Making the Right Power/Performance Tradeoffs for PCIe/NVMe SSDs

Rick Huang

SSD Product Marketing Manager

SiliconMotion Inc.

SSD Getting “Hot”

SATA SSD replacing HDD

- Seq. Read >500 MBps
- Seq. Write >500 MBps

PCIe SSD running faster

- Seq. Read >3000 MBps
- Seq. Write >1500 MBps

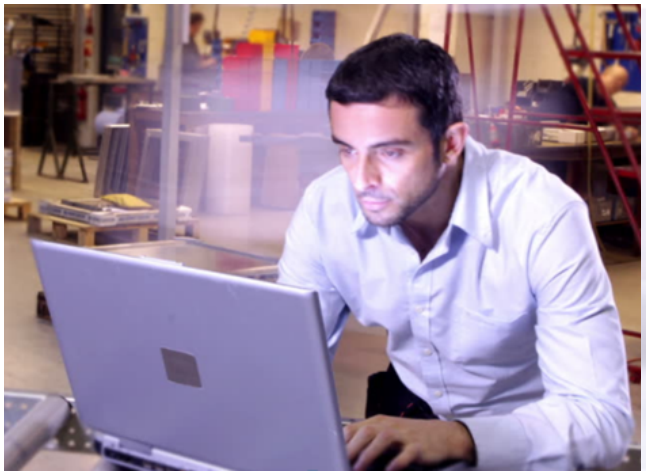
Making SSD "Cool"

- SATA to PCIe, power budget still constrained
  - E.g. max 2.5A @3.3V (**8.25W**) or max 1.6A @5V (**8W**)
- Low power SSD design matters
  - Not suffering thermal issue
  - Extending battery lifetime

Major Powers of Consideration

Active

* MBps per Watt



Sleep

* Lowest power

Average

* Average power

Key Power Index

	SATA SSD (512GB)	PCIe SSD (512GB)
Active (MBps/W) * 100% 128KB Seq. Write	150 e.g. 528MBps @3.5W	?
Sleep (mW) * DEVSLP or L1.2	2.0	?
Average (mW) * MobileMark 2014	96	?

Power Consumption in ASIC

$$Power = P_{Dynamic} + P_{Static}$$

$P_{Dynamic}$: $c \cdot V^2 \cdot f$ — Dominate active power

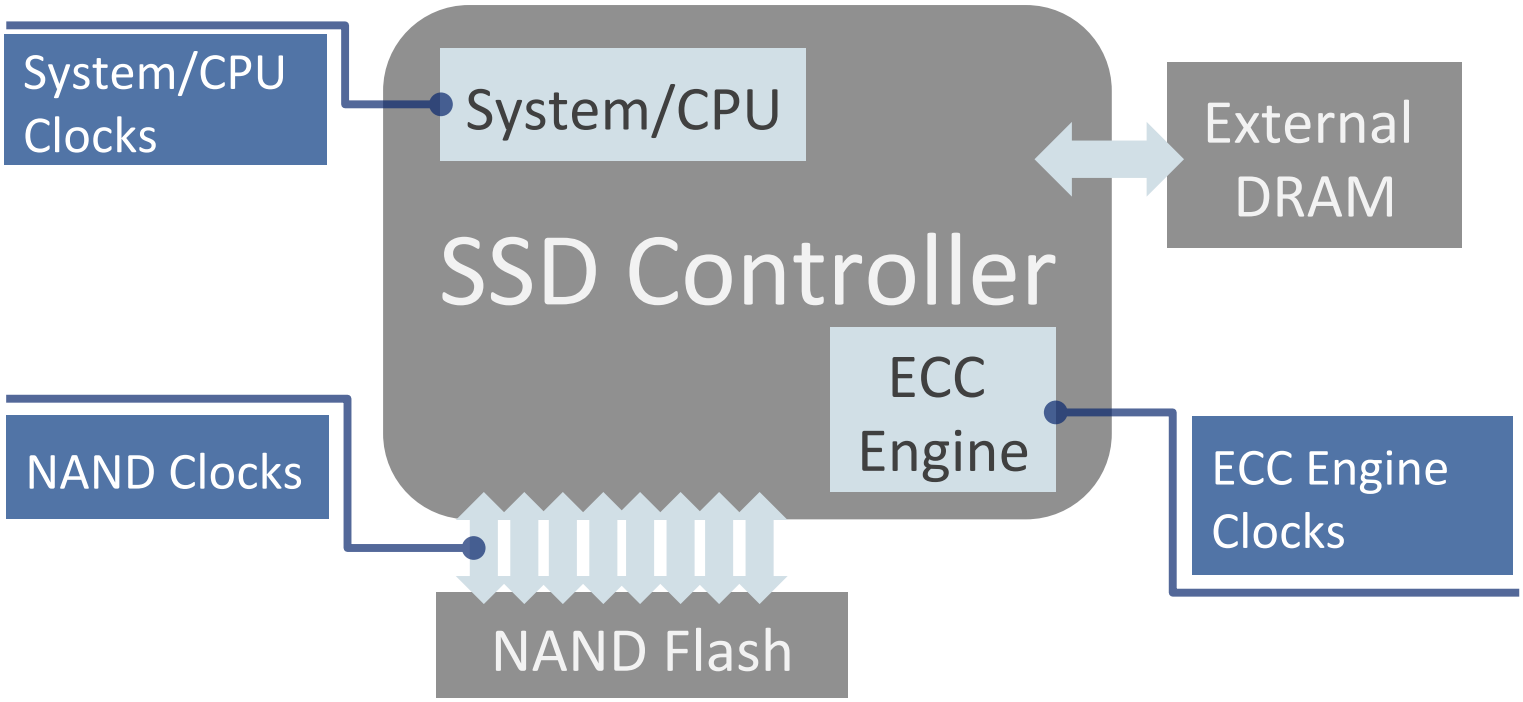
P_{Static}: $V \cdot I_{leak}$ — Dominate sleep power

Active Power – Advanced Process

- $P_{Dynamic} \sim= c \cdot V^2 \cdot f$
  - Lower supply voltage reduces power significantly
  - Higher development cost

	55nm	40nm	28nm	16nm
Voltage (V)	1.2	1.1	0.9	0.75
Power	100%	84%	56%	39%

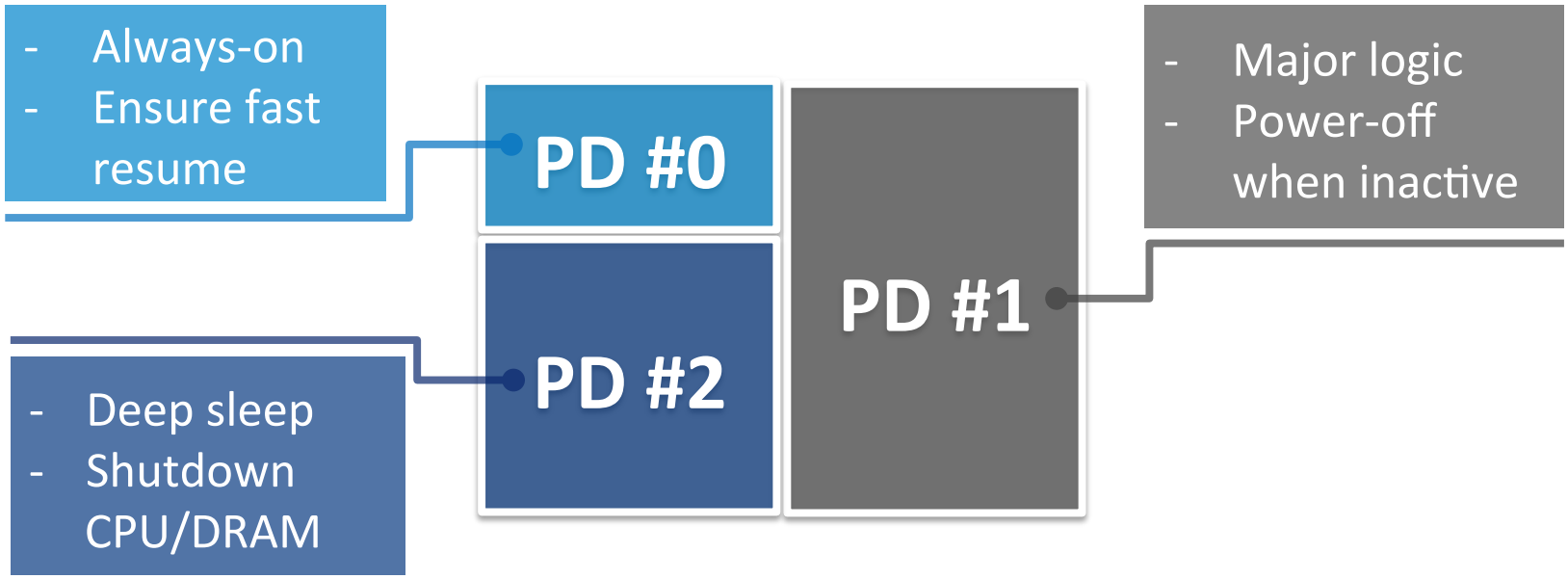
Active Power – Changing/Gating Clocks



Sleep Power – Leakage Power

- P_{Static} ~= $V \cdot I_{Leak}$
  - Lower down supply voltage to reduce power
- Advance process leads to higher cell leakage
  - 55nm (1X) -> 40nm (>1.6X) -> 28nm (>3.2X) @25C
  - Reduce I_{Leak} by power domain partition

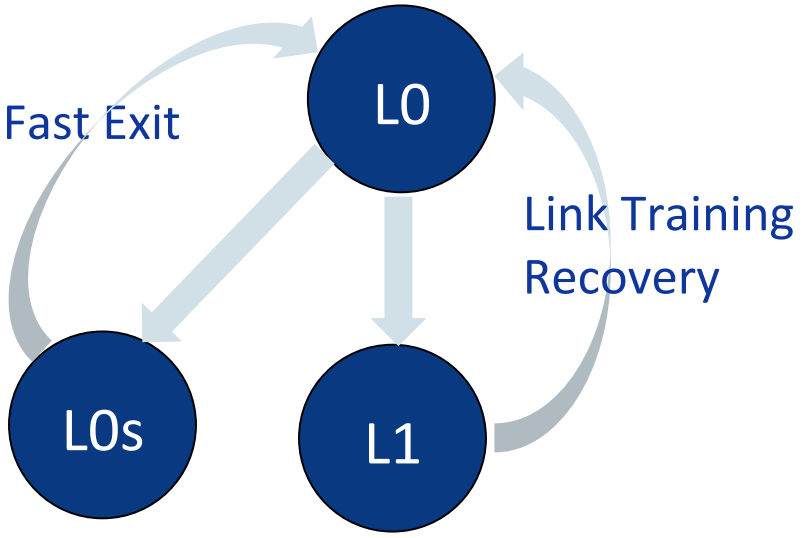
Sleep Power – Power Domains Partition



PCIe/NVMe Power Management

- ASPM – Active State Power Management
  - Dynamic control of PCIe link power reduction
- APST – Autonomous Power State Transition
  - Intelligent control of active and sleep power states

ASPM – PCIe Link Power Reduction



- Hardware-autonomous
  - Dynamic link power reduction without SW intervention
  - Transition time of us level
- Need robust PCIe PHY design and good compatibility
  - Adjust to adapt to host platform

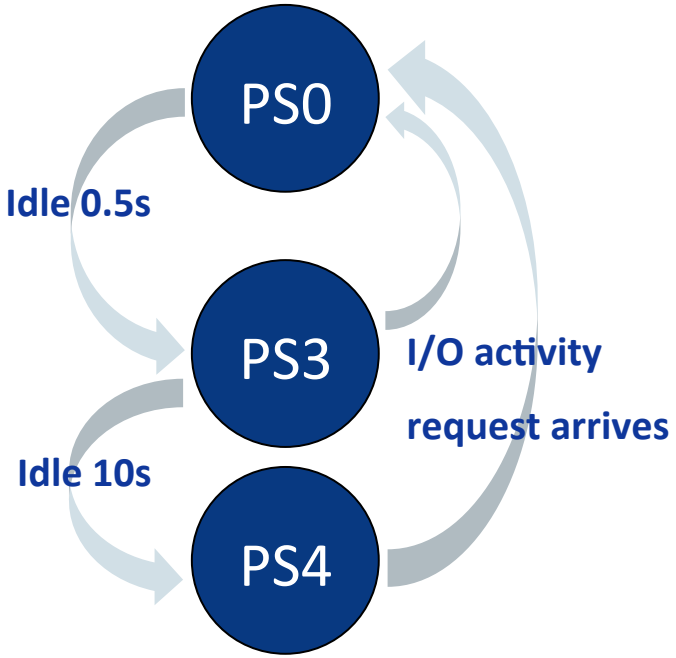
APST – Active/Sleep Power State Transition

Power State Table

Power State	Operational State	Max. Power	Exit Latency
PS0	**Yes**	5W	-
PS1	**Yes**	4W	<10us
PS2	**Yes**	3W	<10us
PS3	**No**	50mW	<10ms
PS4	**No**	2.5mW	<50ms

APST Table

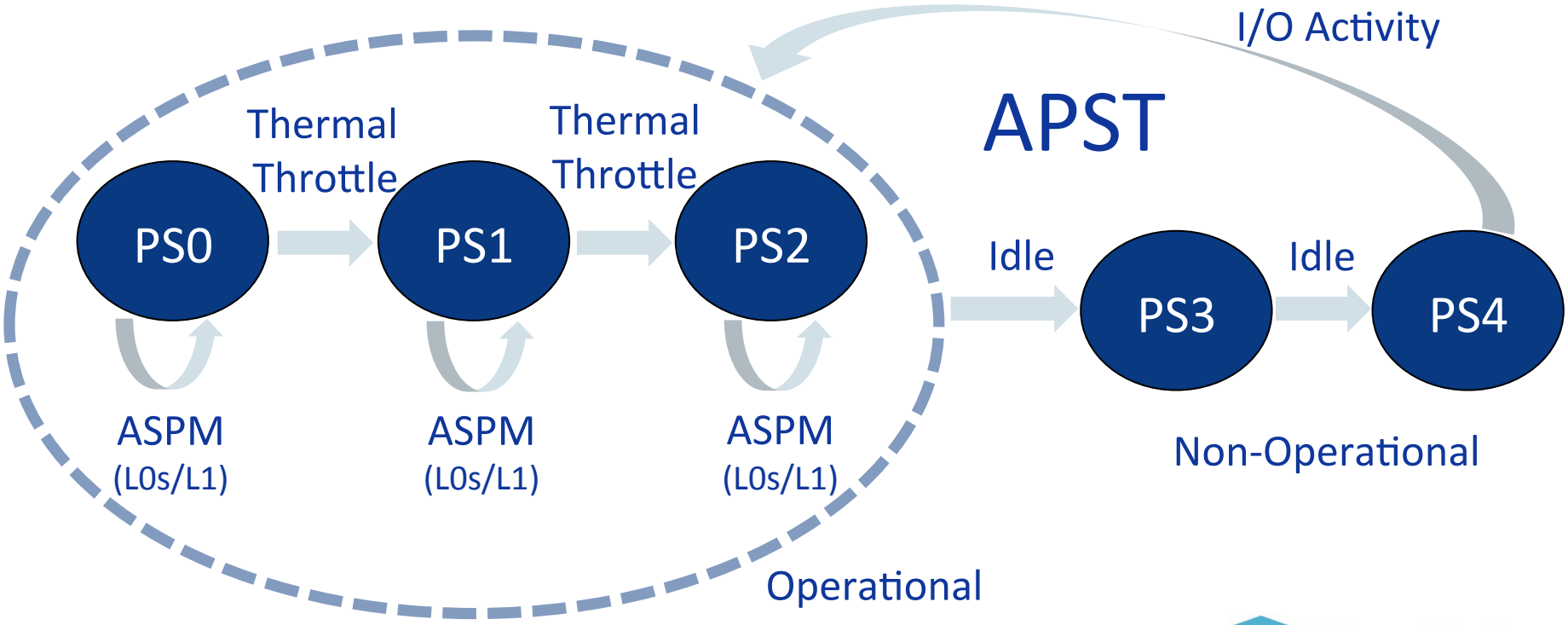
Idle Time Prior to Transition	Transition-to Power State
500ms	PS3
500ms	PS3
500ms	PS3
10,000ms	PS4
-	-



Example of Power States Design

NVMe Power State	PCIe Link Status	Operational	Max Power	Entry Latency	Exit Latency	Note
PS0	**L0 / L0s / L1**	**Yes**	5W	-	-	Full-speed Active
PS1	**L0 / L0s / L1**	**Yes**	4W	<10us	<10us	Light Throttle
PS2	**L0 / L0s / L1**	**Yes**	3W	<10us	<10us	Heavy Throttle
PS3	**L1 / L1.1 / L1.2**	**No**	50mW	<5ms	<10ms	Light Sleep
PS4	**L1.2**	**No**	2.5mW	<10ms	<50ms	Deep Sleep

Power States Transition with APSM/APST



Excellent Power Index of PCIe/NVMe

	SATA SSD (512GB)	**PCIe SSD** (512GB)
Active (MBps/W) * 100% 128KB Seq. Write	150 e.g. 528MBps @3.5W	481 e.g. 2213MBps @4.6W
Sleep (mW) * DEVSLP or L1.2	2.0	2.3
Average (mW) * MobileMark 2014	96	55

Summary

- Low power PCIe/NVMe SSD available
- Low power ASIC/FW techniques for active/sleep
  - Advanced process, gating/changing clocks, and power domains
- PCIe/NVMe power management with ASPM/APST
  - High speed active mode with moderate power to serve I/O
  - Autonomous low power states transition to save power

Thank You

http://www.siliconmotion.com