



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Evaluation Methodology of an NVRAM-based Platform for the Exascale

**Citation for published version:**

Herrera, JFR, Prabhakaran, S, Weiland, M & Parsons, M 2018, 'Evaluation Methodology of an NVRAM-based Platform for the Exascale' 5th International Conference on Exascale Applications and Software, Edinburgh, United Kingdom, 17/04/18 - 19/04/18, pp. 11.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Evaluation Methodology of an NVRAM-based Platform for the Exascale

Juan F.R. Herrera,<sup>1</sup> Suraj Prabhakaran,<sup>2</sup> Michèle Weiland,<sup>1</sup> and Mark Parsons<sup>1</sup>

<sup>1</sup>*EPCC, The University of Edinburgh, UK, j.herrera@epcc.ed.ac.uk, m.weiland@epcc.ed.ac.uk, m.parsons@epcc.ed.ac.uk*

<sup>2</sup>*Intel GmbH, Germany, suraj.prabhakaran@intel.com*

One of the major roadblocks to achieving the goal of building an HPC system capable of Exascale computation is the I/O bottleneck. Current systems are capable of processing data quickly, but speeds are limited by how fast the system is able to read and write data. This represents a significant loss of time and energy in the system. Being able to widen, and ultimately eliminate, this bottleneck would majorly increase the performance and efficiency of HPC systems. The NEXTGenIO project is investigating this issue by bridging the latency gap between memory and disk through the use of non-volatile memory, which will sit between conventional DDR memory and disk storage. In addition to the hardware that will be built as part of the project, the project will develop the software stack (from OS and runtime support to programming models and tools) that goes hand-in-hand with this new hardware architecture. This project addresses a key challenge not only for Exascale, but also for HPC and data intensive computing in general: the challenge of I/O performance.

An application suite of eight memory and I/O-bound applications have been selected alongside a set of test cases for each application, to evaluate the platform's effectiveness regarding I/O performance and throughput. The application suite covers a wide range of fields, from computer-aided engineering to meteorology, computational chemistry, and machine learning. The output of the evaluation will document the benefits of the NEXTGenIO technology, and indicate its impact and future lines of development.

Three measurement scenarios are defined to assess the specific benefits of the NEXTGenIO technology:

- A. Baseline measurement in today's systems.
- B. Measurements on the NEXTGenIO platform without the use of non-volatile memory.
- C. Measurements on the NEXTGenIO platform with the use of non-volatile memory.

The profiling tools Allinea MAP and Score-P are used to collect the metrics needed to evaluate the performance of the applications for each scenario. These tools have been extended to support performance analysis with non-volatile memory.

In our presentation, we will present our methodology for evaluating the NEXTGenIO platform and show early memory and I/O profiling results for the applications. We will discuss how NVRAM will impact the performance of these applications.

## Acknowledgments

The NEXTGenIO project received funding from the EU Horizon 2020 research and innovation programme under grant agreement No 671591. This work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>).