How UFS Storage is Evolving for On-Device AI and Autonomous Vehicles

Kevin Hsu
Sr. Applications Engineering Manager
KIOXIA America, Inc.

FMS the Future of Memory and Storage

KIOXIA

Agenda

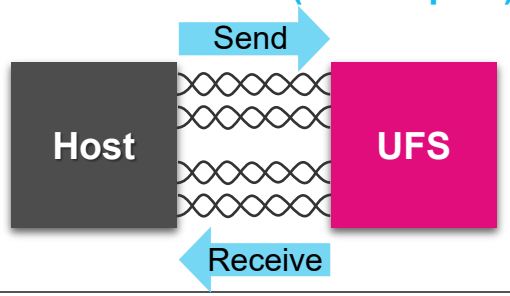
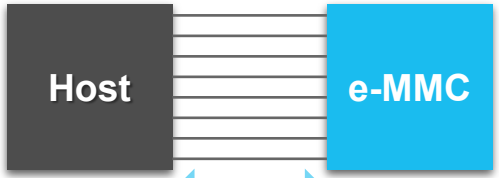
1. UFS introduction and advantages
2. Storage trends in UFS applications
3. New features benefiting automotive
4. On-Device AI
5. Final thoughts

FMS the Future of Memory and Storage

Advantages of UFS[1]

UFS is the leading storage technology for mobile applications

- Fast, efficient serial interface
- Low power consumption optimized for power and performance
- Scalability to higher capacities and speeds

		UFS	e-MMC[2]
Interface	Bus	Serial Interface (Full Duplex) 	Parallel (x8) Interface  Send or Receive Half Duplex
	Speed	**Max. 2320MB/s** (HS-G4(11.6Gbps) x 2lanes with UFS 3.0/3.1) **Max. 4640MB/s** (HS-G5(23.2Gbps) x 2lanes with UFS 4.0/4.1)	**Max. 400MB/s** (HS400 with e-MMC 5.0/5.1)
	Pin Count	1 lane : 6 (4 I/O and 2 control) 2 lane : 10 (8 I/O and 2 control)	11 (8 I/O and 3 control)
Command Set		SCSI	MMC[3]

1. Universal Flash Storage (UFS) is a product category for a class of embedded memory products built to the JEDEC® UFS standard specification. JEDEC is a registered trademark of JEDEC Solid State Technology Association. 2. Electrical component qualification requirements defined by the AEC (Automotive Electronics Council). 3. The MultiMediaCard (MMC) standards are managed by JEDEC. All images and/or graphics within this slide are the property of Kioxia Corporation (KIOXIA) and are reproduced with the permission of KIOXIA.

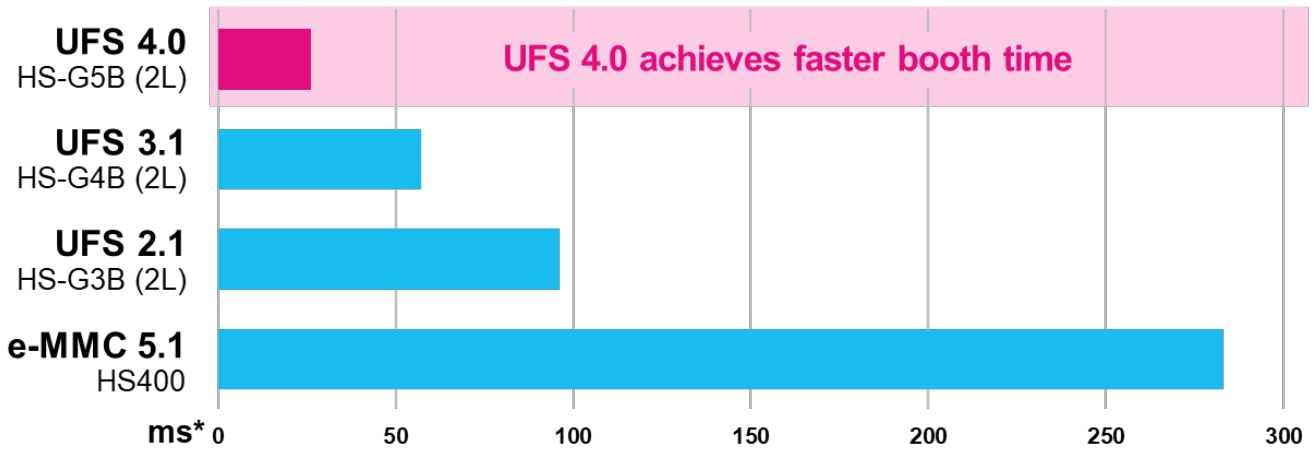
Storage Trend of UFS

UFS is the leading storage interface in smartphones and automotive applications

Application	2022	2023	2024	2025	2026	2027	2028
Smartphones (High End)	UFS	UFS	UFS	UFS	UFS	UFS	UFS
Smartphones (Low End)	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC
Automotive	UFS	UFS	UFS	UFS	UFS	UFS	UFS
Automotive	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC	e-MMC
Automotive							SSD

Key Improvement – Faster Boot Time

Boot Time Compared to Previous Generation Devices



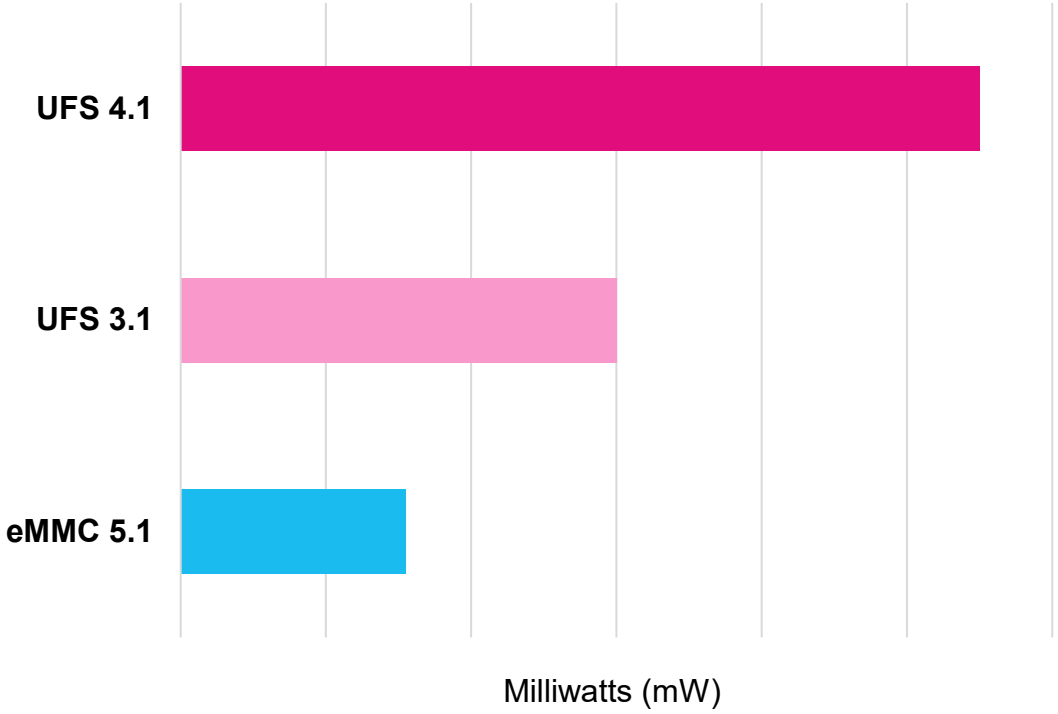
Benefits

- Near instantaneous system boot up in in-vehicle-infotainment systems and smartphones
- Improves user experience
- Enhances automotive cabin experience by reducing latency
- Additional feature called High Speed Link Startup Sequence (HS-LSS) in UFS 4.0 further improves initialization

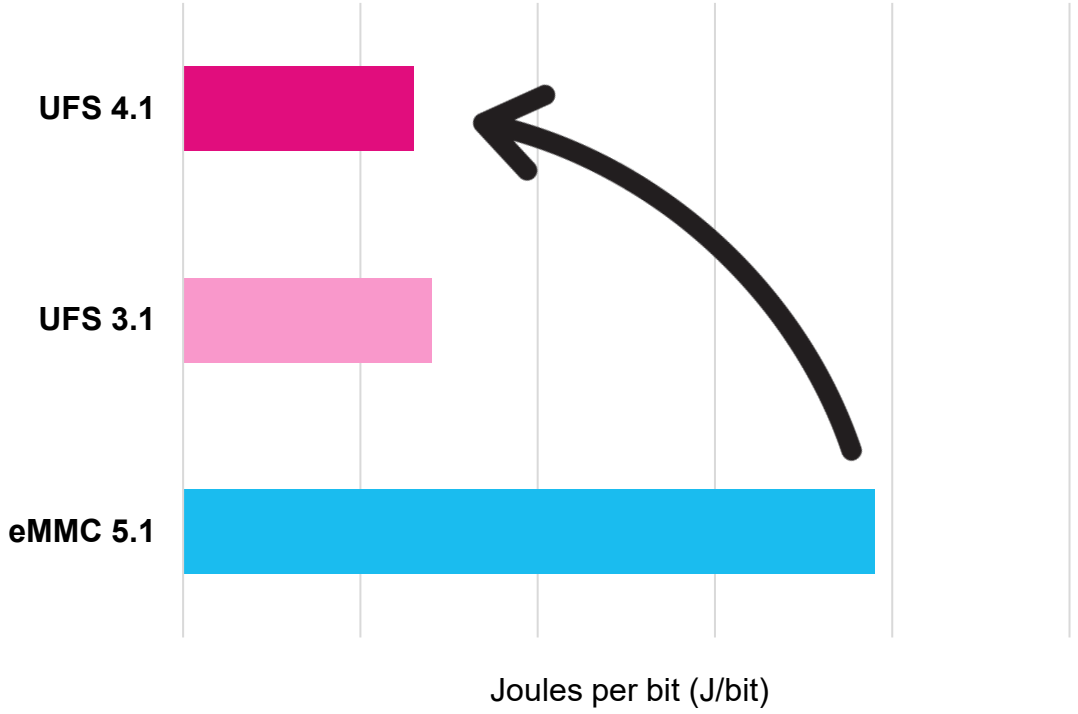
Key Improvement – Lower Energy Consumption

Although overall power consumption proportionally increases with faster interface, the overall energy per bit consumption is lower

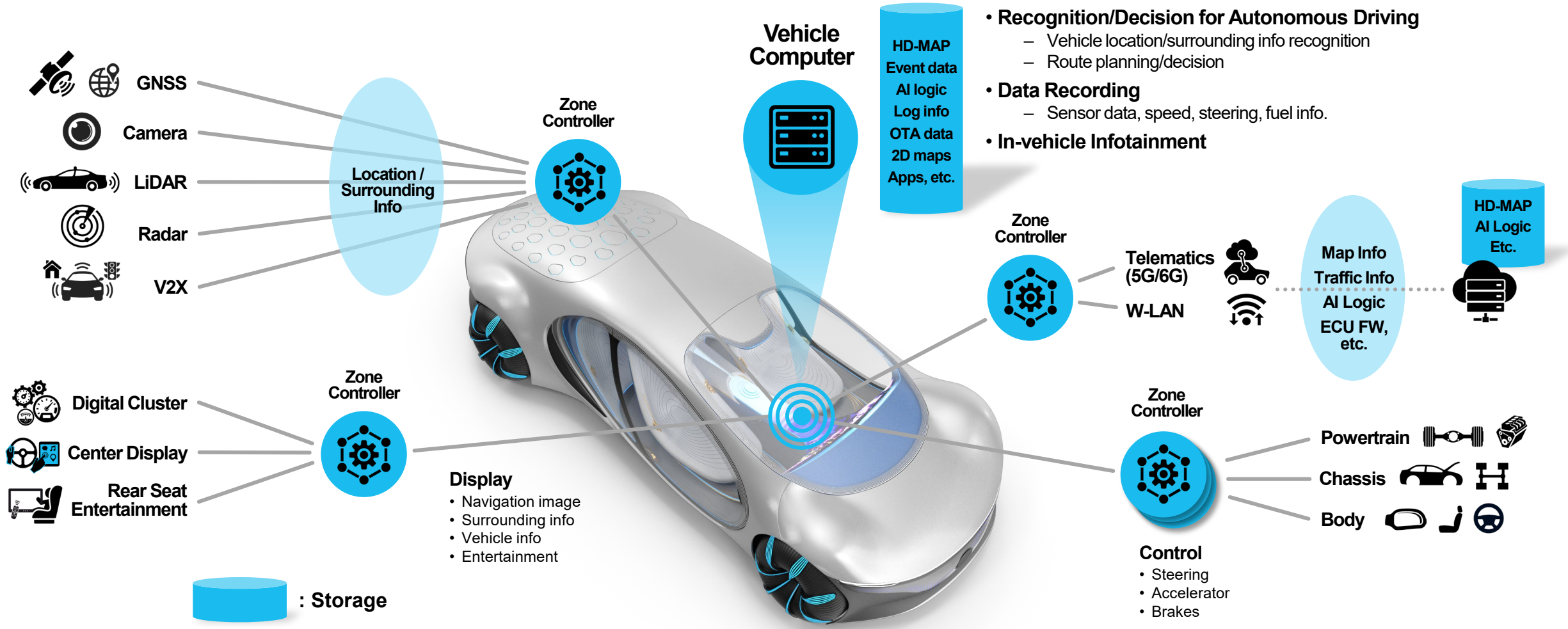
Power Consumption



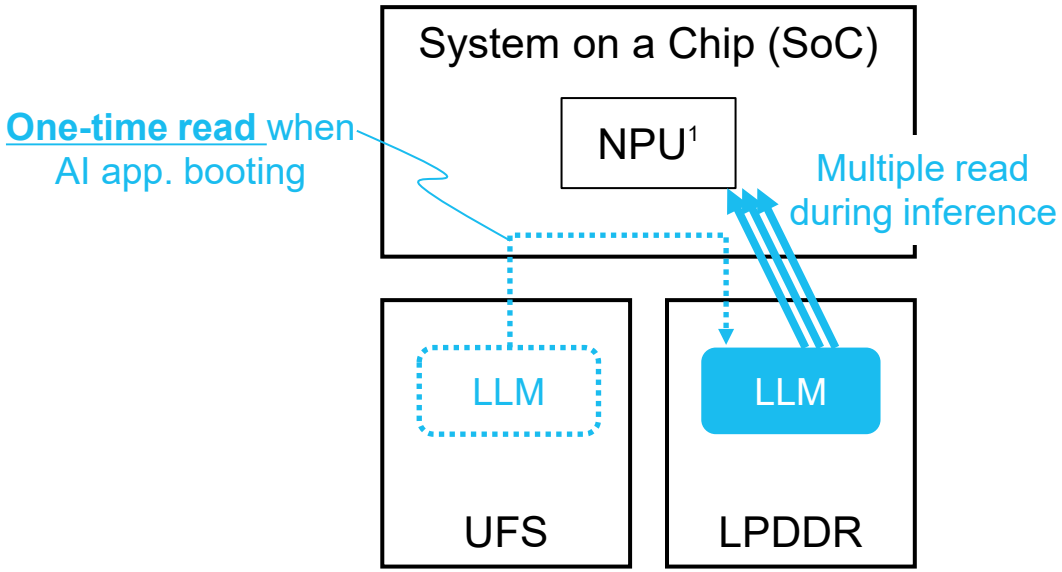
Energy per bit Consumption



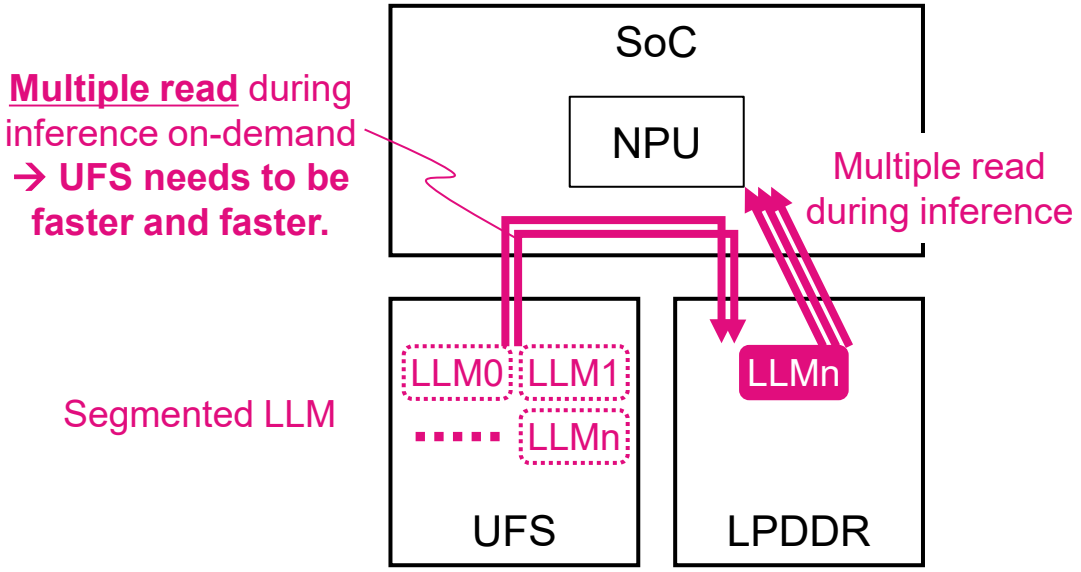
Storage Device in Integrated Automotive System



UFS Usage in On-Device AI



Existing Implementation



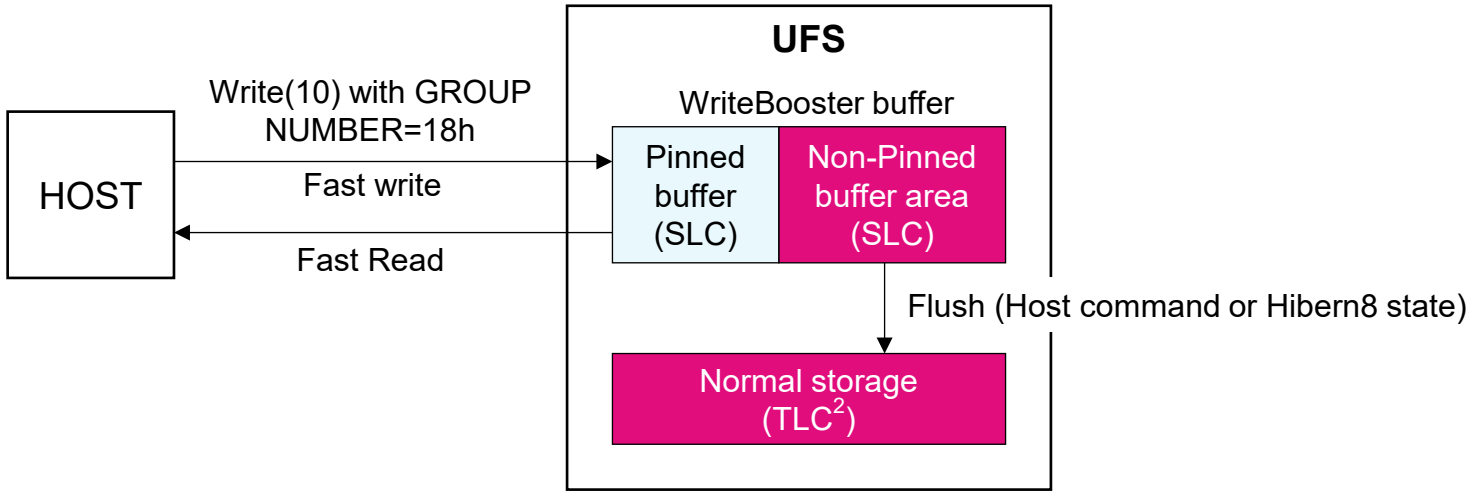
Future Implementation

1. Neural Processing Unit (NPU)

UFS ver.4.1 Major Feature Overview : Pinned Partial Flush Mode of WriteBooster

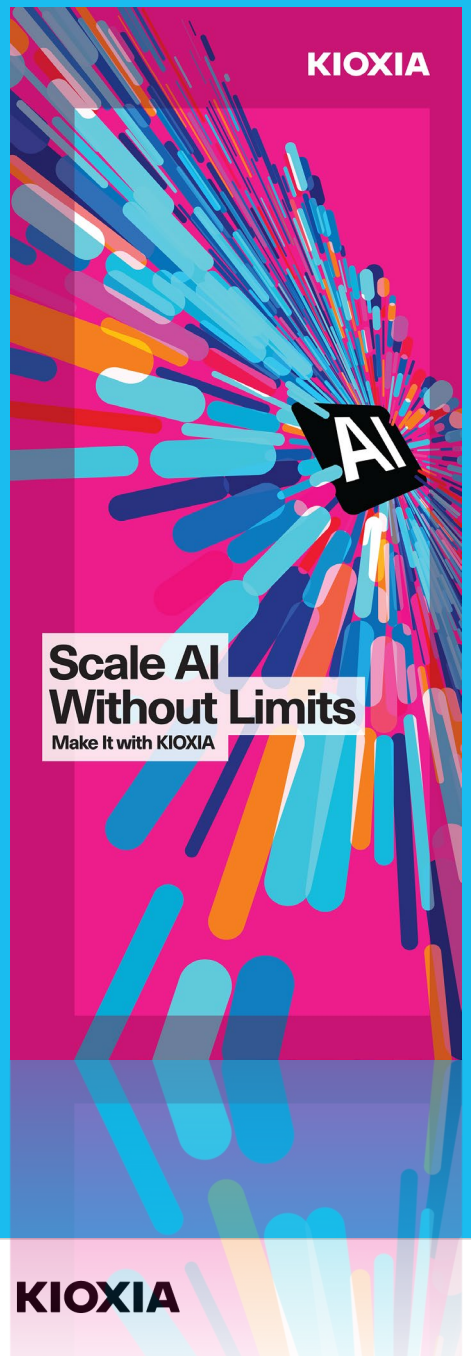
Read latency can be improved by writing frequently accessed data to pinned SLC buffer

- UFS device provides pinned buffer in the WriteBooster buffer
- Data written by SCSI WRITE(10) command with GROUP NUMBER=18h is stored in Pinned buffer
- Data that was written to pinned buffer will not be flushed by WriteBooster flush and be kept in SLC[1] mode so that it improves read latency
- The pinned data is released by setting 00h “No partial flush “ to bWriteBoosterBufferPartialFlushMode or set 1b to fUnpinEn



1. Single-level Cell (SLC). 2. Triple-level Cell (TLC).

Final Thoughts



As mobile and automotive applications increase adoption of LLMs, the need for high bandwidth storage will be critical for storing LLMs to reduce DRAM size (cost and power)

UFS is poised to maintain its leadership as the preferred storage as it is optimized for power, performance, and offers new innovative features to address on device AI

Similar applications such as drones, factory automation, and robotics can also benefit from UFS

Please visit the KIOXIA booth for more information about our latest UFS v4.1!

Definition of capacity: KIOXIA defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1GB = 2^30 = 1,073,741,824 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, such as Microsoft Operating System and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

All company names, product names and service names may be trademarks of their respective companies.

Images are for illustration purposes only.