

Performance vs. Latency

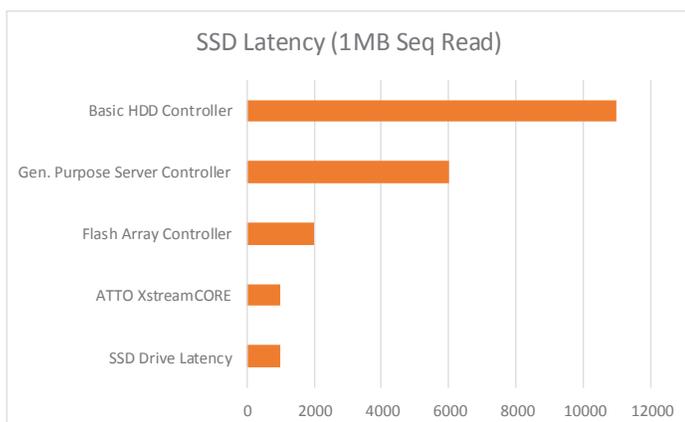
Today's demanding data center environments have an added extra layer of complexity. It used be that all admins needed to worry about was available storage capacity but flash storage has made maximizing performance to these devices another priority. With the rapid integration of flash storage system architects have to make sure that the systems they are building are capable of taking advantage of the performance capabilities of SSD flash storage. One of the key measures of how well flash will integrate is storage latency, the amount of time it takes to read or write data from a storage device and deliver it to an application. Latency directly effects throughput and IOPS performance, high latency will lead to applications waiting for data while lower latency will efficiently feed applications so all server resources can be maximized.

Benefits of SSD based flash storage

- Solid state design leads to lower failures
- No moving parts allows faster transfer of data, more completed transactions
- Faster streaming of data without buffering
- Allows admins to utilize more CPU cycles compared to using hard drives

Latency Comparison

XstreamCORE adds an insignificant amount of latency to native flash transactions thanks to xCORE data acceleration technology.



This chart illustrates the latency associated with each of the components associated with an enterprise storage solution. Using the SSD drive as the baseline, the ATTO XstreamCORE Controller adds <4ms of latency

ATTO xCORE Data Acceleration

With the advent of vendors using commodity servers to build flash architectures there has been a lack of focus on one of the items that make flash storage so beneficial, low latency. ATTO leveraged highly programmable FPGA technology to develop a platform that minimizes latency added to each read and write. One of the key ways we do this is by separating the data and control planes to keep latency at a consistent, deterministic sub-four microsecond rate. xCORE manages the data plane and assures that all reads and writes are processed as fast as possible. All nondata transactions are handled by ATTO exclusive eCORE Control Engine so priority is maintained for all data transfer traffic.

Benefits of delivering the lowest latency possible with ATTO

- Data is more quickly served to host / application servers
- Applications don't wait as long for data to complete their tasks
- More efficient use of CPU / Server resources
- No penalty to network storage
- Leads to faster access to shared storage than local storage

Where can I find xCORE?

xCORE technology is currently available exclusively on ATTO XstreamCORE solid-state storage controllers which connect stranded SAS storage to shared storage networks. XstreamCORE allows storage architects to build solutions that meet today's demanding performance requirements while maximizing the amount of performance available in flash storage.



What is latency?

Computer and operating system latency is the combined delay between an input or command and the desired output. In a computer system, latency is often used to mean any delay or waiting that increases real or perceived response time beyond what is desired.

What does latency mean in the real world?

One would think that speed is the answer to latency. But that is not necessarily so until you can define what speed is and what other factors affect it. For example, any number of factors could have an effect on this speed and storage latency. Network type such as Fibre Channel vs iSCSI, or the disk speed SATA vs SSD, or the controller, or even virtualization on the data pool that drive the speed of the storage. These factors are not mutually exclusive and must be anticipated in order to make decisions about what is your fast storage and what is your slow storage.

Perception vs. reality

We all make the mistake in thinking that a bigger number is better, that more is better. Adding more hardware or software can actually add to the latency issue because each request or command needs to be processed and responded to, thus adding latency. The difference of a few microseconds may not be noticeable by a human but can mean a lot to a data center doing millions of transactions per hour.

100ms of latency can cost you dollars



Amazon sales in 2014 were just under \$89B. Assuming that 1 % of sales cost for every 100ms of latency, 100ms of latency for all of 2014 would've cost the company a staggering \$889M throughout the year. Time is money!



In financial services, the cost of latency is enough to warrant a \$1.5B investment to reduce latency between London and Tokyo by 60ms.



Current estimates from Akamai show that a 1 second delay in page response can result in a 7% reduction in conversions. For an ecommerce site making \$100,000 per day, that adds up to \$2.5 million in lost sales every year.



If a broker's electronic trading platform is 5ms behind the competitions it could cost them at least 1 % of their flow, that's \$4M in revenues per ms (TABB Group)



A 1 millisecond advantage can be worth \$100,000,000 to a major brokerage firm. Considering there are trillions of transactions per day and that prices can change 8+ times per second



Knowing that 1 millisecond can be worth millions of dollars the cost of the first trans-artic ocean fiber optic cable between Tokyo and London resulted in a latency decrease of 38% at a cost of only \$1.5B.



The XstreamCORE Storage Controller adds <4us of latency. System Architects can design solutions using XstreamCORE that will lower OpEx and CapEx while meeting the demands of organizations from SMB to enterprise.