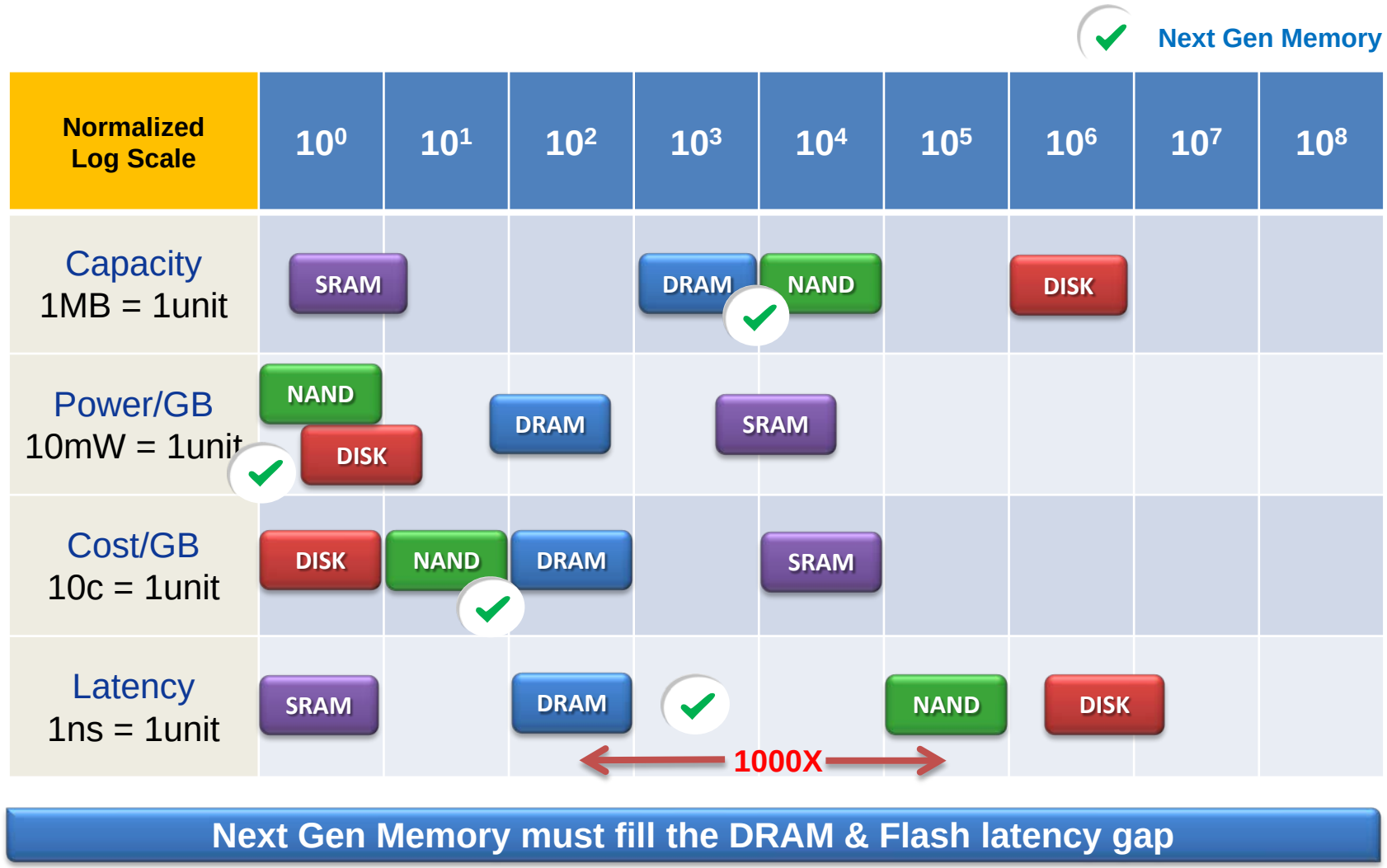
Fine-tuning the Flash Engine

Flash on the Memory Bus Panel

Rakesh Cheerla, SMART Modular Technologies

Comparison of Memory Technologies



Next Gen Memory must fill the DRAM & Flash latency gap

Working with Flash

High Latency, Low BW		4K Pages	Block Erase
100µs read 1.5ms write	1/10X DIMM BW	Byte vs. 4K Page	Erase: 5ms Latency (Worst): 100ms

Simplify Design
1. Remove SATA/PCIe/NVMe
2. Append Writes

Improve DRAM hit rate
1. Flash aware data layout
2. Application aware prefetching

Increase Parallelism

Application Acceleration

Better Caching
1. Large RW cache
2. Write Merging & Data Forwarding

Write less to flash
1. Late Writeback
2. Compress / Dedupe

Erase optimization

Cache erased page

Alternatives
1. Memory Mapped PCIe M.2 Cards
2. Intel Xeon + FPGA Socket ?

Architect around flash limitations for high performance DIMM bus

Summary

- Seamless transition to Next-Gen Memory
- Innovate to mitigate flash challenges
- Think Applications!