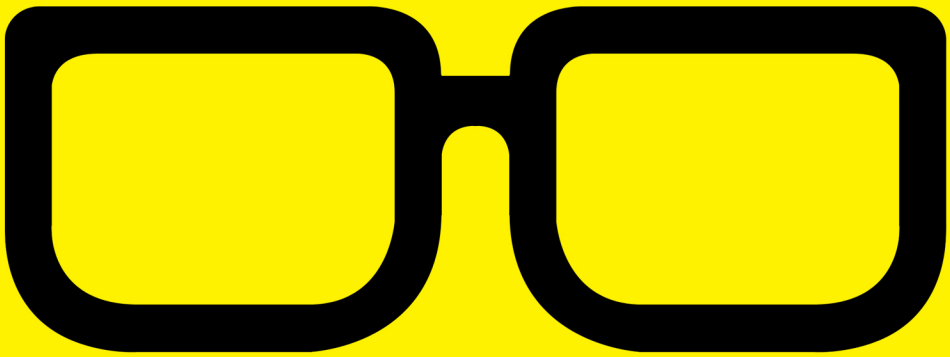
SPONSORED BY

intel Software

GEEK GUIDE



Memory: Past, Present and Future—and the Tools to Optimize It

LINUX™ JOURNAL

Table of Contents

About the Sponsor 4
Introduction 5
A Brief History of Computer Memory 7
Moving Closer to the CPU 8
The Solid-State Drive 10
Non-Volatile Memory Express (NVMe) 13
The Future in Memory Technologies 15
The Need for Speed 16
Time to “Go Parallel” 17
Analyze, Identify and Optimize 18
Conclusion 20

PETROS KOUTOUPIS is currently a senior software developer at **IBM** for its **C**loud **O**bject **S**torage division (formerly **C**leversafe). **H**e is also the creator and maintainer of the **R**apid**D**isk **P**roject (http://www.rapiddisk.org). **P**etros has worked in the data storage industry for more than a decade and has helped to pioneer the many technologies unleashed in the wild today.

GEEK GUIDES:

Mission-critical information for the most technical people on the planet.

Copyright Statement

© 2017 *Linux Journal*. All rights reserved.

This site/publication contains materials that have been created, developed or commissioned by, and published with the permission of, *Linux Journal* (the "Materials"), and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of *Linux Journal* or its Web site sponsors. In no event shall *Linux Journal* or its sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

No part of the Materials (including but not limited to the text, images, audio and/or video) may be copied, reproduced, republished, uploaded, posted, transmitted or distributed in any way, in whole or in part, except as permitted under Sections 107 & 108 of the 1976 United States Copyright Act, without the express written consent of the publisher. One copy may be downloaded for your personal, noncommercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Linux Journal and the *Linux Journal* logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners. If you have any questions about these terms, or if you would like information about licensing materials from *Linux Journal*, please contact us via e-mail at info@linuxjournal.com.

About the Sponsor

Intel® Software and Services Group

The Intel® Software and Services Group (Intel® SSG) is responsible for software and software development at Intel worldwide, with the goal of enhancing computing and connectivity for Intel® architecture across the software ecosystem and providing end-to-end value from the device edge to the data center. With resources spanning more than 20 countries, business relationships with over 7,000 independent software vendors, and a robust developer program that reaches nearly 20 million software, entrepreneurial, enthusiast and hobbyist developers, Intel Software fuels worldwide innovation to help developers take advantage of the latest computing and communications technologies.

Memory: Past, Present and Future—and the Tools to Optimize It

PETROS KOUTOUPIS

Introduction

In recent years, solid-state drives (SSDs) have taken the computer industry by storm. The technology has impressive capabilities. It promises low-latency access to sometimes critical data while increasing overall performance—at least when compared to what is now

becoming the legacy hard disk drive (HDD). With each passing year, SSD market shares continue to climb, and in many sectors, SSDs have been replacing HDDs (for example, personal and mobile computing).

The HDD was first unleashed into the computing world by IBM in 1956. By the 1960s, the HDD became the dominant secondary storage device for general-purpose computers. The primary characteristics that define the HDD are its capacity and performance. The first IBM-manufactured hard drive, the 350 RAMAC, was as large as two medium-sized refrigerators with a total capacity of 3.75MB on a stack of 50 disks. Modern HDD technology has produced disk drives with volumes as high as 12TB, specifically with the more recent Shingled Magnetic Recording (SMR) technology coupled with helium. The sealed helium gas increases potential speed of the drive while creating less drag and turbulence. Being less dense than air, it also allows more platters to be stacked in the same space used by 2.5" and 3.5" conventional disk drives.

The performance of a disk drive is typically specified by the time required to move the drive's heads to a specific track or cylinder and the time it takes for the requested sector to move under the head—that is, the latency. Performance is also measured by the speed that the data is transmitted. Now, although the performance of HDDs has been increasing with newer protocols (Parallel ATA or PATA, Serial ATA or SATA, and even Serial Attached SCSI or SAS) and technologies, it is still a bottleneck to the CPU and, in turn, the overall computer system.

Since its conception, engineers have been devising newer

and creative methods to help accelerate the performance of HDDs, completely replacing them in some cases.

A Brief History of Computer Memory

Before non-volatile memory (NVM) came into the picture, the computing world was first introduced to volatile memory in the form of random access memory (RAM). What RAM introduced was the ability to write/read data to/from any location of the storage medium in the same amount of time. The often random physical locations of a particular set of data did not affect the speed at which the operation completed.

The most notable of RAM technologies is the dynamic random access memory (DRAM). It also came out of the labs at IBM, in 1966, a decade after the HDD. Being that much closer to the CPU and also not having to deal with mechanical components (that is, the HDD), DRAM performed at stellar speeds. Even today, many data storage technologies strive to perform at the speeds of DRAM. Although there was a drawback; as emphasized above, the technology was volatile, and as soon as the capacitor-driven integrated circuits (ICs) were deprived of power, the data disappeared along with it.

Another set of drawbacks to the DRAM technology is its very low capacities and the price per gigabyte. Even by today's standards, DRAM is just too expensive when compared to the slower HDDs and SSDs (discussed below).

Shortly after the debut of DRAM came the erasable programmable read-only memory (EPROM). Invented by Intel, it hit the scene in about 1971. Unlike its volatile counterparts,

Unlike its volatile counterparts, the EPROM offered an extremely sought-out-for industry game-changer: memory that retains its data as soon as system power was shut off.

the EPROM offered an extremely sought-out-for industry game-changer: memory that retains its data as soon as system power was shut off. The EPROM used transistors instead of capacitors in its ICs. It is these transistors that were capable of maintaining state after the electricity was cut.

As its name implies, the EPROM was in its own class of read-only memory (ROM). Data typically was pre-programmed into these chips using special devices or tools, and when in production, its only use was to be read from, at high speeds. As a result of this design, the EPROM immediately became popular in both embedded and BIOS applications, the latter of which stored vendor-specific details and configurations.

Moving Closer to the CPU

As time progressed, it became painfully obvious that the closer you move data (storage) to the CPU, the faster you are able to access (and manipulate) it. The closest memory to the CPU are the processor's registers. The amount of available registers to a processor varies by architecture. The purpose of a register is to hold a small amount of data intended for fast storage. Without a doubt, these registers are the fastest way to access small sizes of data.

Next in line, and following the CPU's registers, is the CPU cache. This is a hardware cache built in to the processor module and utilized by the CPU to reduce the cost and time it takes to access data from the main memory (that is, DRAM). It is designed around a static random access memory (SRAM) technology that is also a type of volatile memory. Like your typical cache, the purpose of this CPU cache is to store copies of data from the most frequently used main memory locations. On modern CPU architectures, there exists multiple and different independent caches (and some of those caches are even split). They are organized in a hierarchy of cache levels: Level 1 (L1), Level 2 (L2), Level 3

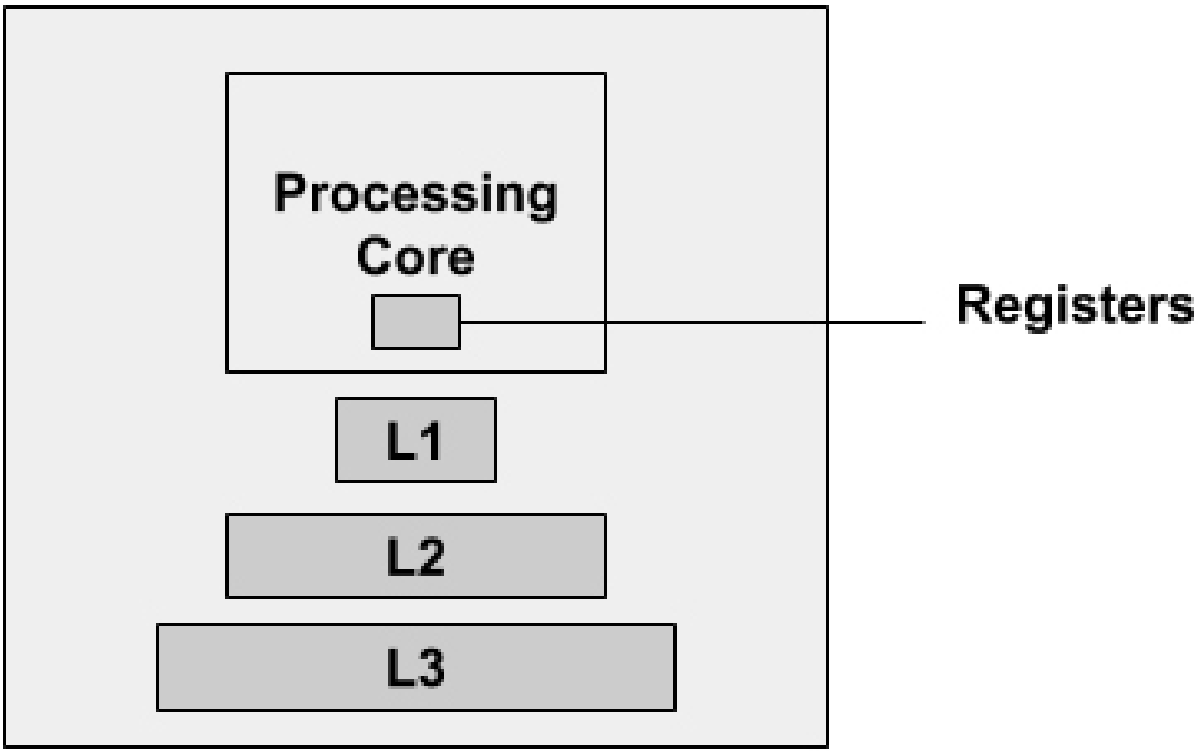


FIGURE 1. A General Outline of the CPU and Its Memory Locations/Caches

(L3) and so on. The larger the processor, the more the cache levels, and the higher the level, the more memory it can store (that is, from KB to MB). On the downside, the higher the level, the farther away its location is from the main CPU. Although mostly unnoticeable to modern applications, it does introduce latency.

The first documented use of a data cache built in to the processor dates back to 1969 and the IBM System/360 Model 85 mainframe computing system. It wasn't until the 1980s that the more mainstream microprocessors started incorporating their own CPU caches. Part of this delay was driven by cost. Much like it is today (all types of) RAM was very expensive.

The Solid-State Drive The performance of a particular storage technology was constantly gauged and compared to the speeds of CPU memory. So, when the first commercial SSDs hit the market, it did not take very long for both companies and individuals to adopt the technology. Even with a higher price tag (dollar per GB) when compared to HDDs, people were able to justify the expense. Time is money, and if access to the drives saves time, it can potentially increase profits. It is unfortunate though that with the introduction of the first NAND-based SSDs, the drive did not move data storage any closer to the CPU. These more modern SSDs would continue to adopt and use the already established SATA, SAS and Fibre Channel (FC) storage interface protocols. Now, although we did not move any closer to the CPU, we did achieve a new milestone in this technology: reduced seek times across the storage media, resulting in significantly less latencies. This is

because the drives were designed around ICs and contained no movable components. Its overall performance was night and day when compared to the traditional HDD.

The first official SSD to be manufactured without the need of a power source (that is, a battery) to maintain state was introduced in 1995 by the company M-Systems. They were designed to replace HDDs in mission-critical military and aerospace applications. By 1999, the Flash-based technology would be designed and offered in the traditional 3.5" storage drive form factor. It would continue to be developed this way until 2007 when a newly started and

FIGURE 2. A SATA SSD in a 2.5" Drive Form Factor

revolutionary startup company named Fusion-io (now part of Western Digital) decided to change the performance-limiting form factor of traditional storage drives and throw the technology directly onto the PCI Express (PCIe) bus. This approach would remove many unnecessary communication protocols and subsystems. The design also would bring us a bit closer to the CPU and produced a noticeable improvement in performance. This new design would not only change the technology for years to come, but it also even brought the SSD into the data center.

The early SSD offerings were relatively small with regard to storage capacities, but as the years passed and the technology improved, this once limiting factor became significantly better. For instance, today, if you have the money to spend, you can invest in a 16TB SSD.

Fusion-io's products eventually would inspire other memory and storage companies to bring somewhat similar technologies to the dual in-line memory module (DIMM) form factor, which plug directly in to the traditional RAM slot of the supported motherboard. These types of modules register to the CPU as a different class of memory and remain in a somewhat protected mode. Translation: the main system and, in turn, the operating system did not touch these memory devices unless it was done through a specifically designed device driver or application interface. Unfortunately, this type of technology has not picked up too much momentum in recent years.

It is also worth noting here that the transistor-based NAND Flash technology still paled in comparison to DRAM performance. We are talking about microsecond latencies

versus DRAM's nanosecond latencies. Even in a DIMM form factor, the NAND-based modules just don't perform as well as the DRAM modules.

Non-Volatile Memory Express (NVMe) The work of Fusion-io continued to inspire in the form of the Non-Volatile Memory Express (NVMe) drives. With the first industry specifications announced in 2011 (http://www.nvmexpress.org), NVMe quickly rose to the forefront of SSD technologies. Remember, historically SSDs were built on top of SATA, SAS and FC buses. These interfaces worked well for the maturing Flash memory technology, but with all the protocol overhead and bus speed limitations, it did not take long for those drives to experience their own fair share of performance bottlenecks. Today, modern SAS drives operate at 12Gbit/s, while modern SATA drives operate at 6Gbit/s. This is why the technology shifted its focus to PCIe. With the bus closer to the CPU and PCIe capable of performing at increasingly stellar speeds, SSDs seemed to fit right in. Using PCIe 3.0, modern drives can achieve speeds as high as 40Gbit/s. Leveraging the benefits of PCIe, it was then that the NVMe was conceived.

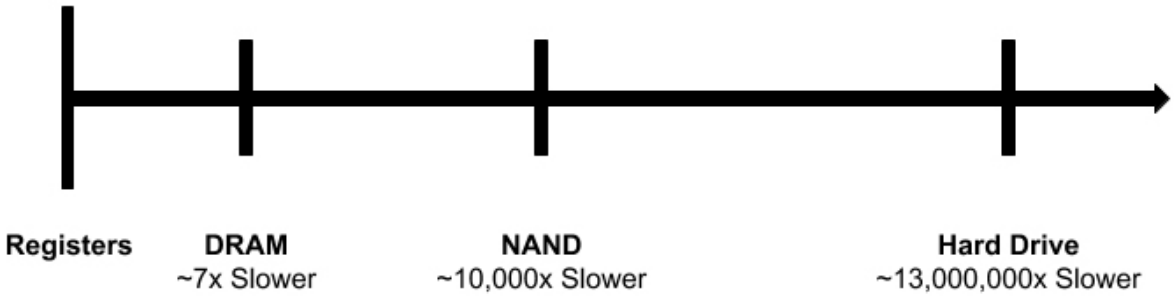


FIGURE 3. An Overview of Data Storage Performance Relative to the Distance from the CPU

What really makes NVMe shine over the operating system's legacy storage stacks is its simpler and faster queueing mechanism. These are called the submission queues (SQs) and completion queues (CQs). Each queue is a circular buffer of a fixed size that the operating system uses to submit one or more commands to the NVMe controller. One or more of these queues also can be pinned to specific cores, which allows for more uninterrupted operations.

Almost immediately, the PCIe SSDs were marketed for enterprise-class computing with a much higher price tag. Although still more expensive than its SAS or SATA cousins,

FIGURE 4. A PCIe NVMe SSD

the dollar per GB of Flash memory continues to drop, enough to convince more companies (enterprise and cloud service providers) to adopt the technology.

The Future in Memory Technologies

As I write this, companies like Intel, Micron, SanDisk and others are constantly researching and experimenting with newer and better ways to store information. Although its early development started in 2012, in 2015, both Intel and Micron unveiled a few teaser details to an upcoming line of persistent memory products through a joint effort. Initially, this new memory technology was branded as 3D XPoint (pronounced cross-point), but it has since been relabeled as the Optane technology by Intel and the QuantX technology by Micron. Note that I continue to refer to this technology as Optane here. Now, fast-forward to the present. We still do not know much about Optane. A lot of its details are shrouded in mystery.

This is what we do know. Optane is a breakthrough NVM technology that combines both memory and storage. It doesn't perform as well as DRAM (4–8x slower), but it does perform significantly better than NAND-based SSDs (10x lower latencies to NAND over NVMe). Although official price listings have yet to be announced, it is supposedly cheaper than DRAM while being somewhat cost-competitive to traditional SSDs. Its capacities scale better than DRAM with 10x the densities. The overall endurance (or lifetime) of the memory technology is noticeably better than that of NAND, with 1000x improvement. These numbers are provided only by

both Intel and Micron. Availability of samples (to select customers) has been very limited.

This better endurance alone is an extremely important evolution. The biggest problem with today's SSDs is that its memory cells can be written and rewritten to only a finite number of times. There is definitely an obvious trend in that the higher the SSD volume's capacity, the worse the endurance. As a result, modern SSD vendors resorted to various tricks (write coalescing and over-provisioning) and unique allocation algorithms (wear-leveling) to reduce cell wear on the drives. Now, how does Optane improve on this? It is still unclear.

Also, how much closer will Optane bring our data to the CPU? And, will it plug in to existing systems? We are unable to answer these questions at the moment. At the time of this writing, it is still unclear whether we even will see any of this technology in the year 2017.

Either way, it should not be much longer until we find out these details, and when we do, chances are that we will need to rethink our current server designs and potentially rewrite our applications accessing this newer form of data storage. This technology alone will definitely redefine the modern-day data center.

The Need for Speed

Focusing on what is available today for both multicore processors and memory technologies, the goal here is to achieve multicore speeds. Time has shown that as demands increase, data set sizes also increase. With larger data set sizes, the longer it will take to retrieve and process that

Now, how can we leverage today's architectures to process those large data sets much quicker? How do we optimize our applications for memory storage technology of both today and the future?

same requested data. Now, how can we leverage today's architectures to process those large data sets much quicker? How do we optimize our applications for memory storage technology of both today and the future?

Time to "Go Parallel" Catering to the high-performance computing (HPC) sector, Intel has launched a series of applications under the umbrella of the Parallel Studio XE suite. The primary purpose of these applications is make your code run faster. Whether your software performs big data analytics, medical imaging, time-critical (financial) analysis, simulations or any other tasks, this suite will aid in identifying the problems with your code and guide you to optimize it for faster results. The suite works with the development tools you probably already are using: GNU, XCode and Visual Studio on Linux, Mac OS X and Windows. The Intel Parallel Studio XE boosts application performance by taking advantage of the always increasing processor(s) core count and vector register width in modern Intel Xeon and Intel Xeon Phi processors and coprocessors. Note that this suite is supported only for Intel architectures.

The suite is designed around the C, C++, Fortran and Python programming languages. Multiple editions of the suite exist, and each incremented level unlocks a new set of tools, starting from the basic compilers and performance libraries to performance profilers and analysis tools and much more. You can read more about this collection of tools at http://software.intel.com and http://goparallel.sourceforge.net.

Analyze, Identify and Optimize When your software is un-vectorized and/or un-threaded, it will under-perform. The fundamental idea behind vector programming is that operations apply all at once to an entire set of values—that is, you operate on whole aggregates of data without resorting to explicit loops of individual scalar operations. Threaded (or multithreaded) code is a technique that allows multiple threads to exist within the context of a single application process. While these threads share the process' resources, they are able to execute independently.

Below is a short list of some of the major features included in the Intel Parallel Studio XE suite:

- The Intel Trace Analyzer and Collector is a graphical utility designed to find your message passing interface (MPI) bottlenecks and potential imbalance quickly, improve code correctness and help to achieve high performance for parallel cluster applications. The built-in and unique MPI Correctness Checker detects deadlocks, data corruptions and errors with MPI parameters, data types, buffers and so on. It is thread-safe and allows you to trace multithreaded MPI applications for event-based

tracing, as well as non-MPA threaded applications.

- The Intel Advisor is a vectorization optimization and thread prototyping tool that offers a platform to prioritize, prototype and predict potential performance gains. With today's processors, it has become increasingly crucial to both vectorize and thread software when attempting to achieve full performance on the processor. Code that is vectorized and threaded can sometimes achieve 175x or more faster performance when compared to code that hasn't been optimized. You need relevant and good data when making design decisions. This is where the Advisor comes into the picture by locating vectorization blockages while giving you the necessary tips for effective vectorization.

- The Intel Inspector provides you with a memory and thread debugger. Find and root cause errors. Debug race conditions and deadlocks quickly. What makes this tool extremely ideal is that you can use normal debug or production builds of your code—that is, there is no need for special compilers or builds. Immediately identify and locate memory leaks/corruption, illegal memory access, uninitialized memory reads, data/heap/stack races, deadlocks and more.

- Intel VTune Amplifier is a performance profiler that accurately collects CPU, GPU, FPU, threading, memory data and more. It is designed to introduce minimal or low overhead with the intent not to modify (or add to) the

application's native execution profile. You then are able to analyze your results with the user-friendly graphical interface, and you can sort, filter, visualize and interpret your data from a local or remote machine. Or, you simply can automate your analysis routines and extract the desired data from the command-line interface (CLI).

Conclusion

When Intel initially announced the 3D XPoint technology, it was stated that two form factors would be supported: 1) the standard SSD PCIe form factor, covering everything from notebooks to servers, and 2) the DIMM form factor, intended primarily for Xeon systems. In the case of using the Optane DIMMs alongside Xeon Phi processors, the memory hierarchy becomes a bit more complex. This is where the Intel Parallel Studio XE will truly shine and ensure that you are accessing that same memory efficiently.

Whether it is with the DRAM of today, the Optane of tomorrow or another game-changing memory technology of the future, Intel's Parallel Studio XE is positioned to ensure that you unlock the Intel Xeon processor's full capabilities and achieve maximum application performance.■