searches for the records
- Load records into memory
- Perform the analysis on those records

* CSD = Computational Storage Device

Moving the Processing to the Data

- Computational Storage
- In-storage Processing
- *In situ* Processing

"Computational Storage"

- Near-data Processing
- Near-memory Processing

General terms for data-centric computing approaches

- Processing-in-Memory
- Neuromorphic Computing

New or existing memory technologies with math hardware built-in

- (HBM* has a Logic Layer)

*HBM: High Bandwidth Memory, widely used with GPUs

The Computing Paradigm is Shifting

Compute-Centric	Data-Centric
Solve differential equations	Analyze petabytes of data
Bottleneck: CPU/Memory	Bottleneck: Storage & I/O
Computational fluid dynamics	Search & mining
Finite element analysis	Network analysis
Multi-body simulations	Digital media creation & transmission
	Modeling (environment, etc.)

Source: IBM

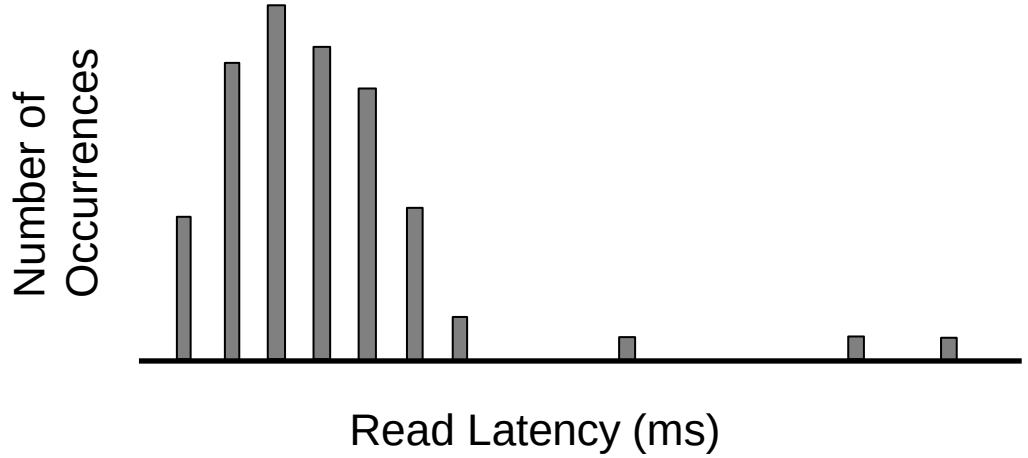
Benefits of Computational Storage

- Frees CPU cycles for other tasks
- Enables parallelism
- Saves time of moving data
  - Reduces latency
  - Reduces network/bus traffic
- Privacy and security

Data Centers
Cloud Services Providers

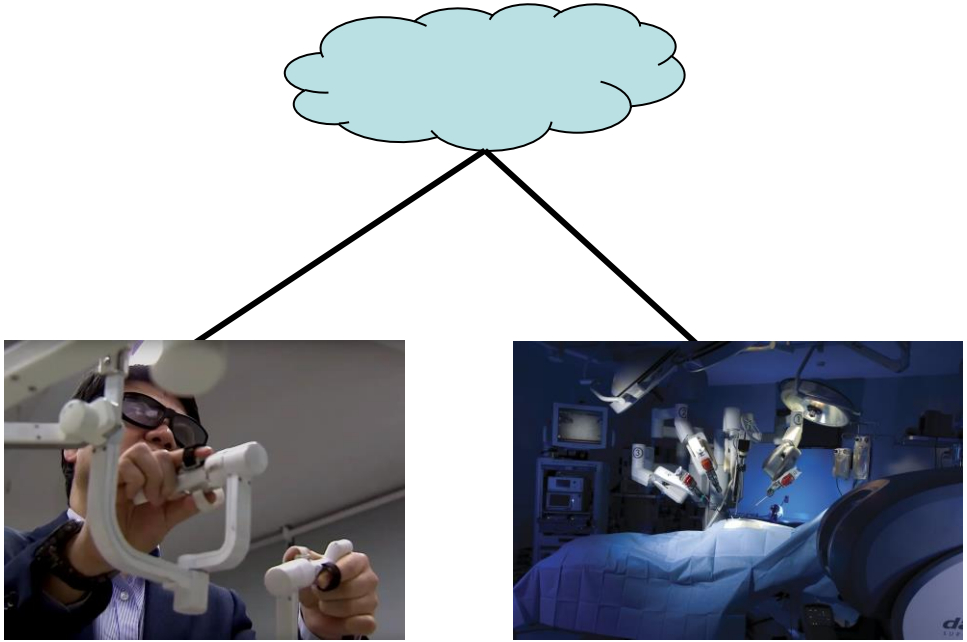
Edge Devices

Predictability: Tail Latency



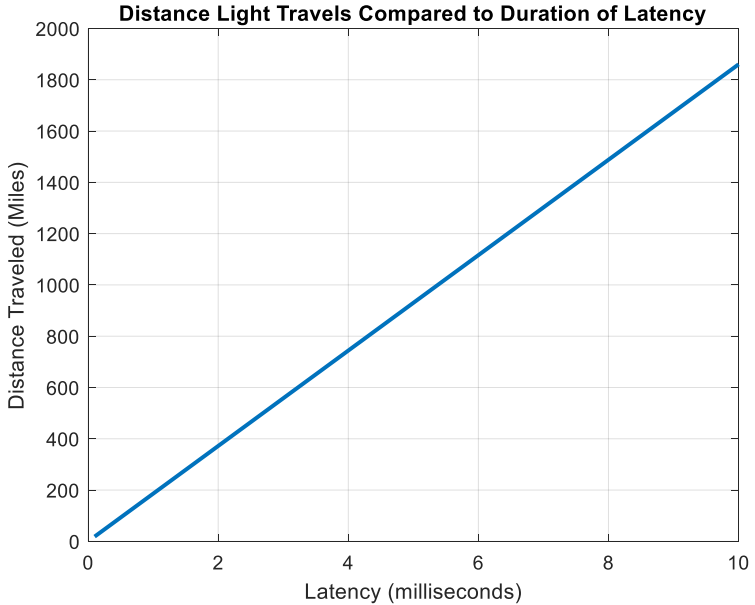
- 5G is supposed to provide reduced and predictable latency

5G, IoT, and the Edge



Source: Business Insider

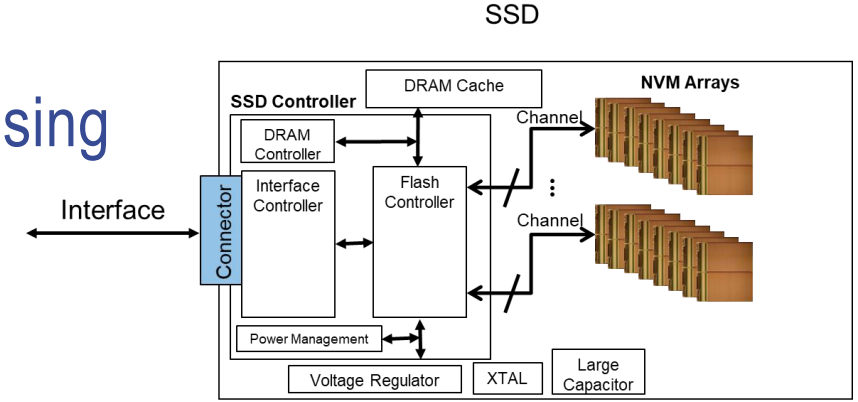
Source: Mercatornet



5G promises the “tactile internet.”

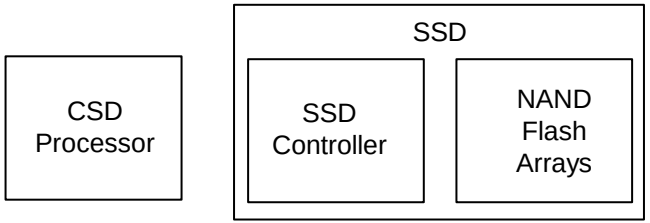
Computational Storage in a Nutshell

- SSD = NAND flash + processing

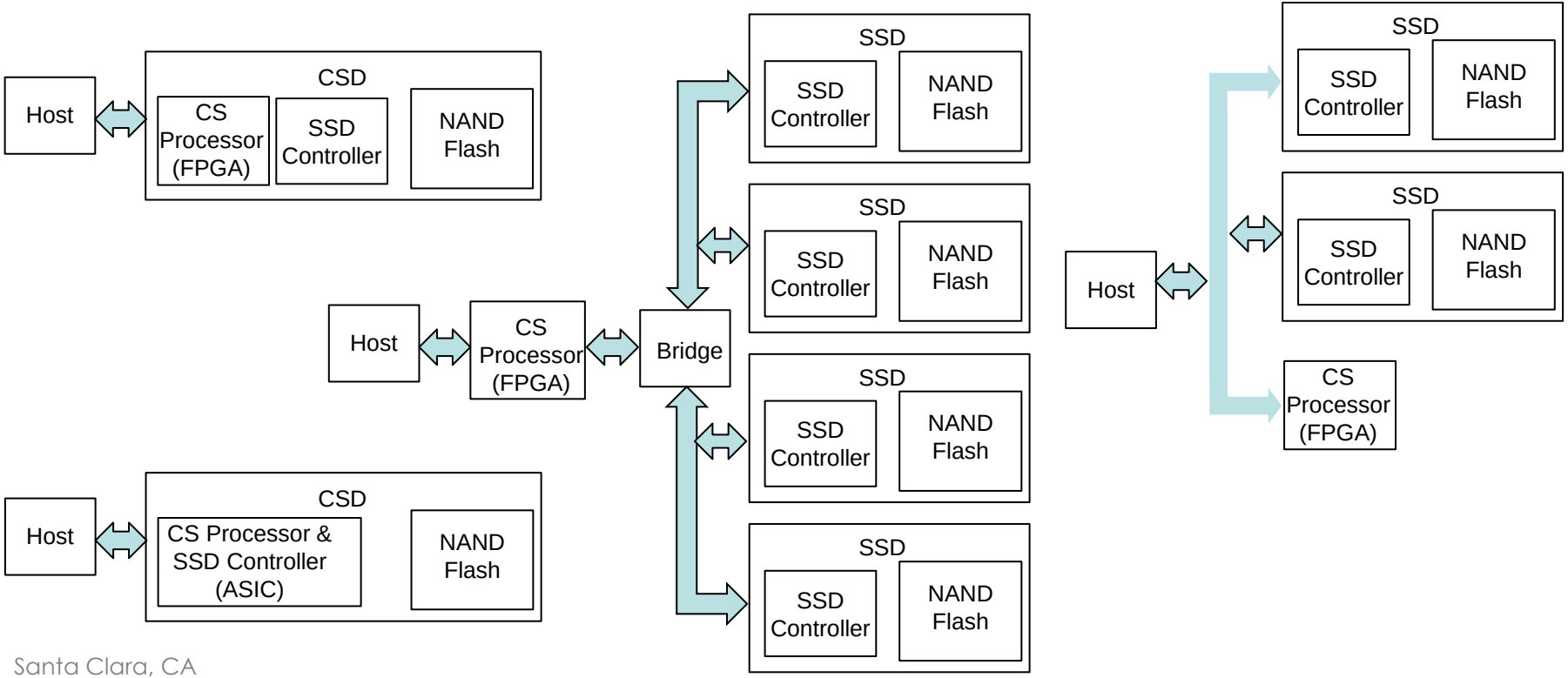


Source: KnowledgeTek's SSD class

- Computational storage = SSD + more processing



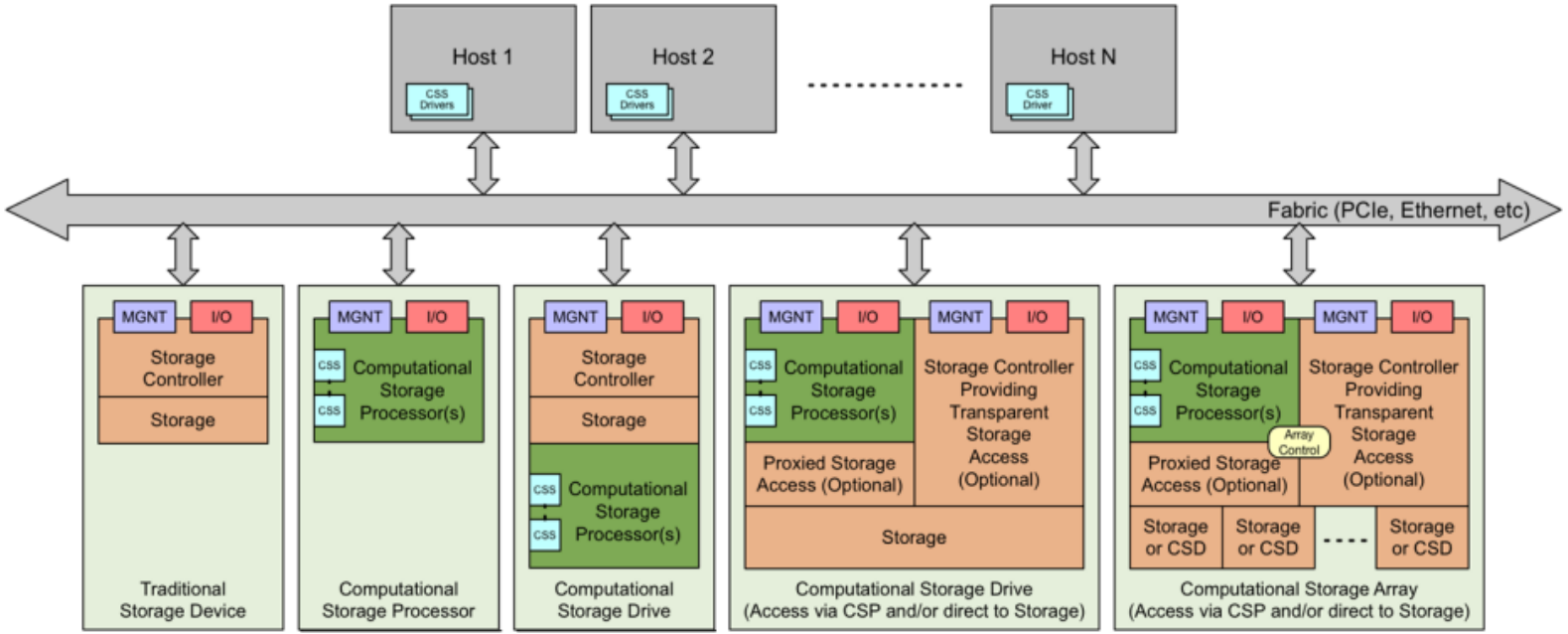
Current Computational Storage Architectures



Summary of Current Computational Storage Device Architectures

- FPGA or ASIC/SoC + Solid-State Storage
  - FPGA added to an SSD
  - ASIC/SoC and NAND flash
  - FPGA and bridge connected to several SSDs
  - FPGA or ASIC/SoC on a bus (PCIe fabric) with SSDs

Computational Storage Devices (CSx)



Source: SNIA

Industry Organizations: SNIA

- Storage Networking Industry Association (SNIA)
  - https://www.snia.org/computational
- SNIA
  - Compute, Memory, and Storage Initiative (CMSI)
    - Computational Storage Special Interest Group (SIG)
      - Computational Storage Technical Working Group (TWG)
- Specification out for public review
  - Computational Storage Architecture and Programming Model v0.5 rev 1
  - https://www.snia.org/sites/default/files/technical_work/PublicReview/SNIA-Computational-Storage-Architecture-and-Programming-Model-0.5R1.pdf

Group met at FMS 2018!

SNIA's Computational Storage (Draft) Spec Defines Capabilities and Actions

- Management
  - Discovery
  - Configuration
  - Monitoring
- Security
  - Authentication
  - Authorization
  - Encryption
  - Auditing
- Operation
  - Mechanisms for storing and retrieving data
  - Data locality information

SNIA's Spec Defines Computational Storage Services (CSS, CSSes)

- Fixed Computational Storage Services (FCSS) Examples
  - Compression
  - Deduplication
  - Encryption
  - Erasure coding
  - RAID
  - Regular expression (search)
- Programmable Computational Storage Services (PCSS)
  - Berkeley packet filter (network traffic analysis)
  - Container
  - FPGA bitstream (for rapid reconfiguration)
  - Operating system

SNIA CMSI Members

Source: SNIA 2019

Industry Organizations: NVM Express

- Nonvolatile Memory Express (NVM Express, “NVMe”)
  - https://nvmexpress.org/membership/nvm-express-working-groups/
- NVMe Computational Storage Task Group
- NVMe specification features for computational storage
- Chairs
  - Stephen Bates, Eideticom
  - Kim Malone, Intel
  - Bill Martin, Samsung

Big Acquisitions that could Affect Computational Storage

- NVIDIA to buy Arm (from SoftBank)
  - $40B USD, 9/13/2020
  - Cores used in most SSD controllers and computational storage processors
  - Expected result: More AI/ML in more places
- AMD to buy Xilinx*
  - $35B USD , 10/27/2020
  - FPGAs used in many computational storage architectures
  - Expected result: More accelerators in more places
- Must clear government hurdles worldwide

* Intel bought Altera (FPGA) for $16.7B, 12/28/2015

In this Session We'll Hear the Latest from

- Stephen Bates, Eideticom
- Andy Walls, IBM
- Neil Werdmuller and Jason Molgaard, Arm

and ChannelScience, too!

- US Department of Energy Proposal
- Computational storage device
- AI/ML customization
- Data triage
- Scientific Instruments
  - *e.g.*, electron microscopes

ChannelScience™
Establish the State-of-the-Art™

Source: Jeol

What to Watch for in Computational Storage

- Coordination between standards bodies
  - SNIA and NVM Express
  - Watch for possible developments at the Industrial Internet Consortium (merged with Open Fog) https://www.iiconsortium.org/
- More start ups emerge from stealth mode
- AI/ML at the edge
- 5G growth
- RISC-V – open source hardware and software
- Data center infrastructure scaling (*e.g.*, AWS AQUA)
- Leverage the parallel nature of NAND flash

Thank You!

If you want to follow-up later, you can reach me at
csobey@channelscience.com

Flash Memory Summit

Everything You Need To Know For Success