Low-Power Design of SSDs

Starblaze
Daniel.Sun
daniel.sun@starblaze-tech.com
Mobile Storage: Power sensitive

CPU: Dynamic Voltage & Frequency Scaling (DVFS) is mature

DATA I/O: Determined dynamic power because traditionally unavoidable

Better Design Goal: Less-frequent charging, longer battery life

Strategy: Low-voltage device design, auto-clock-gating, auto-power-down management
Enterprise storage low power

- EPA statistics: IDC energy cost doubled every 5 years.
- U.S average Power Usage Effectiveness (PUE) now is 1.8~1.9, when new IDC PUE is 1.3
- China PUE now is 2.0-2.5, new IDC is 1.73

US data center energy use from 2000 until 2020
Source: US Department of Energy, Lawrence Berkeley National Laboratory
Energy waste on enterprise storage

Dynamic Power waste

Static Power waste

(Enterprise) SSD

Front End I/O Interface to Host
Internal Mem
Error Correction
CPU Cluster
Back End I/O Interface to Flash

Interface always toggle
Data Buffer
Data Encryption/Decryption
Bus Connection

Memory always full speed
Module always trigger
Larger Track Cell leakage
Flash IO always power on
What NVMe SSD can do

- PCIe L0, L0s, L1 and L2 power mode
  - L0-Active
  - L0s-Low resume latency standby (ASPM)
  - L1-Higher latency, low power “standby” (ASPM)
    - L1.1-Link common voltage maintained
    - L1.2-Link common voltage not required to be maintained
  - L2-Vaux only, deep-energy-saving state
- NVMe SSD Power state can be defined by different levels.
Low power mode consideration on NVMe SSD

<table>
<thead>
<tr>
<th>Power State</th>
<th>PCIe IO</th>
<th>CPU</th>
<th>Internal Mem</th>
<th>External Mem</th>
<th>Flash IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power State0</td>
<td>Full On</td>
<td>Full On</td>
<td>Full On</td>
<td>Full On</td>
<td>Full On</td>
</tr>
<tr>
<td>Power State1</td>
<td>Full On</td>
<td>Throttling</td>
<td>Throttling</td>
<td>Throttling</td>
<td>Throttling</td>
</tr>
<tr>
<td>Power State2</td>
<td>Full On</td>
<td>Halt</td>
<td>Retain</td>
<td>Retain</td>
<td>Off</td>
</tr>
<tr>
<td>Power State3</td>
<td>L1 Standby</td>
<td>Part off</td>
<td>Retain</td>
<td>Off</td>
<td>Off</td>
</tr>
<tr>
<td>Power State4</td>
<td>L1.2 Standby</td>
<td>Off</td>
<td>Retain</td>
<td>Off</td>
<td>Off</td>
</tr>
<tr>
<td>Power State5</td>
<td>L2</td>
<td>Off</td>
<td>Partly Retain</td>
<td>Off</td>
<td>Off</td>
</tr>
</tbody>
</table>
Advanced consideration on SSD power saving

- Lower track, higher voltage driven cell using for lower leakage power. Especially for algorithm IP
- Auto-Clock Gating for all data path IPs to reduce dynamic power on different operation mode
- Auto-Power Down for all flash IO path IPs to reduce IO power drain.
- High density memory with async-clock design, especially for CPU, ease the whole clock tree generating for low power
- Flexible usage of memory sleep and shutdown mode
- Modularize the functional IPs with different power domains.
- Turn off the SOC Top with minimized retained mem/instruction.
Starblaze SSD low power design

- Active Power < 2.5W
- PCIe L1.2 < 5mW
- Random read up to 600K IOPS
- Random write up to 600K IOPS
- Sequential read up to 3.5GB/s
- Sequential write up to 3GB/s
- Read latency 75us
- Write latency 8.5us
Conclusion

• All low power design strategy applied on mobile device can be applied on SSD controller

• PCIe ASPM mode or PCIe PM mode definition greatly helped us define the high efficiency, low power consumed SSD controller, which means NVMe SSD is more promising than competitors

• Be careful about clock tree, memory, IO
Q & A

• Booth #649