Using Non Volatile Memories to build energy and cost efficient clusters

ONKAR PATIL, North Carolina State University, USA
LATCHESAR IONKOV, Los Alamos National Laboratory, USA
JASON LEE, Los Alamos National Laboratory, USA
FRANK MUELLER, North Carolina State University, USA
MICHAEL LANG, Los Alamos National Laboratory, USA

Non-Volatile Memory (NVM) is a byte-addressable, high density and high latency memory technology that expands the memory hierarchy by another level. It adds to the capacity of main memory and expands the address space of applications but lowers access speeds.

Intel's Optane DC Persistent Memory Modules is a NVM device in DIMM form. It has the ability to persist the data stored in it without the need for power. It has up to 8x the capacity of DDR4 DRAM modules, which can expand the byte-address space up to 6 TB per node although it has up to 6x higher write latency [1; 2].

Today's scientific computing applications require High Performance Computing (HPC) clusters to have high number of compute and memory resources. Applications like VPIC, ACME and GROMACS support datasets that can reach up to petabytes. The training datasets for large scale Artificial Intelligence applications also require similar memory resources. These applications are run on some of the largest HPC systems in the world, e.g., the Summit supercomputing facility at Oak Ridge National Laboratory [3]. The Summit supercomputer has 4806 nodes with 512 GB of DDR4 DRAM on each node with additional 96 GB of HBM2 GPU memory. That accounts for a total main memory size of 2.67 PB. These nodes are connected by a high-speed interconnect, which uses Mellanox EDR Infiniband switches in a fat tree topology. The components in Summit consume 13 MW at its peak, which is a significant number for power consumption of a supercomputer and results in increased operating costs in addition to its already high initial setup costs around $200 million. To run larger problems in the future, we will need the memory address space to scale along with core count scaling to stay within reasonable energy and cost budgets. The hypothesis is that NVM can increase the memory density of a compute node in HPC systems, which enables large problem sizes to be run on fewer nodes. This can help reduce the cost of setting up and operating a supercomputer by lowering the power consumption and reducing the number of components required to build a supercomputer. Although, NVM has lower access latencies, due to its lower dependency on communication infrastructure, NVM creates a novel trade-off between performance and cost of operation.

To test this hypothesis, we conduct experiments on a single node with Intel's 48-core Cascade lake processor with Optane DC PMMs and 4 nodes with Intel's 44-core Skylake processor with no NVM connected over Mellanox EDR 100 GB switches. The node with NVM has 192 GB of DDR4 DRAM and 1.6TB of Optane DC NVRAM. The nodes without NVM have 384 GB of DDR4 DRAM, which adds up to equivalent amount of memory to the node with NVM. All the nodes are connected to rack mounted power distribution units to obtain measurements of the total energy consumption over all components in a node. The VPIC benchmark is run on both the node configurations under strong and weak scaling to assess the performance characteristics and power consumption using LIKWID. We use the 'lpi' input deck. The problem sizes of VPIC range from 300 GB to 1 TB which are allocated agnostic to the memory technology and we scale the processes from 1 to 96 using MPI.

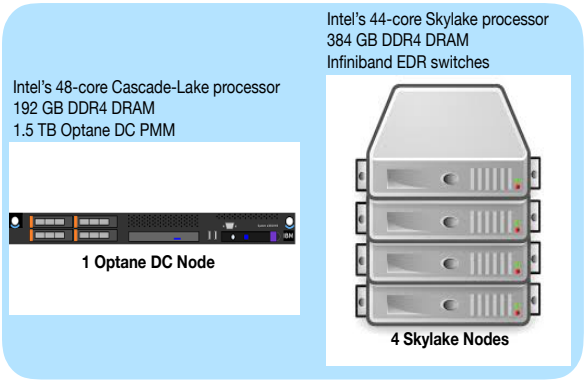

Fig. 1. Our experimental setup. Left: Single node with Optane DC PMMs Right: 4 nodes with equivalent amount of DRAM

We observe that the execution times for the node with NVM and nodes without NVM are similar with no significant difference for most executions in both strong and weak scaling cases. For both configurations, the bandwidth achieved by the NVM node is lower than the traditional nodes barring the single process executions. In spite of this, the NVM delivers comparable performance to the traditional nodes. This is because VPIC is a compute-bound application that optimizes cache hits and does not fetch from memory that often. For weak scaling, beyond 48 processes, the execution time doubles up due to oversubscribing of CPUs which fills the fetch queues on the memory controllers and instills backpressure. We observe that the NVM node consumes up to 3x less energy than the traditional nodes due to low power consumption of Optane DC and fewer auxiliary components. These energy savings are observed for most executions.

These results open up many opportunities for future use of Optane DC PMMs in HPC. In the future, we wish to develop novel policies for the same that can optimize the performance of NVM. We will look into static and dynamic approaches to optimize memory allocations for HPC applications executing on DRAM-NVM hybrid memory systems. NVM has potential to reduce the amount of energy and physical resources needed for future HPC systems effectively building cost and energy efficient clusters that

Authors' addresses: Onkar Patil, Department of Computer Science, North Carolina State University, USA, opatil@ncsu.edu; Latchesar Ionkov, Ultrascale Research Center, Los Alamos National Laboratory, USA, lionkov@lanl.gov; Jason Lee, Ultrascale Research Center, Los Alamos National Laboratory, USA, jasonlee@lanl.gov; Frank Mueller, Department of Computer Science, North Carolina State University, USA, mueller@cs.ncsu.edu; Michael Lang, Ultrascale Research Center, Los Alamos National Laboratory, USA, mlang@lanl.gov.

0730-0301/2018/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

support large problem sizes. Optane DC PMMs provide a great opportunity to extend memory address spaces of HPC systems without compromising on performance and delivering energy savings.

CCS Concepts: • **Computer Systems Organization → Architecture**; • **Computing methodologies → Massively parallel and high-performance simulations**.

Additional Key Words and Phrases: Optane DC, NVM, persistent memory

ACM Reference Format:
Onkar Patil, Latchesar Ionkov, Jason Lee, Frank Mueller, and Michael Lang. 2018. Using Non Volatile Memories to build energy and cost efficient clusters. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 2 pages. https://doi.org/10.1145/1122445.1122456

REFERENCES

[1] J. Izraelevitz, J. Yang, L. Zhang, J. Kim, X. Liu, A. Memaripour, Y. J. Soh, Z. Wang, Y. Xu, S. R. Dulloor, J. Zhao, and S. Swanson, "Basic performance measurements of the intel optane DC persistent memory module," *CoRR*, vol. abs/1903.05714, 2019.

[2] O. Patil, L. Ionkov, J. Lee, F. Mueller, and M. Lang, "Performance characterization of a dram-nvm hybrid memory architecture for hpc applications using intel optane dc persistent memory modules," in *Proceedings of the fifth ACM/IEEE International Symposium on Memory Systems*, pp. 288–303, ACM/IEEE, 2019.

[3] J. Hines, "Stepping up to summit," *Computing in science & engineering.*, vol. 20, no. 2, pp. 78–82, 2018.