Accelerator-integrated Storage I/O

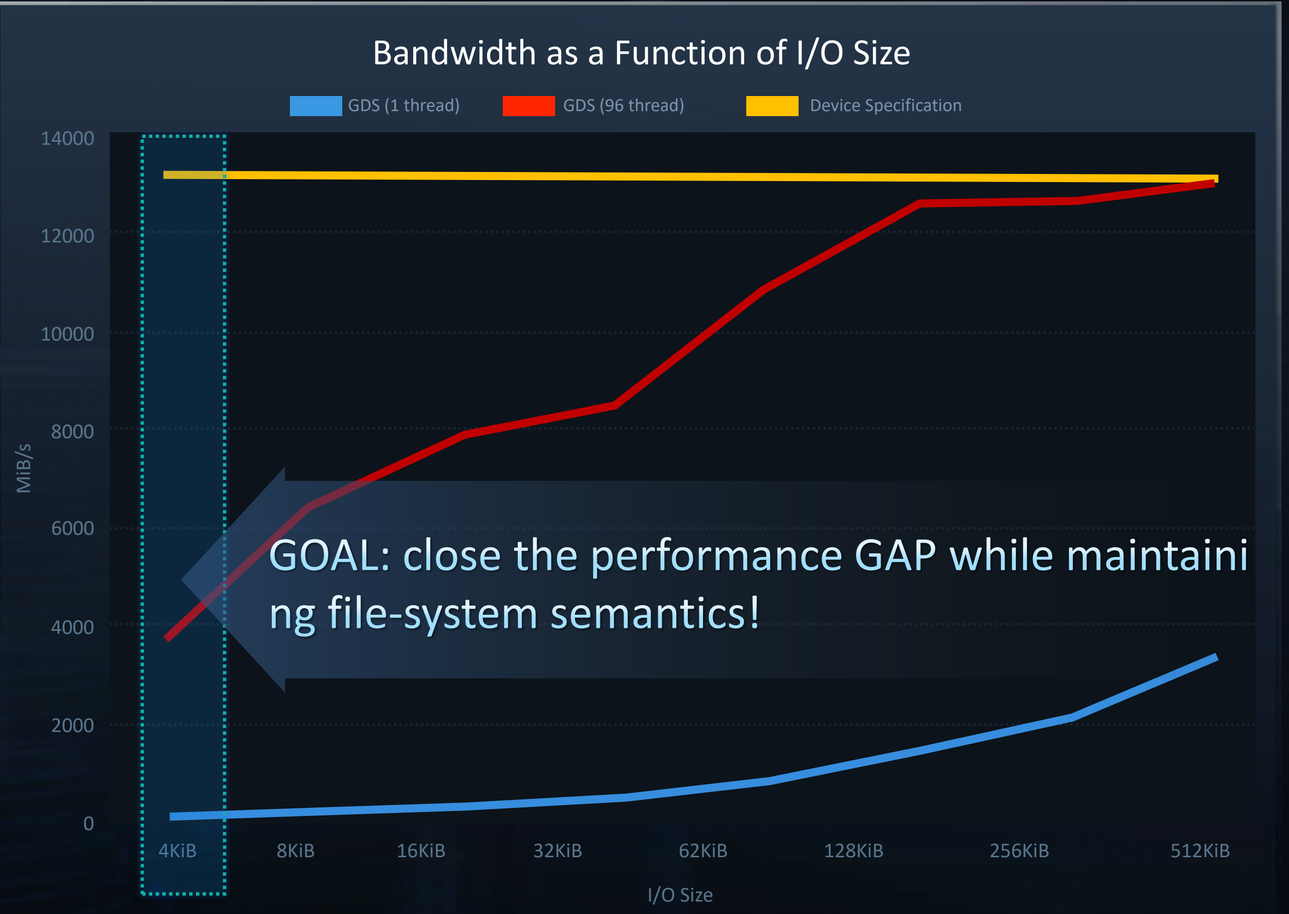
Eliminating software bottlenecks through HW/SW co-design
enabling efficient and scalable I/O

The Problem: Small I/O Sizes

Demand is surging for efficient high-rate access with small I/O size <= 4K. Stemming from training with large amounts of small files, inference vector-embeddings, VectorDB search and graph-neural-networks.

All use-cases hitting the Achilles heel of the state-of-the-art in the consumption of storage for GPUs.

- Large amounts of small I/O
- Software-bottlenecks
- Hindered scalability
- Impacted use-cases include:
  Training, Prediction, Vector Search, and GNN

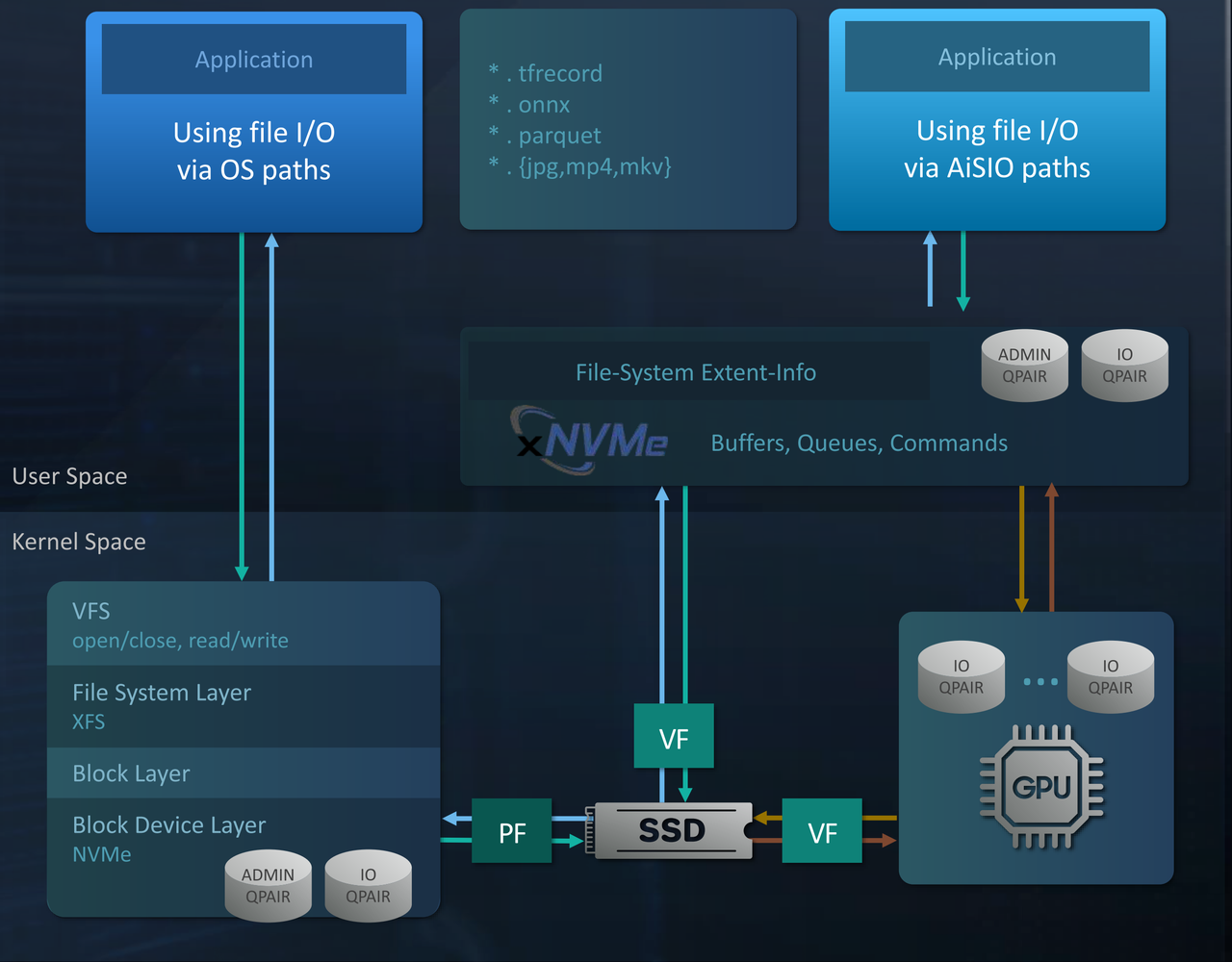


Random read in a PCIe Gen5 setup with a NVIDIA H100 GPU and a PCIe Gen5 SSD in the market. Theoretical peak performance 14.000MB/sec about 13.352MiB/sec. Evaluation clearly illustrates the challenge for state-of-the-art for small I/O sizes. Experiments conducted with infrastructure and hardware provided by Samsung Memory Research Center (SMRC).

The Solution : AiSIO multi-path in Software and Hardware

The AiSIO open-source system software orchestrates multiple I/O paths in SW and HW. Interoperating the conventional OS storage stack with File-Systems on top of inbox drivers, extracting file-extent information to be used by GPUs via their hardware accelerated I/O path with NVMe driver running on the GPU. P2P DMA along the shortest data-path and eliminating memory copies on the fast-path.

- Efficient and scalable block and file I/O
- Hardware enabled multi-path via SR-IOV
- Open-Source and open eco-system
- Standard interfaces: PCIe and NVMe
- Integrates with proprietary, and open, compute devices



Performance Evaluation Part 1/2 : Data-movement using Block I/O

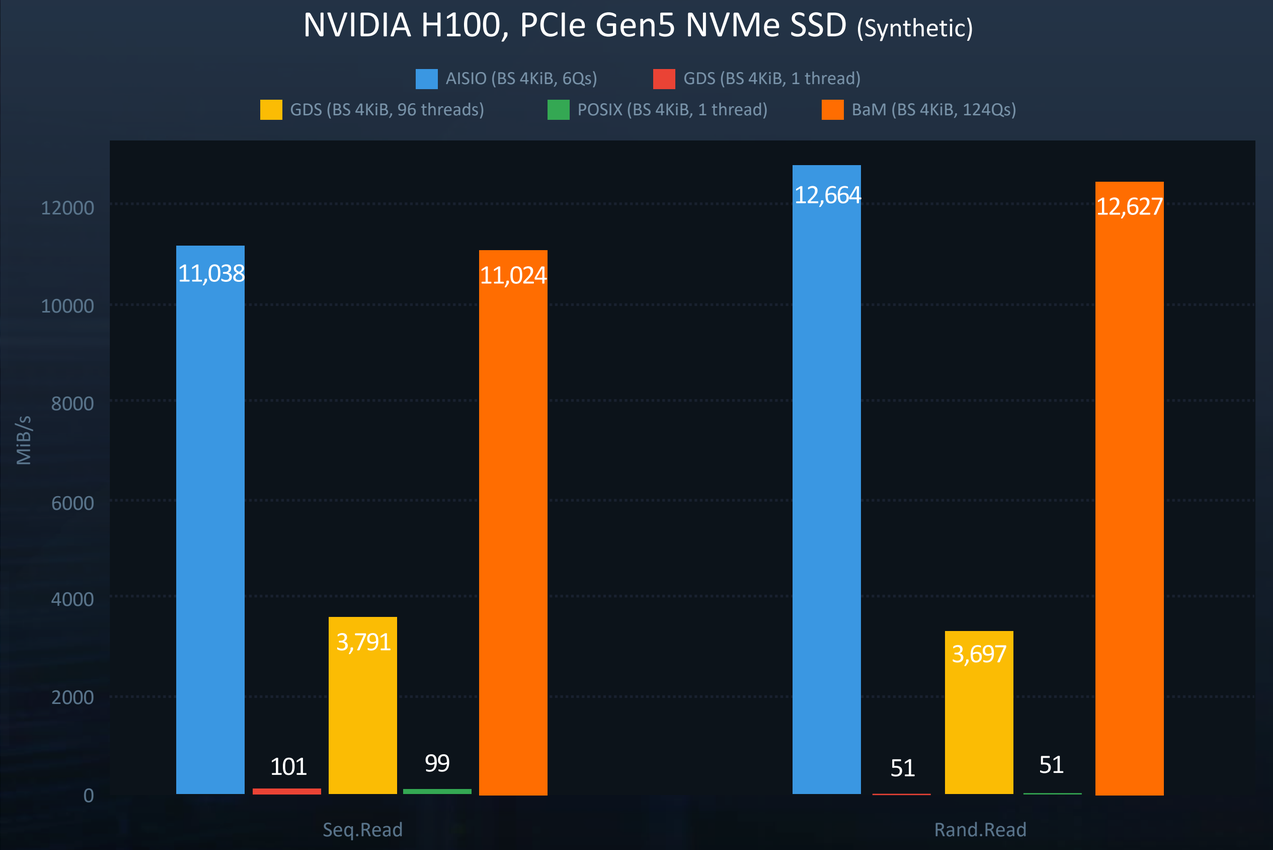
Comparing AiSIO to conventional I/O (POSIX), state-of-the-art in industry (NVIDIA GDS) and cutting-edge research (BaM)

Match the performance of cutting-edge research without sacrificing file-support!

Performance without sacrificing interoperability!

Functions in parallel with BaM; built on the same foundation extended to unlock support for File I/O, without compromise

Synthetic 4K block read as representative for the Small I/O problem in AI workloads



PCIe Gen5 setup with a NVIDIA H100 GPU and a PCIe Gen5 SSD in the market. Theoretical peak bandwidth of 14.000MB/sec about 13.352MiB/sec. POSIX is CPU-initiated, does read() into host memory, then memcpy() to into GPU memory. GDS is CPU-initiated but utilizes P2P for data-transfer. AiSIO is GPU-initiated, running setup and scheduling on the CPU including file-system interaction for extent-info, with P2P data and control on GPU. BaM is GPU-initiated, running setup and scheduling on the CPU, with P2P data and control on GPU, but without support for files. Experiments conducted with infrastructure and hardware provided by Samsung Memory Research Center (SMRC).

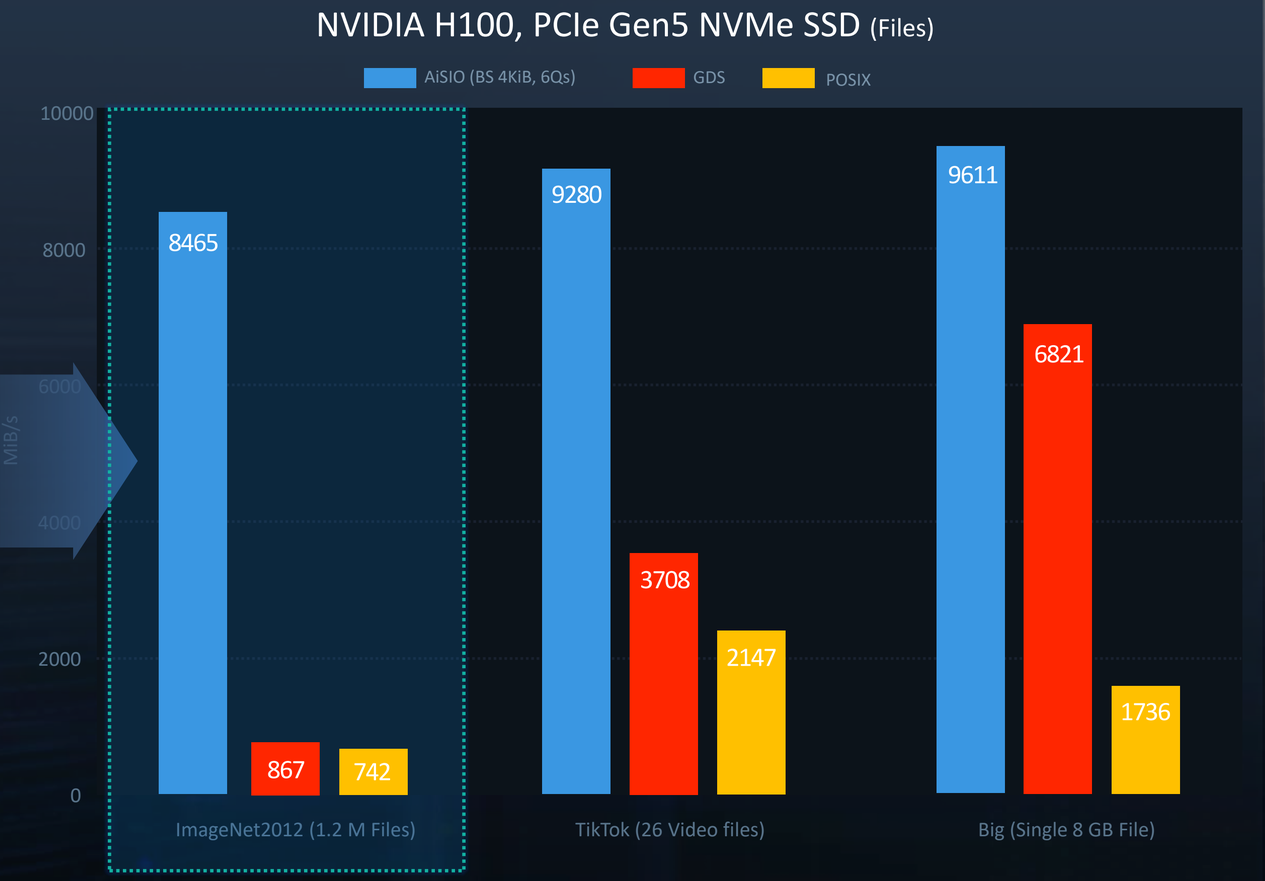
Performance Evaluation Part 2/2 : Data-movement using File I/O

Comparing AiSIO to conventional I/O (POSIX), state-of-the-art in industry (NVIDIA GDS) and cutting edge research (BaM)

9.7x higher bandwidth than state-of-the-art in industry for small I/O

1.4x higher bandwidth than state-of-the-art in industry for large I/O

ImageNet2012 dataset as representative of the Small I/O problem in AI workloads.



PCIe Gen5 setup with a NVIDIA H100 GPU and a PCIe Gen5 SSD in the market. Theoretical peak bandwidth of 14.000MB/sec about 13.352MiB/sec. POSIX is CPU-initiated, does read() into host memory, then memcpy() to into GPU memory. GDS is CPU-initiated but utilizes P2P for data-transfer. AiSIO is GPU-initiated, running setup and scheduling on the CPU, with P2P data and control on GPU. The ImageNet2012 dataset consists of an avg. file size of 100KiB, whereas the TikTok dataset is 200MiB. BaM is not represented here as it does not work with files. Experiments conducted with infrastructure and hardware provided by Samsung Memory Research Center (SMRC).

AiSIO: Accelerator-initiated Storage I/O

Eliminating software bottlenecks through HW/SW co-design enabling efficient and scalable I/O

- Efficient and scalable block and file I/O
- Hardware enabled multi-path via SR-IOV
- Open-Source and open eco-system
- Standard interfaces: PCIe and NVMe
- Integrates with proprietary, and open, compute devices

SEE THE DEMO!