Choosing the Optimal Micron Enterprise SSD for Machine Learning Platforms

Real-World Training Systems Don't Run in Isolation — High-Speed Ingest Without Compromising Training Time Is the Ideal Model

Overview

There is a litany of enterprise SSDs one can choose for storage in machine learning (ML) platforms. Choosing the optimal SSDs for these platforms may be far more complex than data sheet specifications and interface rates. In this brief we use common ML benchmarks to compare the relative training rates during simultaneous ingest for ML platforms based on the following Micron® enterprise SSDs:

- Performance NVMe™ (9300 family)
- Mainstream NVMe (7300 family)
- Mainstream SATA (5300 family)

We discuss four key topics:

1. Isolated and production platforms are different
2. Simultaneous training and ingest is desirable
3. Data ingest rates vary widely by SSD
4. Ingest I/O size affects new data set ingest rate

Additional techniques to make ML benchmarks more accurately reflect real-world results are available in our technical brief, "Micron's 9300 NVMe™ SSD Brings Performance to Immense Machine Learning Training Datasets."

Fast Facts

Real-world model training excels with parallel execution.

New training data ingest rates vary by SSD type and interface.

Smaller ingest I/O sizes benefit lower-performing SSDs.

Adding Fast Storage Has Little Effect on Isolated Systems

As shown in our technical note, "Micron's 9300 NVMe™ SSD Brings Performance to Immense Machine Learning Training Dataset," when the training set fits into system memory, the training set is read from storage once (only for the first epoch). All subsequent re-reads are directly from memory. This results in very low storage I/O and reduces the advantages of very fast storage.

Figure 1 shows that when running an isolated training benchmark, SATA and both mainstream and performance NVMe SSDs perform very similarly. This is expected as it is a result of the very low storage I/O typically seen when benchmarking isolated systems. When run in isolation (as many ML training benchmark systems are run), SATA SSDs are able to provide sufficient throughput to support an I/O intensive training workload. For example, the Image Classification workload shown below requires a total storage throughput of 1.2 GB/s.

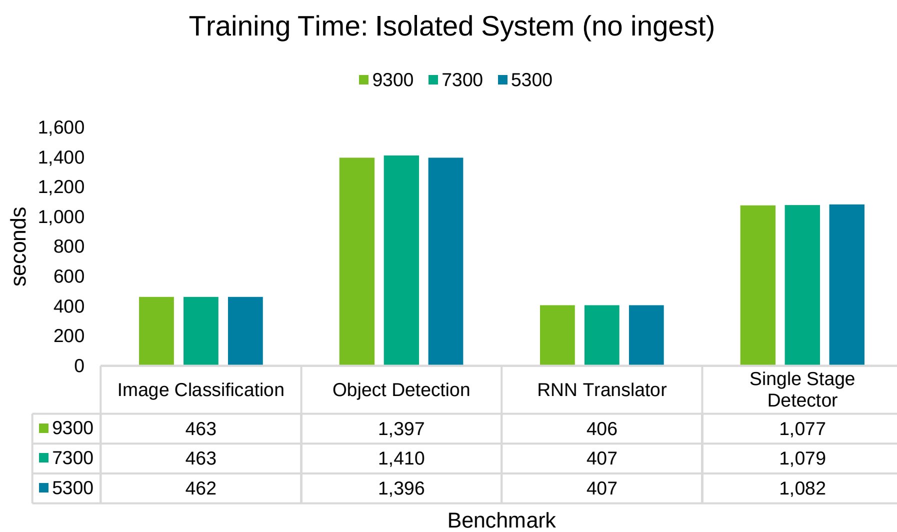

	Image Classification	Object Detection	RNN Translator	Single Stage Detector
9300	463	1,397	406	1,077
7300	463	1,410	407	1,079
5300	462	1,396	407	1,082

Figure 1: Isolated Training

Simultaneous Training and Ingest: Real-World Use

The results in Figure 1 may not reflect how ML training is implemented. Real-world training systems are often not isolated and training isn't the only task running (that is, storage is not idle except when reading in the training data set). Actual deployments tend to be more complex.

- Training is not a one-pass-then-complete process
- Multiple models may need to be trained
- Data sets are very large
- Data sets need to be cached in (copied to) local storage, used, then evicted in preparation for the next data set
- As the data set is evicted, the next data set should be cached into local storage; the faster this can be done, the faster the next training can begin. This operation model is more time efficient.

Figures 2a and 2b illustrate an important difference between isolated and-real world training operating modes.

In Figure 2a (isolated training), the ingest and training processes are serial: Model A training data is ingested, then ingest stops as Model A is trained. Once Model A training completes, Model A training data is discarded and Model B training data is ingested. Once complete, Model B is trained. This serial process repeats for all models.

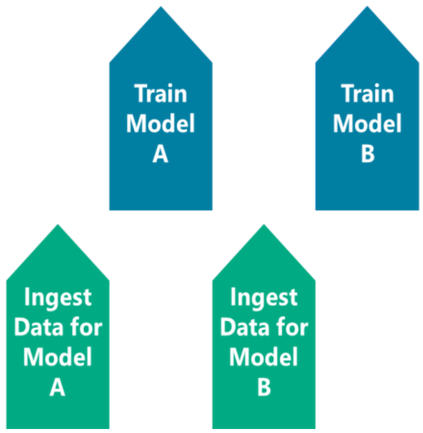

Figure 2a: Isolated Training: Serial

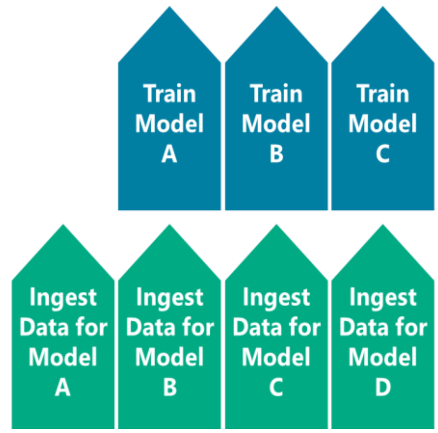

Figure 2b: Real World Training: Parallel

Figure 2b shows actual training processes that are more parallel. In real-world parallel training, we ingest data for training Model A. As Model A is trained, we simultaneously ingest the training data set for Model B. Training and ingest run at the same time — a parallel process that occurs for all models.

Supporting simultaneous training and ingest requires storage devices that can sustain a significant ingest rate while the training occurs, helping optimize GPU resource utilization.

Data Ingest Rates Vary by SSD

To measure data ingest rates during training for each SSD type, we used flexible I/O (FIO) to generate large block, sequential write I/O (128KB transfer size at 32 threads) to the platform's storage while the training workloads ran. We compared the resulting data ingest rate for each SSD type. Figure 3 shows that NVMe SSDs support the parallel data ingest + model training process described in Figure 2b.

	Image Classification	Object Detection	RNN Translator	Single Stage Detector
9300	11,003	10,935	11,248	11,035
7300	6,506	6,687	6,699	5,590
5300	594	879	958	338

Figure 3: NVMe, SATA Data Ingest While Training

Ingest I/O Size Affects Some Benchmark Results

When the ingest I/O size can be controlled, smaller ingest I/O sizes show faster benchmark results (for some benchmarks). We tested ingest I/O sizes shown in Table 1.

Indicator	Ingest I/O Size
	0K (no ingest)
	4K
	128K
	1M

Table 1: Ingest I/O

Figure 4 shows how object detection results are affected by each of the ingest I/O sizes in Table 1. Similar height shaded vertical bars indicate that the ingest I/O size has negligible effect on the object detection benchmark execution time (including results with no ingest, shown at far left).

Figure 4 also shows that the two NVMe SSDs (9300 and 7300) are affected more by a 4K ingest I/O size than the 5300 SATA SSD. Most other ingest I/O sizes do not appreciably affect object detection training time.

A 4K I/O size for new data set ingest has the most effect on object detection benchmark completion time.

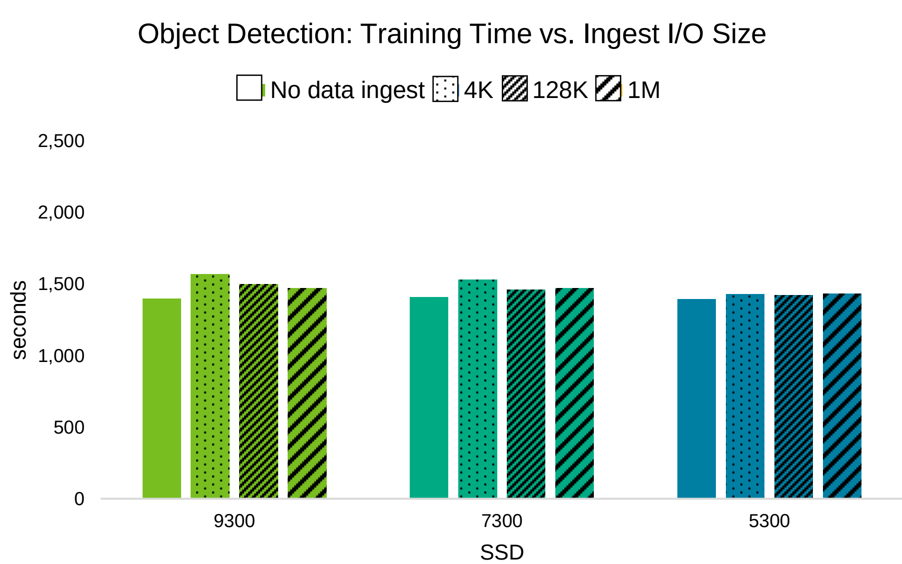

Figure 4: Object Detection Benchmark Results

Single-stage detector benchmark results are different. Here we see that only the performance NVMe SSD (9300) shows little-to-no effect with any ingest I/O size, while the mainstream NVMe SSD (7300) shows a moderate training time increase only with the 1M data ingest I/O size (4K and 128K ingest I/O size training rates are similar to the no-ingest training rate for this SSD).

The SATA SSD (5300) shows significantly longer benchmark times with 1M ingest I/O size. While the 128K and 4K ingest I/O sizes increase the SATA benchmark times, a 1MB data ingest I/O has the greatest effect.

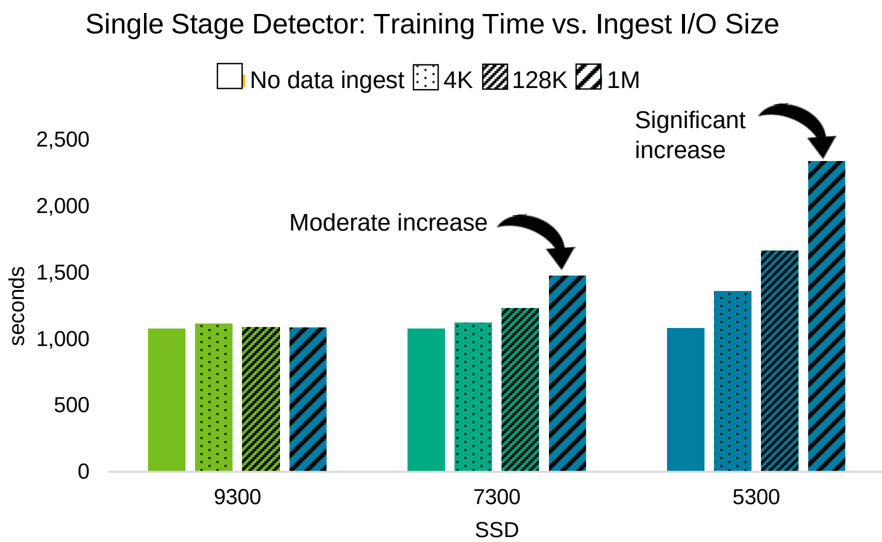

Figure 5: Single-State Detector Benchmark Results

Conclusion

ML training benchmarks, like MLPerf, help ML teams and platform designers glean an understanding of how their configurations will perform. But those benchmarks are typically run on isolated training platforms. In real-use deployments, ML platforms are not typically isolated – they perform additional operations. Ideally, one of those operations is ingesting the next training set as rapidly as possible while the current training executes.

Ingesting data while training can affect benchmark completion times for mainstream NVMe and SATA SSD-based platforms. Tuning the new data set ingest I/O size can have a noticeable effect on SATA SSDs, with a smaller effect on mainstream NVMe SSDs. Performance-focused NVMe SSDs show very little effect.

Learn More

For more details on enterprise SSDs:

- Micron 9300 Performance NVMe: www.micron.com/9300
- Micron 7300 Mainstream NVMe SSDs: www.micron.com/7300
- Micron 5300 SATA SSDs: www.micron.com/5300

For Micron's latest technology insights, visit www.micron.com/insight. Follow on Twitter (@MicronStorage) and connect with us on LinkedIn.

How We Tested

The MLPerf benchmark (https://mlperf.org/) is undergoing rapid development. As of this document's publication, MLPerf Training v0.6 (posted on 07/10/19) was the most recent version available (results from the prior v0.5 benchmark revision were published at the end of 2018). Because of this fluidity, it is imperative to demonstrate that MLPerf results are reproduceable and show strong correlation with results published by others.

In December of 2018, the first benchmark results for MLPerf were submitted by Intel, Google and Nvidia. The results measured the performance of different ML algorithms on the submitters' various hardware, using time to train to an accuracy threshold with their metrics.

Micron has been using similar benchmarks in our Austin, Texas performance engineering lab to help us understand how ML training stresses storage resources. To ensure our results were close to existing published results (hence, validating our test process), we compared our MLPerf results to the commercial results noted above.

Figure 6 shows that Micron's test results (in blue) strongly correlate with results submitted by these commercial entities (in gray). Correlation values = absolute value (Micron results/Commercial results); a correlation value of 1 indicates complete correlation.

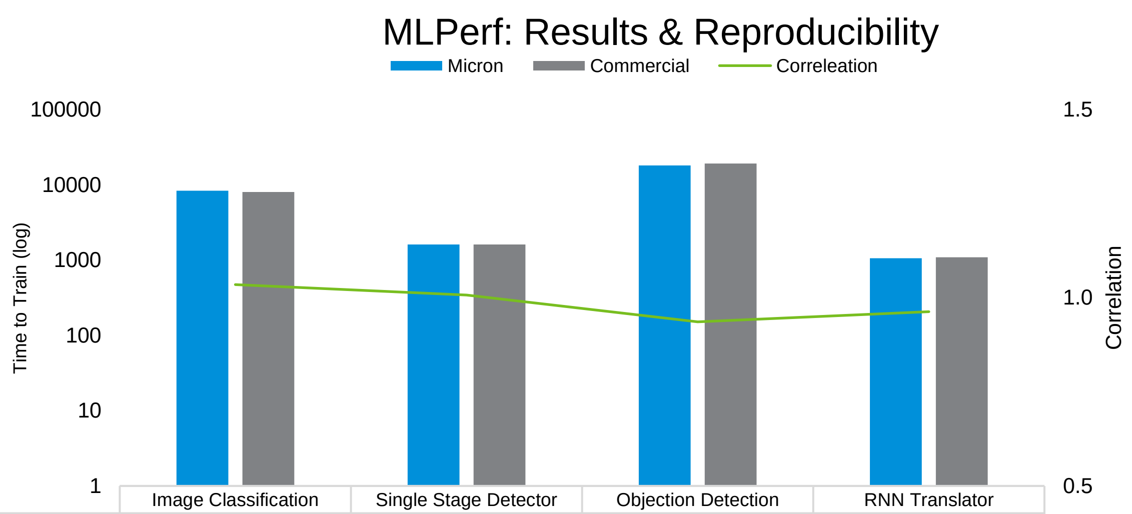

	Image Classification	Single Stage Detector	Objection Detection	RNN Translator
Micron	8360	1626	18138	1058
Commercial	8076	1614	19374	1098
Correleation	1.04	1.01	0.94	0.96

Figure 6: Verifying Reproduceable Results

The correlation values in Figure 6 range from a low of 0.94 to a high of 1.04. This narrow distribution indicates very good correlation results.