How consumer product like Google's Pixelbook benefit from NVMe Storage?

Zhiping Yang, Ph. D.
Google Inc.
Oct. 17, 2017
Agenda

• Consumer Hardware at Google
• Consumer Product’s Storage Wishlist
• Why PCIe NVMe in Consumer Product?
  • Performance/Power/Thermal management
• Key Focus Areas for 1113 BGA SSD
  • Small size, UFS/eMMC backward compatibility
• Cost/Power Reduction, More Choices/Features
• Conclusion
Consumer Hardware at Google
Introducing a few new things made by Google
Google Pixelbook: Designed to do everything you want to do.

- Solid state drive (128GB, 256GB, or 512GB NVMe)
- A 7th Gen Intel® Core™ processor
- All-day battery
- 8GB or 16GB RAM
- Charge for 15 minutes to get up to 2 hours of use
Consumer Product’s Storage Wishlist

- Low Cost
- Small Space
- Smooth Transition
- High Performance
- Smart Thermal/Power Management
- Low Power Consumption
Why PCIe NVMe in Consumer Product?

- Scalable software stack across client platforms
- One less platform interface to develop and validate
- Flexibility in system design
- Low latency interface
- Drivers for all major OSes
- Host Memory Buffer provides DRAM-less solutions
- Scalability to next generation NVM
- PCIe is widely adopted and well positioned for the future

PCIe and NVMe have clear benefits that are seen in the datacenter and client. These benefits can also be realized in consumer product.
Bandwidth/Latency Improvement With PCIe

PCIe NVMe SSDs provides highest throughput and lowest latency

*UFS, eMMC and names associated with it are property of JEDEC.
Bandwidth from each interface are max theoretical that account for encoding overhead.
Latency values are estimated based on storage access to and from memory.
Latency Improvement Details

SSD NAND technology offers ~100X reduction in latency versus HDD

NVMe* eliminates 10-20 µs of latency today

Next Gen NVM needs NVMe to deliver 4KB operations in under 10 µs

Source: Storage Technologies Group, Intel. Comparisons between memory technologies based on in-market product specifications and internal Intel specifications. *Other names and brands may be claimed as the property of others.
NVMe SSD Power

Two metrics of power: Active and Standby

- **Active Power**: Power consumed when doing something
- **Standby Power**: Power consumed when doing nothing

Active Power consumption is about energy consumed. PCI Express\(^*\) (PCIe\(^*\)) is competitive. L1 Substates in PCIe reduce standby power to range acceptable for mobile.

NVMe offers APST (Autonomous Power State Transition) to manage latency and power trade-offs.

<table>
<thead>
<tr>
<th>Item</th>
<th>PCIe Gen3</th>
<th>PCIe Gen2</th>
<th>UFS Gear3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Speed [Gbps]</td>
<td>8</td>
<td>5</td>
<td>5.83</td>
</tr>
<tr>
<td>PHY Overhead</td>
<td>128/130,1GB/s</td>
<td>8/10, 500MB/s</td>
<td>8/10, 583MB/s</td>
</tr>
<tr>
<td>Active Power [mW]</td>
<td>60 (L0)</td>
<td>46 (L0)</td>
<td>58 (HS)</td>
</tr>
<tr>
<td>Standby Power [mW]</td>
<td>0.11 (L1.2)</td>
<td>0.11 (L1.2)</td>
<td>0.2 (Hibern8)</td>
</tr>
<tr>
<td>MB/mJ (higher better)</td>
<td>14-18</td>
<td>8-12</td>
<td>8-12</td>
</tr>
</tbody>
</table>

UFS and names associated with it are property of JEDEC.

1. pci-sig.com: "L1 PM Substates with CLKREQ, Revision 1.0a"
2. Source: SanDisk\(^*\). Data based on PHY power estimates of PCIe vs. MPHY.

NVMe SSDs can be power competitive in Small Form Factor.
NVM Express* SSD Thermals

- Two ways to manage thermals in NVMe
  1. Change power states: This allows the host system to run at a higher power/performance until the device or system becomes limited.
  2. Host controlled Thermal Management: This allows the system to decide when the device needs to throttle from a thermal perspective.

<table>
<thead>
<tr>
<th>Workload</th>
<th>Perf / Power</th>
<th>Time to Throttle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst case</td>
<td>1600MB/s @ 4W</td>
<td>~60s to heavy throttle</td>
</tr>
<tr>
<td>Sustained</td>
<td>1000MB/s @ 2W</td>
<td>Light throttle only</td>
</tr>
<tr>
<td>TDP</td>
<td>500MB/s @ &lt;1W</td>
<td>Never</td>
</tr>
</tbody>
</table>

Source: NVM Solutions Group, Intel. Simulated Thermal constraints based on the PCIe* BGA 16x20mm package

NVMe SSDs provides burst performance while being able to throttle back if needed.
Key Focus Areas for 1113 BGA SSD

- Reduces the platform area in half compared to 16x20
  - 150mm² vs. 320mm²
- 11.5x13 only supports x2 PCIe for now
- The same MLB can support either NVMe or eMMC¹/UFS¹
  - It provides a smooth transition from eMMC/UFS based storage to NVMe storage.
  - eMMC/UFS and NVMe together provide good performance, price, features, and capacity coverage to meet different customers’ needs.
- Optional SPI interface for better system integration

1. eMMC and UFS specs are managed by JEDEC

NVMe innovations enable scaling into smaller form factors delivering new differentiated platforms.
Cost Reduction

- Controller cost has to go down
  - Cost is a function of NAND, controller, and package/assembly/test.
  - Higher capacity SSD means a lower percentage of cost from controller
- Cost vs Performance trade-off
  - Lower performance & Less feature mean lower cost.
- Volume is critical to drive cost down
- Host Memory Buffer (HMB) (less cost with more performance)
  - Support in Linux kernel has been proposed

PCIe BGA SSD cost can be competitive to eMMC depending on capacity and performance targets
Power Consumption Reduction

- Power/battery life is critical for consumer product
- PCI L1.2 power is still higher than existing eMMC/UFS solutions
- Power vs Performance trade-off
  - Lower performance & less feature mean lower power.
- Optimize NVMe/PCIe PHY design for consumer product
  - Shorter distance, within the same device/board
- Looking for more power saving features

PCIe BGA SSD power can be competitive to eMMC/UFS
More Suppliers & Choices

- LongSys’s 1113 BGA NVMe announcement
  - Spec was approved in February 2017 by PCI-SIG
- More vendors in the pipeline for FY17/18
Conclusion

- 11x13 BGA NVMe consumer devices are coming!
  - New usages require better storage.
  - Offers many advantages over existing solutions

- Continue to push for lower cost, lower power consumption, and more choices.

- Call to action
  - Consider 11x13 PCIe BGA devices for your consumer product
  - Contribute to 11x13 development to make it more attractive
  - Support of smaller capacity (32/64GB)
Google Consumer Hardware Is Hiring!

- [https://careers.google.com/hardware/](https://careers.google.com/hardware/)

Our Consumer Hardware team researches, designs, and develops new technologies and hardware to make users' interaction with computing faster, more powerful, and seamless. Whether finding ways to capture and sense the world around us, advancing form factors, or improving interaction methods, our Consumer Hardware team is making people's lives better through technology.
Backup
Available NVMe Form Factors

- M.2 Module: Capacity, Size, cost, backward compatibility
- 16x20mm BGA SSD: Capacity, Size, backward compatibility
- 11.5x13mm BGA SSD: **Right Solution**
Why Size Matters?

- 16x20mm does not fit in smaller form factor devices.
  - Need for a smaller performance SSD
- M.2 storage ~15% of platform area
- Allows for smaller motherboard and better product

Credit: iFixit
# SPI ROM Interface for Secure Boot

<table>
<thead>
<tr>
<th>Interface</th>
<th>Signal Name</th>
<th>I/O</th>
<th>Description</th>
<th>IO Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPI ROM</td>
<td>WP_L(^2)</td>
<td>I</td>
<td>Write protect signal to prevent writes from occurring to SPI NOR. Active low.</td>
<td>1.8 V</td>
</tr>
<tr>
<td></td>
<td>SPI_CLK</td>
<td>I</td>
<td>SPI clock. Max frequency is 50MHz.</td>
<td>1.8 V</td>
</tr>
<tr>
<td></td>
<td>SPI_MOSI</td>
<td>I</td>
<td>Master Out Slave In signal for SPI NOR.</td>
<td>1.8 V</td>
</tr>
<tr>
<td></td>
<td>SPI_MISO</td>
<td>O</td>
<td>Master In Slave Out signal for SPI NOR</td>
<td>1.8 V</td>
</tr>
<tr>
<td></td>
<td>SPI_CS_L</td>
<td>I</td>
<td>Chip select for SPI NOR. Active low.</td>
<td>1.8 V</td>
</tr>
<tr>
<td></td>
<td>SPI_18</td>
<td>I</td>
<td>+1.8 V supply. Optional voltage supply if SPI NOR included in package.</td>
<td>1.8 V</td>
</tr>
</tbody>
</table>

1 Optional  
2 WP_L can be used for NVMe hardware write protection as well.