High Performance FTL for PCIe/NVMe SSDs

Ying Y. Tai, Ph.D.
Senior Director
VIA Technologies, Inc.
yingytai@viatech.com
Growing popularity of NAND flash memory and SSD:
- High read/write performance
- Low power consumption

SSD has changed the storage landscape on a wide range of applications: from embedded devices to data center.
## The Trend

<table>
<thead>
<tr>
<th>Host Interface</th>
<th>SATA III (6Gb/s)</th>
<th>PCIe (3GB/s/Lane)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAND Technology</td>
<td>Planar</td>
<td></td>
</tr>
<tr>
<td>NAND Interface</td>
<td>400MB/s, 533MB/s</td>
<td></td>
</tr>
<tr>
<td>NAND Density</td>
<td>MLC</td>
<td></td>
</tr>
<tr>
<td>RAM</td>
<td>SSD DRAM</td>
<td>Host Memory Buffer</td>
</tr>
<tr>
<td>ECC</td>
<td>BCH</td>
<td></td>
</tr>
</tbody>
</table>

**Performance vs. Cost**
Flash Translation Layer

- The bridge between the host and the storage media.

---

Host → FTL

- Address Mapping
- Wear Leveling
- Scheduling
- Power Failure Protection

FTL → Flash

- Bad Block Management
- Garbage Collection
- Flash Management
- ... Others
- Power Failure Protection

---
## FTL Recipe

<table>
<thead>
<tr>
<th>Minimize</th>
<th>Maximize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery Time</td>
<td>Parallelism</td>
</tr>
<tr>
<td>CPU Overhead</td>
<td>Garbage Collection Efficiency</td>
</tr>
<tr>
<td>Write Amplification</td>
<td>Wear Leveling Efficiency</td>
</tr>
<tr>
<td>Data Loss</td>
<td></td>
</tr>
<tr>
<td>Latency</td>
<td></td>
</tr>
</tbody>
</table>
2 Aspects

• Parallelism
  o Multi-plane operations (internal)
  o Multi-channel architecture (external)

• Write amplification
  o Garbage collection is the main contributors
  o Write workloads
Parallelism

- Multi-channel interleaving
- Grouping blocks into a superblock


B. Peleato et al., “Analysis of trade-offs in v2p-table design for NAND flash,” Asilomar 2012

Additional relevant: on write mapping: garbage-collection process involving additional read and write operations. This effect is known as write amplification and it results in both endurance and write-thruput degradation. Different garbage-collection policies lead to different write amplification pro-
Superblock

- Block management is based on the granularity of a superblock.
Superblock Management

• Pros:
  o Superblock spans multiple channels for concurrent write.
  o Block management overhead is reduced.
  o Conductive to RAID 5 implementation.

• Cons:
  o GC efficiency depends on data locality (hot/cold and sequential/random data).
Write Amplification

- Over-Provisioning
- I/O applications
- GC efficiency
- Data integrity writes
- Hot and cold data separation:
  - Algorithm (internal)
  - Host hints (external)
Superblock GC

- Workload: 4K random write with 18% hot data.
Real Workload Traces

<table>
<thead>
<tr>
<th>Server</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>usr</td>
<td>User home directories</td>
</tr>
<tr>
<td>proj</td>
<td>Project directories</td>
</tr>
<tr>
<td>prn</td>
<td>Print server</td>
</tr>
<tr>
<td>hm</td>
<td>Hardware monitoring</td>
</tr>
<tr>
<td>rsrch</td>
<td>Research projects</td>
</tr>
<tr>
<td>prxy</td>
<td>Firewall/web proxy</td>
</tr>
<tr>
<td>src1</td>
<td>Source control</td>
</tr>
<tr>
<td>src2</td>
<td>Source control</td>
</tr>
<tr>
<td>stg</td>
<td>Web staging</td>
</tr>
<tr>
<td>ts</td>
<td>Terminal server</td>
</tr>
<tr>
<td>web</td>
<td>Web/SQL server</td>
</tr>
<tr>
<td>mds</td>
<td>Media server</td>
</tr>
<tr>
<td>wdev</td>
<td>Text web server</td>
</tr>
</tbody>
</table>

Write Workload Characteristics

Data Size

Data Hotness
Real Workload Trace GC

Flash Memory Summit 2016
Santa Clara, CA
Data Classification

• Logical block addressing (LBA):
  o Frequency: hot/cold data
  o Recency: workload changes

• Hot and cold data separation
Host Hints: NVMe

Data Set Management (DSM) Hints:

- Access Size
- Written in Near Future
- Sequential Write
- Sequential Read
- Access Latency
- Access Frequency

- Typical Read and Write
- Infrequent Read and Write
- Infrequent Write, Frequent Read
- Frequent Write, Infrequent Read
- Frequent Read and Write
• Proposed by Samsung.
• A storage interface to inform (hint) SSDs about the data.
• The host system opens “streams” for different write requests.
• Data in a stream is written together to a related physical NAND flash space and separated from the data in other streams.

Host Hints: Other Interface

Summary

• PCIe with NVMe enables higher performance SSDs
• FTL: parallelism and write amplification management is critical
• Hot/cold data separation:
  o Data classification: adjusting with write workloads
  o Host hints: NVMe and other storage interfaces
• VIA VT6745 PCIe/NVMe SSD controller with FTL turnkey solution is the best option for high performance SSDs.
We Are @ Booth #818

VIA Technologies @ Flash Memory Summit

<table>
<thead>
<tr>
<th>Time</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forum E-21 Wed, Aug. 10 8:30 ~ 10:50am</td>
<td>High-Throughput LDPC Solution for Reliable and High Performance SSD</td>
</tr>
<tr>
<td>Forum M-22 Wed, Aug. 10 3:50 ~ 6:15pm</td>
<td>SSD Flash Management for 3D NAND Flash Memory</td>
</tr>
<tr>
<td>Session 302-D Thurs, Aug. 11 9:45 ~ 10:50am</td>
<td>High Performance FTL Architecture for PCIe/NVMe SSDs</td>
</tr>
</tbody>
</table>