OpenPOWER™

Using NVM Express SSDs and CAPI to Accelerate Data Center Applications in OpenPOWER Systems

Stephen Bates PhD, Technical Director
PMC

Teaser

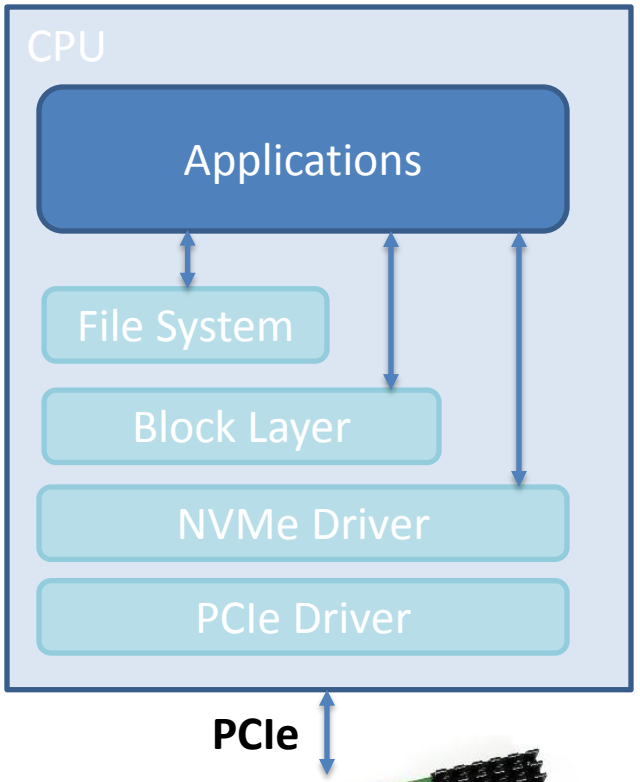
Process your data at 3GB/s with minimal CPU loading. And the code is open-source!

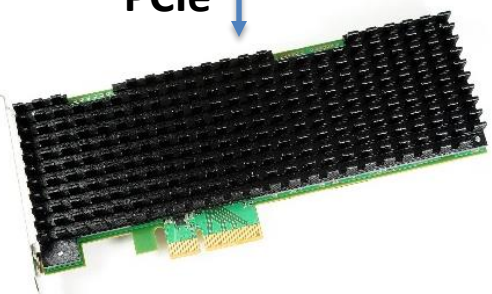
Outline

- What is NVM Express?
- What is CAPI?
- Hardware Setup
- Low-level Performance Data
  - NVM Express SSD performance
  - P8<-> AFU Performance
- A Data-Center Application: String Search and Substitution
- Summary

What is NVM Express?

- NVM Express runs over PCIe
- High Bandwidth and low latency
- Support for multi-core and virtualization
- In-box driver in most OSes



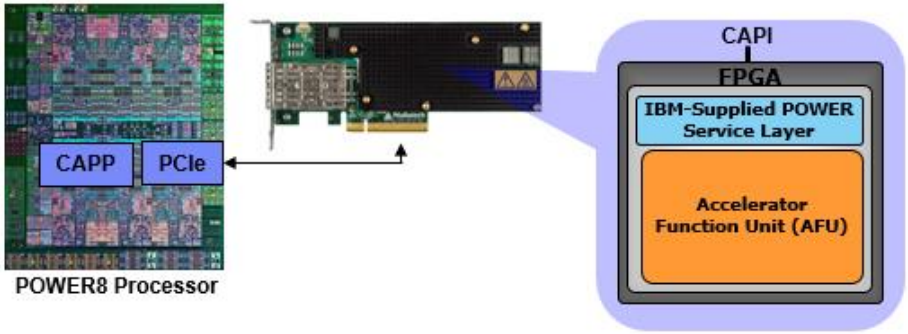


Samsung SM1715 NVMe SSD uses PMC Flashtec Controller

CAPI Technology

What is CAPI?

- CAPI connects the memory subsystem of a Power8 to IO devices via HW assisted PCIe
- Simplifies the programming model and driver for P8<->AFU communication

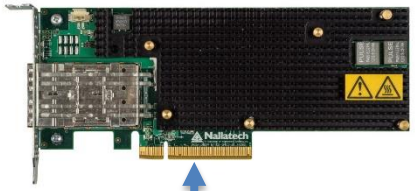


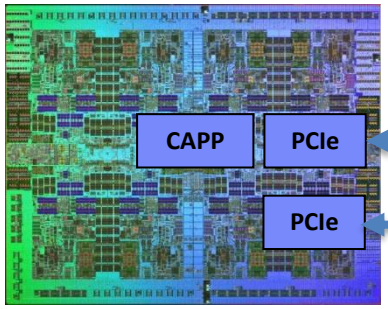
The PSL and AFU can be implemented inside an FPGA (e.g. the Altera Stratix in the Nallatech CAPI card) or inside an ASIC (e.g. the Mellanox ConnectX-4).

The AFU can perform any data-manipulation task and either return results or manipulated data to P8 memory.

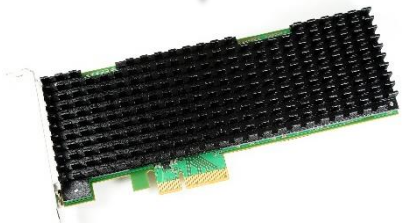
Hardware Setup

- IBM Power8 Server, S822L
- Ubuntu, kernel 3.18.0-14-generic
- Nallatech 385 CAPI card
- Samsung SM1715 1.6TB NVM Express SSD

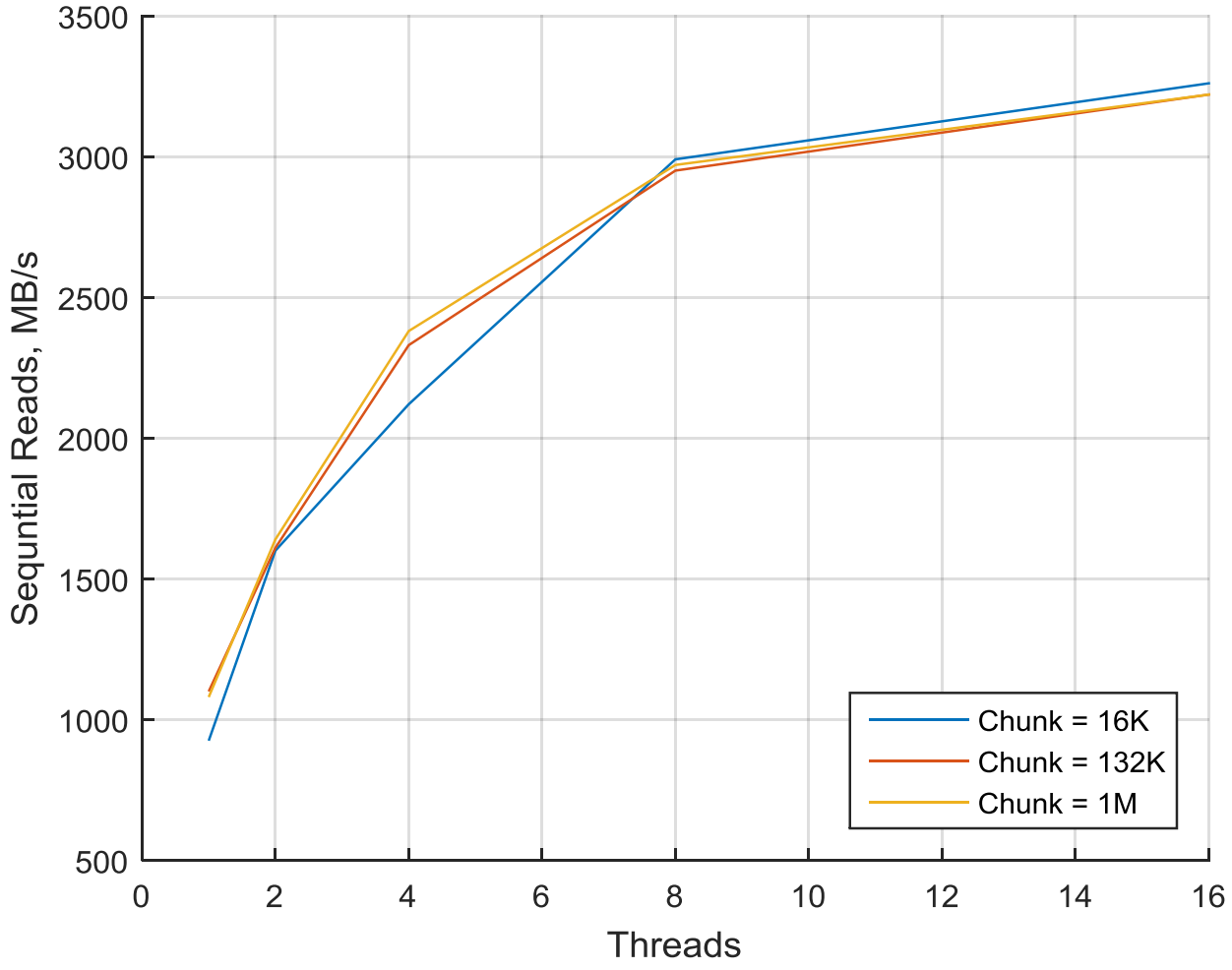




Power8 Processor



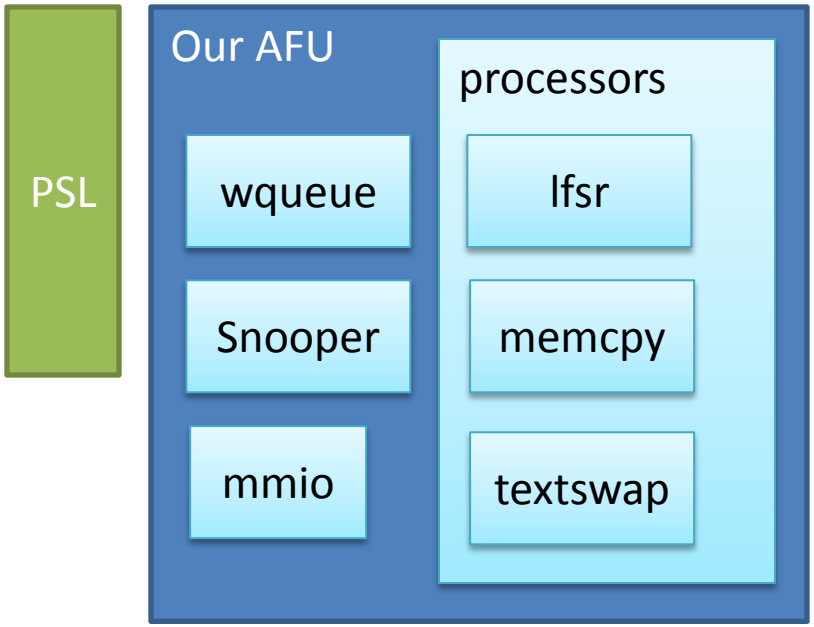
Performance – NVMe SSD



fio, ext4 file-system, in-box NVMe driver

Accelerator Functional Unit

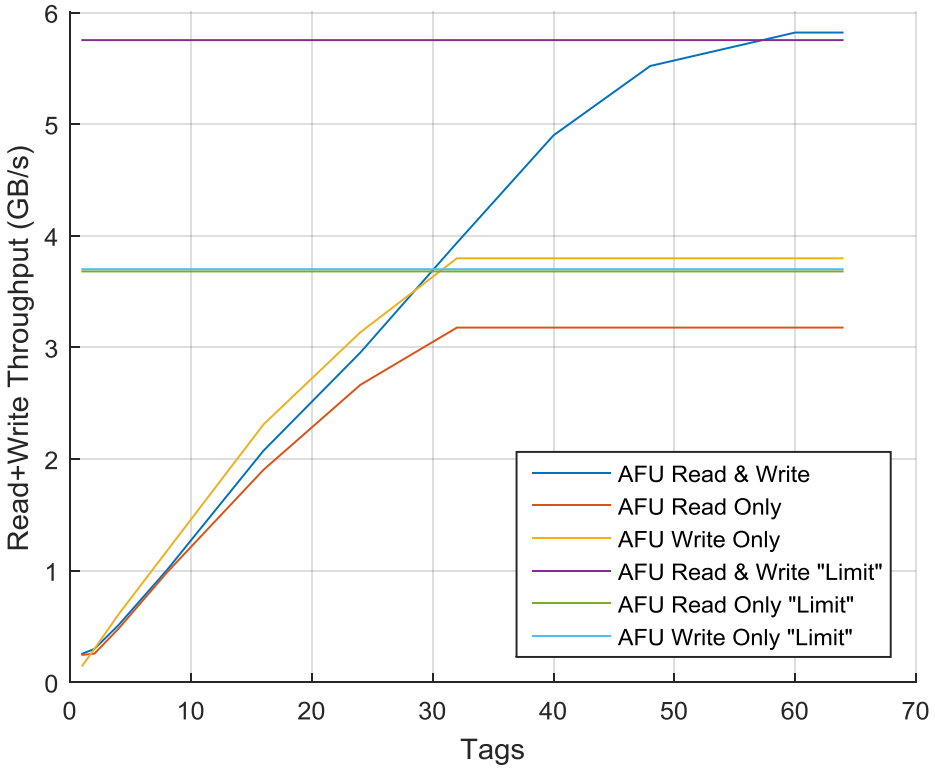
- We wrote a AFU to do low-level performance testing and a simple demo
- AFU monitors queue in memory, processes jobs as they are placed on queue
- A snooper allows for debugging and performance analysis
- Easy to drop in new processing blocks



Our AFU consumes about 30% of the logic resources and 11% of the memory resources on a Stratix V (5SGXMA7H2F35C2).

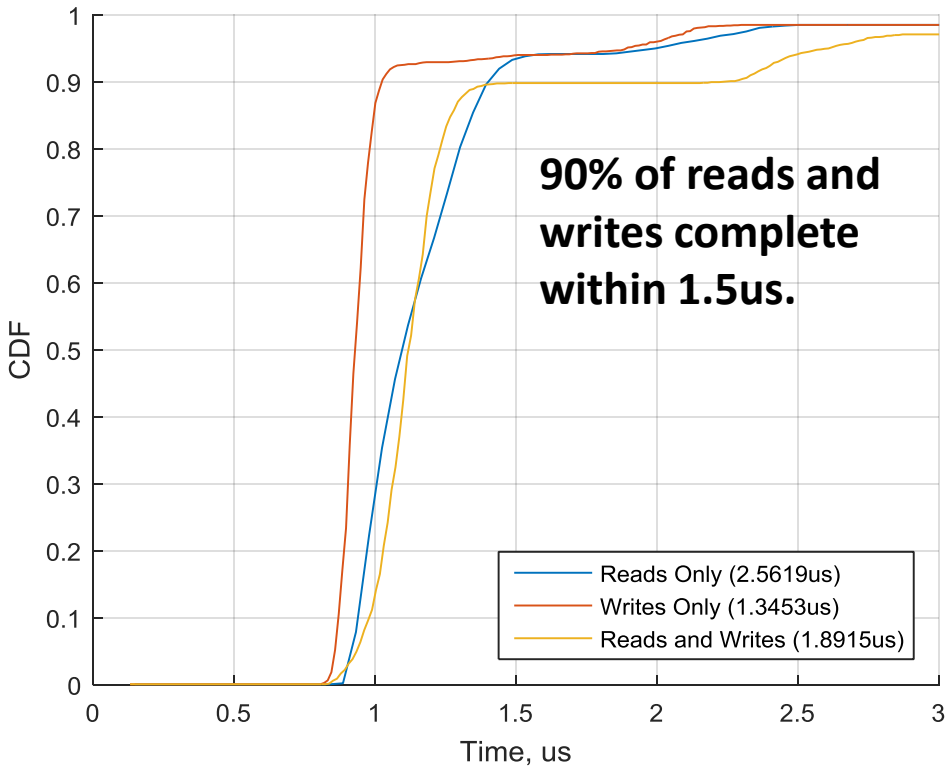
Performance – P8<->AFU

- Moving data between P8 memory and the AFU involves AFU initiated reads and writes
- CAPI allows out of order completions and the AFU must handle this
- A tag and credit based system is used for flow control



Performance – P8<->AFU

- Since the data can reside in a cache, DRAM or even on another CPU the command response time can vary
- Here we plot the PDF for reads, writes and mixed workloads



Text Search Application

- We can combine the NVMe SSD and the AFU to perform search on large data-sets
- In our example we augment the AFU to return pointers to string match locations
- This allows both pattern matching and pattern substitution/annotation to be performed
- This work is easily extended to more complex data processes (e.g. encryption, DNA sequencing)

Device	GB/s
HDD	80MB/s
SAS-SSD	237MB/s
NVMe-SSD	**2950MB/s**

See the demo at the Nallatech Booth #1010!

Summary

High Throughput
Low and Consistent Latency
Low CPU Utilization
Easy Programming Model

Try for Yourself!
https://github.com/sbates130272/capi-textswap.git