

## Set 1 - Performance Measures

Issued: March 1, 2023

Understanding the characteristics of the platform on which performance tests are executed is of fundamental importance since it gives a context in which to read performance results. The primary objective of this exercise is to learn how to characterize computing hardware performance.

### Peak Performance and System Memory Bandwidth

The number of executed floating point operations (FLOP) is a measure used to characterize the costs of scientific software, e.g. computational fluid dynamics codes, finite element analysis programs, computational chemistry and computational biology packages.

The peak floating point performance, hereafter simply called peak performance, is a measure of the quantity of FLOP that a machine can execute in a given amount of time. Typically, the peak floating point performance (PP) in FLOP/s can be computed as in the following:

$$PP [FLOP/s] = f [HZ = cycle/s] \times c [FLOP/cycle] \times v [-] \times n [-], \quad (1)$$

where  $f$  is the core frequency in CPU cycles per second (given in Hz),  $c$  the number of FLOP executed in each cycle,  $v$  the SIMD width (in number of floats) and  $n$  the number of cores.

The exact values for the above features can be found at the hardware specifications or given by the system administrators. For example, for the Euler II cluster<sup>1</sup> we can find the specifications of the Intel Xeon E5-2680v3 here: <https://ark.intel.com/products/81908/>.

Typical floating point performance values are reported in GFLOP/s or TFLOP/s. Floating point performance is also used to assess the performance of a given algorithm implementation (<http://en.wikipedia.org/wiki/FLOPS>).

The bandwidth of the system memory is a measure of the speed of data movement from the system to caches and vice-versa. As the problems considered here in fit on a single node, we are mostly interested in quantifying the DRAM bandwidth.

Given the specifications of the DRAM memory, the theoretical memory bandwidth (PB) in B/s can be computed as follows:

$$PB [B/s] = f_{DDR} [Hz = cycle/s] \times c [channel] \times w [bit/channel/cycle] \times 0.125 [B/bit], \quad (2)$$

where  $f_{DDR}$  is the DDR clock rate,  $c$  the number of memory channels and  $w$  the bits moved through a channel per cycle (typically 64 bits). Typical bandwidths are reported in MB/s, GB/s or TB/s ([http://en.wikipedia.org/wiki/Memory\\_bandwidth](http://en.wikipedia.org/wiki/Memory_bandwidth)).

<sup>1</sup><https://scicomp.ethz.ch/wiki/Euler>

Memories based on the DDR (Double Data Rate) technology, such as DDR-SDRAM, DDR2-SDRAM, and DDR3-SDRAM<sup>2</sup>, transfer two data per clock cycle. As a result, they achieve double the transfer rate compared to traditional memory technologies (such as the original SDRAM) running at the same clock rate. Because of that, DDR-based memories are usually labeled with double their real clock rate. For example, DDR3-1866 memories actually work at 933 MHz transferring two data per clock cycle, and thus are labeled as being a “1,866 MHz” device, even though the clock signal does not really work at 1.866 GHz.

## Question 1: Performance Measures

- a) According to the wiki of the Euler cluster, each of the two sockets on a node hosts an Intel Xeon E5-2680v3 12-core Haswell CPU capable of delivering 480 GFLOP/s each. In addition, the theoretical memory bandwidth for each of the two sockets can reach 68.3 GB/s. Try to justify the above numbers.
- b) Try to find the nominal peak performance and the theoretical memory bandwidth for your laptop or desktop machine.
- c) Sometimes the nominal peak of a platform is not sufficient as you might want a more realistic upper bound on what performance you can effectively reach with a program. The objective of this exercise is to get as high as possible with your current knowledge.  
Write a small C/C++ benchmark code to measure the peak single precision performance of a given platform. Report the method you used to measure the peak performance. Try your code on your laptop or desktop system. Can you reach the theoretical peak performance reported before? Please explain your observations.

---

<sup>2</sup>[http://en.wikipedia.org/wiki/DDR3\\_SDRAM](http://en.wikipedia.org/wiki/DDR3_SDRAM)