Inside Track
Executive Brief

FREEFORM DYNAMICS

The non-volatile memory opportunity

Why NVM matters

in association with

FUJITSU

The convergence of compute and data

In the earliest days of computing, data storage was almost an afterthought: the computation and the "answer" mattered much more. However, as the use of computers spread outside of government and research and into the world of business, it rapidly became clear that the long-term storage of data was essential. Indeed, data can now be one of an organizations' most valuable resources.

Today, not only are compute and data storage almost equally important, but much of the technology development in IT has been to bring compute and data ever closer. In an age where data analytics are used to make real-time or near real-time business and operational decisions, getting data in and out of processors faster can trigger significant business success.

Non-volatile memory, or NVM, implemented in the form of NVDIMMs – memory modules that can be used in a computer – is the latest in a long series of technologies designed to bring data ever closer to the processor. This paper looks at why NVM matters, and the opportunities available for your organization in adopting NVDIMMs.

Introducing NVM

You can think of NVM as a memory/storage hybrid. It is useful because it gives us memory that sits very close to the CPU, but which is able to retain its data even if power is lost or switched off.

To understand why this matters, let's look at what volatility is in a computing context, and perhaps more importantly, why we need to find ways around it.

Processor, memory and storage

Most computers have three main elements: a processor (CPU) that runs programs and manipulates data, memory for the CPU to work with, and storage where data is held long-term, even when the system is switched off.

In this model, the memory sits as close to the CPU as possible. These days, it is usually DRAM (dynamic random-access memory) implemented as plug-in DIMMs (dual inline memory modules). However, DRAM is very expensive if you need enough to handle extremely large data sets, such as those typically required for 'in-memory computing'. More importantly, DRAM is volatile, so it only works when the system has power. If the power fails or is switched off, data held in volatile memory is lost unless it can first be saved to persistent, or non-volatile, storage.

Volatile vs non-volatile

That's why non-volatile storage devices exist, such as tape, hard disk and SSD or Flash. These devices retain data even when they are powered off, but they are many orders of magnitude slower than DRAM. That means delays whenever the system needs to get data back into memory, or write it from memory to permanent storage. As the figure below shows, NVM can bridge the gap between expensive volatile memory and storage.

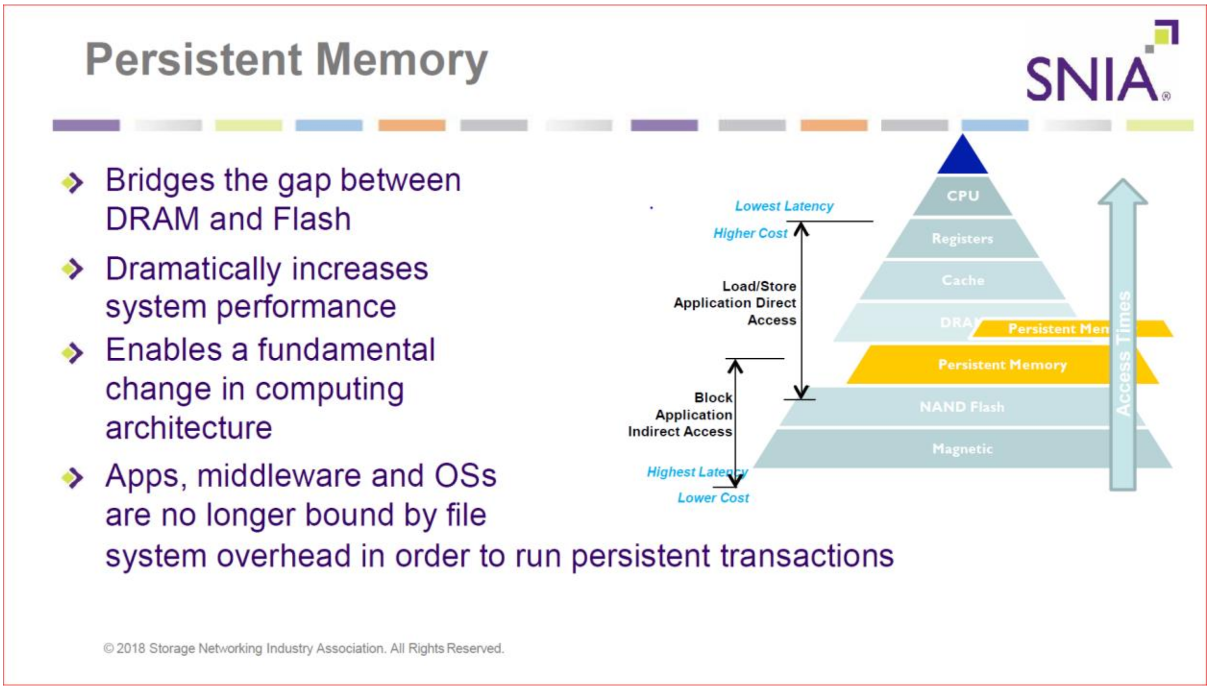

Figure Courtesy of the Storage Networking Industry Association

Essentially, NVM will allow CPUs to work with very, very large sets of data in memory, cost-effectively and with higher performance than is possible with older storage solutions. Non-volatile memory technologies under development include magnetoresistive random-access memory (MRAM), resistive random-access memory (RRAM), phase-change memory (PCM), low latency NAND Flash, and 3DXPoint, pronounced "three D cross point".

Intel Optane

In this paper, we will use Optane™ DC Persistent Memory (OPM) as our NVM example because it is the first to market in significant volumes. OPM is Intel's brand name for the Optane NVDIMMs that it developed under the code name Apache Pass, and it differentiates this NVM application from Optane's earlier use in SSDs.

Of particular interest in the NVM context is that Optane requires far fewer instructions to function than older memory types. Combined with the fact that no I/O is required, this reduces latency and improves system resource utilization.

Also significant is that Optane can work in two architectural modes. The first is memory mode, where it acts like a very large memory cache close to the CPU. The second is App Direct mode where the NVM operates as a very large byte-addressable memory, accessed directly by the CPU and its Optane-compatible operating system and file system. It is also possible to deploy Optane in mixed mode, combining elements of the two approaches.

Why is NVM important?

First, organizations today generate ever-increasing amounts of data, and they want to process more and more of it, as quickly as possible. But they also want to do that without huge costs. Optane-based NVDIMMs can help here.

A second factor is the consistency of storage performance that NVM can provide, which in turn makes application performance more predictable. This comes from the fact that, like SSDs but unlike hard disk storage, it uses silicon chips as its base, not a limited and mechanical system.

Note that there are also NVM configurations which implement the NVM as a large, closely-coupled, block device (drive) with DRAM cache, but which is not 100% persistent. You may need to discuss with your supplier which configuration option is best for your needs.

Where NVM can win

As we have said, NVM bridges the gap between fast but expensive memory and slow but cheap and high-capacity persistent storage. We therefore expect early adoption of NVM in general, and of Optane NVDIMMs in particular, in a number of key use cases.

- **In-memory databases:** The use of these technologies continues to ramp up rapidly as mainstream enterprise applications migrate to take advantage of them. NVM makes fast access to very large volumes of data both practical and affordable. NVM's persistent nature is especially valuable in scenarios where data accuracy and completeness is a legal essential.
- **Big Data and real-time analytics:** Enterprises and service providers want to exploit the information they have as quickly as possible and use it to make immediate decisions, not merely for retrospective reports.
- **HPC:** High performance computing requires fast processors and fast access to often very large volumes of data, and its use is spreading beyond research into operational enterprise use.
- **Very large VMs and dense populations of containers:** Virtualization as a primary deployment model for applications and services is very well established, but as virtualized platforms have expanded, so has their requirement for ever larger memory spaces. NVM will make it possible to run more VMs per system. The same need will apply as container usage expands.
- **Very fast storage:** In the data center, storage is no longer the 'essential but under-appreciated' element that it once was, and access to fast storage is now

a major factor driving system selection. NVM helps address this need, both as a fast, high capacity cache in front of traditional storage, and on its own.

- **Machine Learning and AI:** Machine learning is often most effective when very, very large sets of data are used for training models. The requirement for fast, high volume, cost-effective storage is clear.

As NVM becomes established and as the IT professional community gains experience in how to exploit it successfully, more use cases will be discovered. The speed at which this happens will partly depend on how quickly the price premium of NVM storage versus SSDs etc. decreases.

Ecosystem requirements

So far we have looked at NVM in isolation – as a technology in itself. However, like any other IT development, non-volatile memory can only be deployed if it is properly supported by the surrounding IT environment.

For example, the processor's companion chipset must understand NVM and include the additional instructions required to use NVDIMMs successfully, and without risking data corruption. This is especially important in the event of power failure – the chipset must interact with NVM to ensure all data is stored permanently when the power goes away.

Similarly, the processor's operating system and the file systems also need to support NVM. At a chipset level, Intel's latest generation of processors and chipsets, codenamed Cascade Lake, are the first to fully support 3DXPoint / Optane NVDIMM operations.

Operating system suppliers such as Microsoft, Red Hat, SUSE and Canonical are prepared for NVM. Relational database suppliers and enterprise software developers such as SAP, Oracle and SQL Server are also ready. NoSQL vendors such as MongoDB, Redis and RocksDB are prepared too, as are Big Data suppliers such as Cloudera, and major virtualization platforms such as VMware and Open Source virtual machine managers such as KVM.

In other words, it is an ecosystem play. NVM and Optane could exist on their own, but the benefits would be very limited without significant additional work. Fortunately, the ecosystem already has a very firm base from which to continue growing.

In conclusion

Permanent storage has evolved massively, but even SSDs remain separate from the processors and memory that actually use the data. This physical separation and the need for I/O channels to link memory to storage results in additional latency, meaning slower applications and services.

With non-volatile memory, our processors can use permanent storage as memory – memory that retains its data even when the power is switched off. In latency terms, this new persistent memory fits between traditional storage and DRAM, allowing it to sit much closer to the CPU than storage does, or even to replace the traditional DIMMs with NVDIMMs.

NVM is more affordable than DRAM for large memory capacities. This makes it attractive for use in many workloads, but it relies on the surrounding ecosystem providing good NVM support. It also requires IT architects and system operators to learn how to design and use NVM technology effectively and efficiently. And, of course, it adds yet another layer and aspect to data protection.

About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we help busy IT and business professionals get up to speed on the latest technology developments, and make better-informed investment decisions.

For more information, and access to our library of free research, please visit www.freeformdynamics.com or follow us on Twitter @FreeformCentral.

About Fujitsu

Fujitsu is the leading Japanese information and communication technology (ICT) company offering a full range of technology products, solutions and services. Approximately 140,000 Fujitsu people support customers in more than 100 countries. We use our experience and the power of ICT to shape the future of society with our customers.

For more information, please visit www.fujitsu.com.