The goal of the TADAAM project is to design and build a stateful system-wide service layer for HPC systems. This layer will be twofold. First, it will abstract low-level features of the system (e.g., topology, network, resource usage) and of the software stack (e.g., threads, data, runtime system). Second, applications will be able to register their needs and behaviors thanks to a carefully designed API. With these two sets of information, the layer will optimize the execution of all the running applications in a coordinated fashion and at system-scale.

Context

Scientific priorities:

- Post Moore's law computer (Heterogeneous and non-volatile memories)
- Extreme-scale computing for data intensive science (Data access pattern modeling)

Scientific Research Context:

High Performance Computing currently sees a large redesign of its hardware platforms. Beside many-core architectures and accelerators, the memory subsystem is dramatically changing with the arrival of new memory technologies. High-bandwidth memories emerged several years ago with the new requirements of HPC runtime systems. Platforms may now embed two kinds of memories with different performance characteristics. Hence researchers are working at convenient ways to decide where to allocate application data buffers. The issue will get harder in the next year with the arrival of non-volatile memory that can be used as normal memory. With persistence but lower performance, it bridges the large gap between memory and storage. It makes data placement decisions even harder since performance and access patterns have to be even more carefully involved in the decision process.

Assignment

The goal of this thesis is to develop tools and strategies for addressing upcoming complex memory subsystems in high-performance computing and big data. Current applications are manually tuned to allocate their data buffers to specific kinds of memory, for instance by statically identifying which buffers are bandwidth critical, or by comparing offline the performance of different placements. Upcoming persistent byte-addressable memory (NVDIMM) will significantly complexify this issue by adding yet another level of performance (slower than normal memory) and asymmetric read/write performance. This thesis aims at providing hints to users and to the software stack to better adapt existing HPC applications to these technologies that will be used in most HPC platforms in the next years.

Memory characteristics can be described using low-level benchmarks or using high-level abstractions. The roofline model is a well-known low-level approach that we already extended to better describe NUMA architectures and high-bandwidth memories. We plan to study the extension of this model to non-volatile memory and see how it can help understand and optimize HPC kernels on new memories. It includes developing and/or updating benchmarks to evaluate the new characteristics of these emerging memory technologies, and comparing the results with the performance of existing kernels.

However it is sometimes hard to precisely match observable application performance with the predictions of such models because they cannot take into account some characteristics such as data access patterns, different application phases, etc. Higher-level models try to abstract out application needs and hardware capabilities in rather simple ways. They are less precise but they provide easier-to-user criteria for users and runtimes to find out how to place data in different kinds of memories.

Main activities

The Inria TADAAM team notably focuses on defining such abstractions and modeling hardware platforms. Hence this thesis will also consist in defining abstractions for the new memories. For
instance, we will study criterias such as the amount of read and write access, the data being temporary or not, its reuse pattern, etc. They will be studied on a set of existing benchmarks to see how these characteristics can be used to infer whether a buffer should be allocated in a persistent, normal or high-bandwidth memory.

Then we will focus on new challenges brought by persistent memory that can be used as storage and as normal memory. Instead of using normal memory for computation with storage access at the beginning and end of computation phases, NVDIMMs will enable direct computation in the persistent memory where data files are stored. This removes the need to copy data between normal memory and storage, but slightly decreases access performance since NVDIMMs are slightly slower than normal memory. Hence there is a need to revisit how applications access and compute data by finding a tradeoff between these higher access costs and removing the copies between normal memory and storage. Both high-level coarse-grain algorithmics and benchmark-based models will be used to characterize when directly using persistent memory can be beneficial to applications.

This thesis will benefit from an ongoing collaboration with Intel which is expected to distribute NVDIMMs in the near future. In the meantime, NVDIMM performance and behavior can be emulated in software. The thesis will also use simulations for modeling even more memory kinds with different characteristics.

Keywords:
- Non-volatile memory
- Heterogeneous memory
- Data placement
- Modeling data access patterns
- High performance computing

References:
1. Nicolas Denoyelle, Brice Goglin, Aleksandar Ilic, Emmanuel Jeannot, and Leonel Sousa. *Modeling Large Compute Nodes with Heterogeneous Memories in the Cache-Aware Roofline Model*. In 8th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems, PMBS@SC 2017, Lecture Notes in Computer Science, Denver, CO, USA, November 2017. Springer. [http://hal.inria.fr/hal-01622582](http://hal.inria.fr/hal-01622582)
2. Brice Goglin. *Exposing the Locality of Heterogeneous Memory Architectures to HPC Applications*. In The Second International Symposium on Memory Systems Proceedings (MEMSYS16), pages 30-39, Washington, DC, October 2016. ACM. [https://hal.inria.fr/hal-01330194](https://hal.inria.fr/hal-01330194)
5. Jeff Layton. *How Persistent Memory Will Change Computing*. [http://www.admin-magazine.com/HPC/Articles/Persistent-Memory](http://www.admin-magazine.com/HPC/Articles/Persistent-Memory)

Skills

**Required knowledge and background:**

- Master degree in computer science
- Basic knowledge of parallel architectures and memory architectures (multicore, NUMA, caches, etc.)
- C programming
- Unix environments

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport

Remuneration

1982€ / month (before taxs) during the first 2 years, 2085€ / month (before taxs) during the third year.