Flash Memory Summit

Zoned Storage for UFS

Presenter: Bart Van Assche, Google.

About Myself

- Working on block storage for the Google Pixel team.
- Member of the JEDEC UFS committee.
- Active contributor to the Linux kernel.
- Obtained a Ph.D. in computer science.

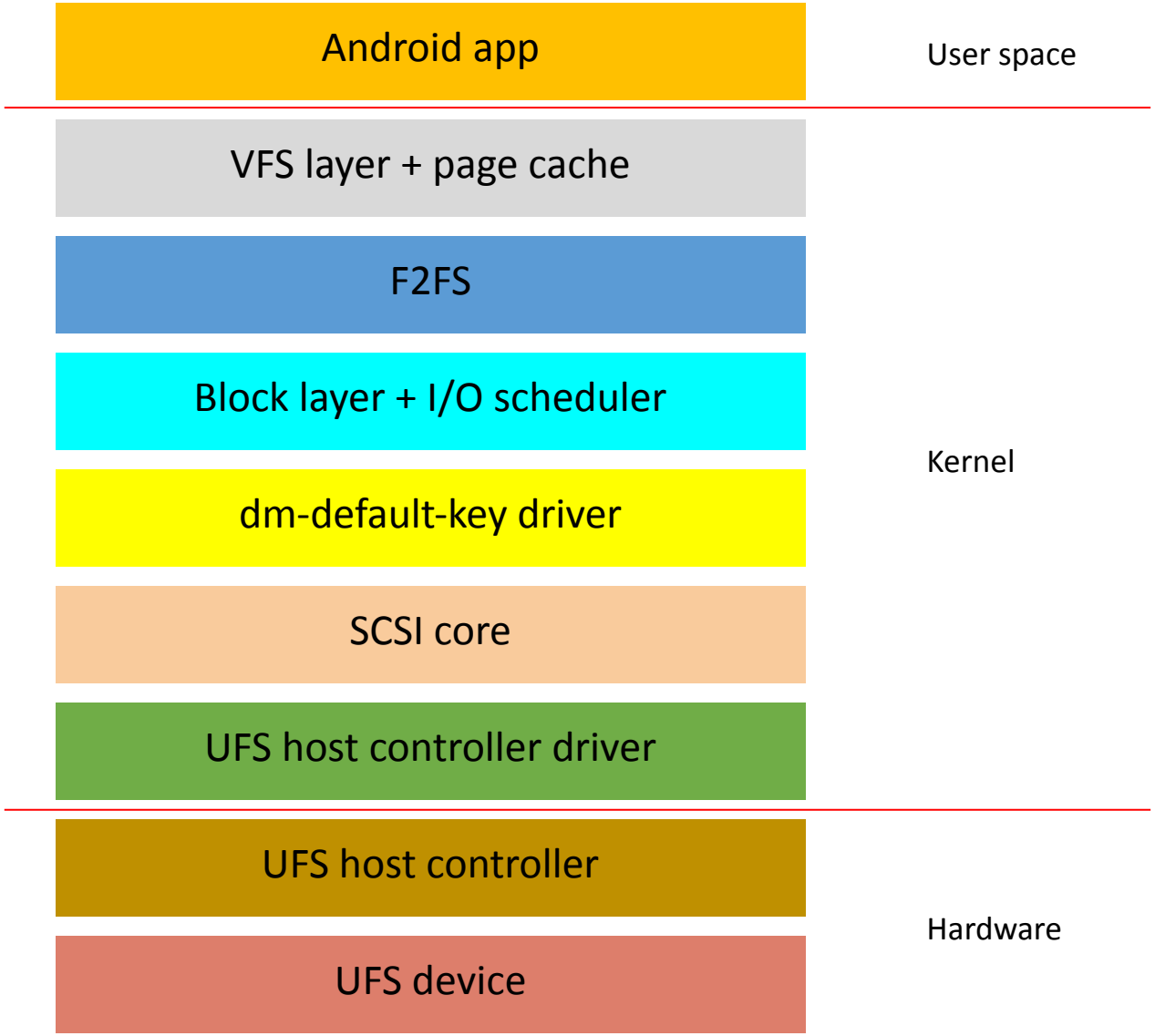
Universal Flash Storage (UFS)

- Flash storage specification for mobile phones, digital cameras and consumer electronic devices.
- Supports high data transfer speeds, low power consumption and small form factors.
- Improves upon eMMC and SD standards.
- The UFS 1.1 standard was published in 2012.

Differences with enterprise SSDs

- **Form factor**: UFS devices are smaller.
- UFS devices are optimized for low **power** consumption.
- Enterprise SSDs **cost** more than UFS devices.
- Enterprise SSDs offer higher **performance.**
- Enterprise SSDs have a higher **capacity**.

Block storage in Android (1/2)



Block storage in Android (2/2)

- F2FS is a log-structured file system and hence performs garbage collection.
- UFS devices use NAND flash.
  - NAND flash has append/erase characteristics.
  - UFS devices maintain a Flash Translation Layer (FTL).
  - UFS devices perform garbage collection internally.

Block storage challenges

- UFS device L2P (logical to physical) table may exceed SRAM size. This may double the number of I/O operations for random I/O.
- UFS device garbage collection may cause latency spikes.

Solution: zoned storage

- Larger L2P entries and hence a much smaller L2P table. L2P table fits again in SRAM. Hence, I/O bandwidth increases and I/O latency decreases.
- Host software can optimize read performance by allocating a contiguous LBA range per file.
- UFS device overprovisioning is no longer required. Filesystem overprovisioning is still required.
- Garbage collection is moved from the UFS device to the host. Hence, write amplification is reduced.

Plan - JEDEC and T10

- JEDEC Zoned UFS (ZUFS) standard has been completed on 2023-07-25. The approach is based on ZBC-2:
  - One zoned logical unit per UFS device.
  - All zones in the zoned logical unit have the sequential write required zone type.
  - All zones in the zoned logical unit have the same size.
  - No gap zones.
- Work with T10 on adding data temperature support in SBC-5. Data temperature support is expected to be added in the next SBC-5 draft (2023).

Zoned storage support in the Linux kernel

- NVMe supports writing at the write pointer and also supports a zone append operation.
- SCSI only supports writing at the write pointer.
- Linux kernel block layer may reorder any pair of pending commands.
- mq-deadline scheduler serializes zoned writes per zone.
- Restricting the queue depth to one per zone results in suboptimal performance.

Plan - Linux kernel

- Reduce reordering
  - Modify the Linux kernel block layer such that reordering is reduced.
  - Modify the UFS host controller driver such that it no longer reorders SCSI commands.
- If the UFS device reports that write commands have been reordered (UNALIGNED WRITE), resubmit these commands in LBA order.

Summary

- Standardization of zoned storage support for UFS devices nears completion.
- Linux kernel zoned storage performance is being improved.

Are there any questions?