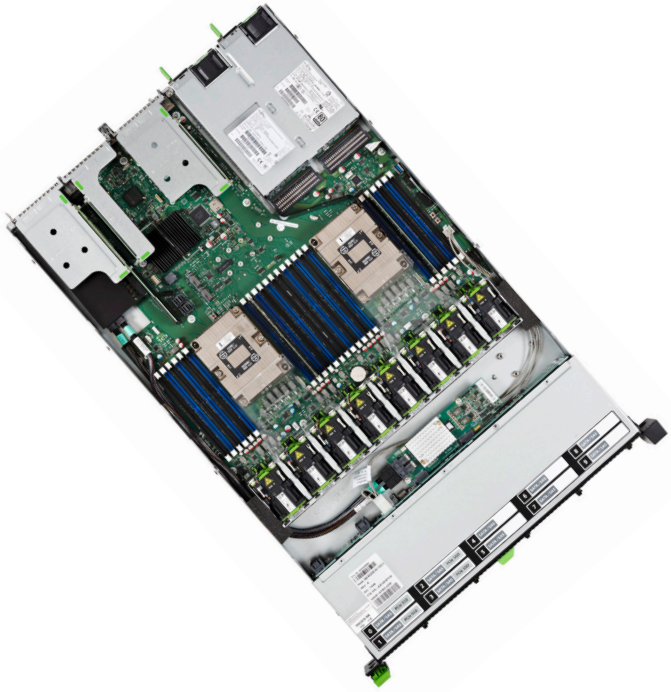
nextgenio newsletter

THE NEXTGENIO HARDWARE HAS LANDED!

We are very proud to be able to showcase the NEXTGenIO prototype system at the ISC 2018 show, utilising the revolutionary Intel® Optane™ DC Persistent Memory NVRAM technology. The prototype, running a realistic simulation of an HPC workload, is available to view on the **FUJITSU** booth, E-1020.



intel OPTANE™ DC PERSISTENT MEMORY

The prototype system, which the project has been developing for the past two and a half years, is the first step towards **true I/O for the Exascale**. The use of non-volatile memory is key as this allows us to bridge the gap between memory and storage in traditional HPC systems. The prototype system will be used to explore how the use of NVRAM technology can **improve I/O intensive** high-performance scientific computing applications.

As **demand on I/O** increases from across scientific and industrial computing sectors, the need for a **new class of memory** to cope with this demand has been greater than ever. The demonstration running on the prototype is typical of the kinds of applications that Intel® Optane™ DC Persistent Memory can **greatly benefit**: typical use cases include weather and climate modelling, which are extremely I/O intensive workloads. Weather forecasters often need to perform several forecasts per day, which produces an **enormous amount of data** for the system to analyse and store. If this data can be kept in NVRAM, the I/O load on the system drops significantly, allowing the process to be completed faster.

NEXTGenIO: Bridging the Gap

LET'S TALK ABOUT MEMORY...

There is a growing need for a new class of memory across to handle modern computing paradigms which require access to large datasets quickly. This is true across scientific computing disciplines - AI, Big Data analytics, Engineering, Deep Learning, as well as within the traditional HPC community. The needs of each community are different, but the solution is the same: eliminating the I/O bottleneck by the use of revolutionary non-volatile memory technology - in NEXTGenIO's case, working with Intel to utilise the Intel® Optane™ DC Persistent Memory NVDIMMs.

FUJITSU

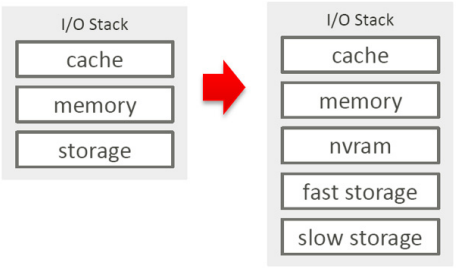
THE CURRENT SITUATION

Most modern IT infrastructures keep their data storage separate from compute; this is good in terms of security but means that users must continually access their data from disk during their compute operations. Conventional storage cannot keep up with the speeds DRAM offers, so time is wasted as the system waits for the data to be loaded from disk into DRAM. As an operation becomes more I/O intensive (such as a CFD or Data Analytics package), the inability to quickly access data significantly impacts on application performance.

USING NVRAM

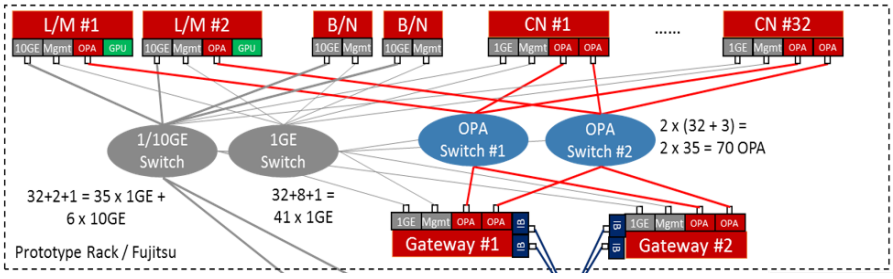
In a system utilising Intel® Optane™ DC Persistent Memory, which is available in capacities up to 512GB per DIMM, large data sets (such as, for example, multiple genome sequences for analysis, or collections of geometries for CFD) can be held as if they were in DRAM at all times, allowing fast access to the dataset for the running programme. NVRAM is slightly slower than conventional DRAM, but much, much faster than the usual PCIe SSDs, or even spinning disks, that would normally make up a storage system. Thanks to the much larger capacities of NVDIMMS, the data only needs to be loaded to the NVRAM modules once, and can then be written back to conventional storage when complete, eliminating the speed differential between disk and DRAM, and thus eliminating the bottleneck.

The introduction of NVRAM significantly alters the conventional memory and storage hierarchy, as seen in the diagram. This enables multiple powerful features, such as the ability of vehicle designers to alter simulation parameters on-the-fly, or weather forecasters to quickly run the many short simulations per day that they require to produce accurate forecasts and predict the probability of different weather conditions.



THE NEXTGENIO SYSTEM

The goal of the NEXTGenIO project has been to build a system from the ground up - both hardware and software - that offers significantly faster I/O than traditional HPC systems, and is also compatible with current applications. There are many codes and software suites that users in I/O intensive sectors simply need to continue using for many reasons, and it isn't practical to completely re-write these. Instead, a full system software stack has been developed which sits between the user and the system. This includes a job scheduler (SLURM), an application object store as alternative to conventional file systems (dataClay), which allows data objects to be stored in persistent memory without the need for serialization, and a multi-node NVRAM file system (echoFS). The compute nodes of the NEXTGenIO system are dual-CPU Intel® Xeon® SP nodes of up to 56 cores, each with 192GB of conventional DRAM and 3TB of NVRAM.



Recent Publications

The project has recently had papers accepted at two major international meetings. These papers cover different aspects of the work being done as part of the project.

24th International European Conference On Parallel and Distributed Computing

A Methodology for Performance Analysis of Applications Using Multi-layer I/O

Authors: Ronny Tschüter, Christian Herold, Bert Wesarg, and Matthias Weber

Abstract.
Efficient usage of file systems poses a major challenge for highly scalable parallel applications. The performance of even the most sophisticated I/O subsystems lags behind the compute capabilities of current processors. To improve the utilization of I/O subsystems, several libraries, such as HDF5, facilitate the implementation of parallel I/O operations. These libraries abstract from low-level I/O interfaces (for instance, Posix I/O) and may internally interact with additional I/O libraries. While improving usability, I/O libraries also add complexity and impede the analysis and optimization of application I/O performance.

In this work, we present a methodology to investigate application I/O behavior in detail. In contrast to current methods, our approach explicitly captures interactions between multiple I/O libraries. This allows to identify inefficiencies at individual layers of the I/O stack as well as to detect possible conflicts in the interplay between layers. We implement our methodology in an established performance monitoring infrastructure and demonstrate its effectiveness with an I/O analysis study of a cloud model simulation code. In summary, this work provides the foundation for application I/O tuning by exposing inefficiency patterns in the usage of I/O routines.

23rd International Workshop on High-Level Parallel Programming Models and Supportive Environments on the 32nd IEEE International Parallel and Distributed Processing Symposium

Visualization of Multi-layer I/O Performance in Vampir

Authors: Hartmut Mix, Christian Herold, and Matthias Weber

Abstract
Nowadays, high performance computing systems provide a wide range of storage technologies like HDDs, SSDs or network devices. With the introduction of NVRAM, these systems become more heterogeneous and finally provide a complex I/O stack that is challenging to use for applications. However, parallel programs have to efficiently utilize available I/O resources to overcome the scalability problem. Typically, performance analysis tools focus on investigating computation efficiency, executed program paths, and communication patterns. However, these tools only visualize I/O performance information of single layers of the I/O stack. To fully understand the I/O behavior of an application, it is necessary to investigate the interaction between the layers.

This work introduces new visualizations of I/O performance events and metrics throughout the complete I/O stack of parallel applications. We implement our approach on the basis of the performance analysis tool Vampir. We extend its timeline visualizations with performance details of I/O operations. Further, we introduce a new timeline view which depicts I/O activities on each layer of the used I/O stack as well as the interaction between layers. This view enables application developers to identify I/O bottlenecks across layers of a complicated I/O stack. We demonstrate our I/O performance visualization approach with a case study of a cloud model simulation code. Thereby, we analyze the I/O behavior in detail, including information of all involved multi-layered I/O libraries.

NEXTGenIO at ISC 2018

PARTNER BOOTHS

F-930

J-622

A-1412

K-500

B-1251

E-1020

EVENTS

The best place to come and talk to us at ISC is at any of our partners' booths, but you will be able to find us at different events throughout the conference as well.

ON BOOTH

Technische Universitat Dresden, who provide software solutions and tools as well as analysis of future Exascale Software for the project, will be displaying on-demand presentationson: an overview of the Score-P/Vampir performance tools and in-depth discussion of upcoming NEXTGenIO features.

TALK

Adrian Jackson, EPCC: "*NEXTGenIO: Exploiting non-volatile memory for HPC*" at the Workshop "*HPC I/O in the Data Center*", Thursday, June 28th, 9am - 6pm

BIRDS-OF-A-FEATHER SESSION

The project will be co-hosting a BoF session on the Tuesday of the conference, titled, *"Multi-Level Memory and Storage for HPC and Data Analytics".*
The speakers for this session will be:
Steve Pawlowski, Micron
Marcelo Cintra, Intel
Keeran Brabazon, ARM/Allinea
Sai Narasimhamurthy, Seagate
Tiago Quintino, ECMWF
Mark Parsons, EPCC
Dan Laney, LLNL
Christian Herold, TU Dresden
This session will take place from **3.45-4.45pm** in **Substanz 1&2.**