Flash Evolution Demands Controller Innovations

Roddy, Marketing Director
YEESTOR Microelectronics Co., Ltd
The Trend of NAND Flash

- Flash density growth will continue
- 3D NAND layer, 64L > 96L > 120L > 190
- NAND flash cell structures continue to advance, increasing storage density. This flash evolution demands new controller innovations.
Internals of Flash controller

- 1, Flash Translation Layer
- 2, Flash Error Handling
- 3, Flash Retention Management
- 4, SPOR Management
- 5, Optimize NAND trim parameters
Challenge and Innovation

**Challenge**
– High throughputs
– Low latency and high IOPs
– NAND reliability deteriorates under different types of stresses
– Von Neumann architecture limited controller innovation

**NAND controller Innovation:**
– Ultra Fast Flash IO
– MRAM Based Flash Controller
– Open Channel
– In Storage Computing
– Utilizing machine learning to optimize NAND trim parameters with minimal overheads
– Open ecosystem
Ultra-Fast Flash IO

- **High bandwidth interfaces**
  - 533MT/s, 667MT/s, 800MT/s..., 1066MT/s? 1200MT/s? 1600MT/s?

- YMTC Xtacking enables DRAM like high I/O speed With Xtacking, the periphery circuits which handle data I/O as well as memory cell operations are processed on a separate wafer using the logic technology that enables the desired I/O speed and functions

- Flash Controller have support Ultra-Flash Flash IO to reduce latency and improve IOPS

- Ultra Flash IO can help controller reduce flash channel and simplify the system design

- New PCIe, UFS controller design will support it
STT-MRAM SSD Architecture

- Improves reliability and manufacturability, supercapacitors is reduced
- Increases write performance with a write buffer
- Improve data management, reduce write amplification to NAND
Open Channel is architectural innovation

I/O Isolation

I/O isolation provides a method to divide the capacity of the SSD into a number of I/O channels that map the parallel units of the device.

Predictable latency

Predictable latency is achieved by having control in the host over when, where and how I/O are submitted to the SSD.

Software-Defined Non-Volatile Memory

By integrating the SSD flash translation layer into the host, workload optimizations can be applied either within a self-contained flash translation layer, file-system integration or applications themselves.
In Storage Computing

Operators Supported In ISC Accelerator

- Search/Scan
- MAX/MIN/Average/Sorting
- Deserialization/Subsample
- List Intersection
- Vector/Matrix arithmetical & logical operators
In-Storage Computing

- NAND Move the processing not the data
- ISC is an Ultimate approach for Acceleration I/O intensive Applications
- Significant reduction of amount and latency of data transfer
- Off-loads computing to the storage devices. Speeds up processing and utilizes otherwise unused processing capacity on the storage devices.
- Ultimately it improves overall system performance and power efficiency, which is particularly important in mobile devices.
Flash controller using machine learning
Flash controller using machine learning

Flash-specific SMART (Self-Monitoring, Analysis, and Reporting Technology) attributes to conduct an in-depth analysis of Flash reliability in a production environment. Big data are collected for training.

We leverage machine learning technologies, specifically data clustering and correlation analysis methods, to discover groups of Flash which have different health status and relations among Flash-specific SMART attributes.

DSP based on machine learning simplify the flow of ECC, characterization data into error models that are then used to generate application-specific LLR tables for the LDPC engine.
Open Storage Ecosystem

**IO Intensive APP Acceleration**
- Neural Network
- Big Data Analytics
- SQL Processing

**Storage-Computing Optimization**
- Predictable latency
- I/O Scheduling
- I/O Isolation
- Key-Value Store
- Atomic Write

**In Storage Computing**
- AI Accelerator
- Data Search & Compare
- File Compression
- Data Tagging
- AIOT

**Trusted Computing**
- Trust Root
- Trusted Computing
- Electronic Signature

**Open Source FTL/Flash Infrastructure**
- Parallel Flash Control Unit
- Open Channel Command Parser
- FTL Accelerator
- ISC Accelerator

**Open Source HW Abstract Layer**
- Multi-Core CPU
- Data Pre-Processing Engine
- Compression
- Application Specific Co-Processor

**Open Source Security Interface**
- TCG-OPAL stack
- OSCCA stack
- SM2/3/4
- AES/RSA/SHA/TRNG

Flash Memory Summit 2019
Santa Clara, CA
1. Offloads host compute to the storage device is the trend. In the future, it will be standard to allow host to download application to the device for the device processing.

2. In-storage compute reduces I/O traffic between storage and host.

3. Software-Defined Non-Volatile Memory is the new convergence of storage and memory, which previously were two separate computing domains. Storage is a multi-layer software implementation outside of the realm of Computer Architecture.

4. It demands architectural innovation for improving flash memory reliability, shortening the access path to data, reducing latency, and alleviating the bottlenecks at I/O ports.

5. Computational storage demands that the flash controller adds value as a compute engine.

6. Innovative flash controller features that leverage modern NAND flash technologies to deliver flexibility, scalability, and reliability for data processing applications are described.

7. Open flash controller ecosystem is trend.
Thank you