Introduction

The tools for comprehension of High Performance Computing from theory, hardware and software give us the basis to go toward optimizations and benchmarking. We show that hybrid architectures seems to be the way to reach exascale in few years but many optimizations need to be done to fit the energy envelope. Our study focus on the behavior of accelerators compared to classical processor on irregular application. We showed that the downside of accelerators seems to be the synchronization and regularized work. In this part we propose our metric putting forward accelerators behavior on irregular problems.

This part is decomposed in three chapters: The first one details the mains bottlenecks, pitfalls and walls in HPC. We give explanations on our problems choices for the chapters two and three of this part.

The first problem for our metric characterizing the behavior of accelerators and more specifically GPUs focus on irregular computations. We choose the problem of Langford, known in our laboratory, and show how we can take advantage of the accelerator for this kind of problem.

The second problem we choose is focused on irregular memory accesses and communications. It is based on a benchmark we presented in the previous part, the Graph500 benchmark. s

Domain scientists, chemists, physicists, meteorologists, etc. always need better and more accurate simulations. This is why HPC always need better architecture and faster computation models. As presented on the previous parts several methods allows the computational power to grow from processors optimization but also memories and communications. Those limitations to reach tomorrow supercomputers are called *walls*. We start this chapter by describing them and were they are faced and in which benchmark.

We then give details on realistic domain scientists problems and irregularity. From those observation we propose the metrics that will be presented in the two other chapters of this part.

0.1 Walls

In this section we describe the main walls in HPC. Starting from memory wall, communication wall, power wall and finally computational wall.

0.1.1 Memory Wall

THis problem is targeting for the first time in [WM95]. The author explains that:

> We all know that the rate of improvement in microprocessor speed exceeds the rate of improvement in DRAM memory speed, each is improving exponentially, but the exponent for microprocessors is substantially larger than that for DRAMs.

The new release of accelerators also have an impact on the memory wall, the memory of the host processors and devices accelerators cannot be accessed directly and copies from one to the other are requested.

0.1.2 Communication wall

The multiplicity of racks, nodes and

0.1.3 Power wall

The energy consumption of nowadays and future supercomputers is the main wall in HPC. Indeed, an exascale supercomputer could be construct using several petascale supercomputer but, with todays architectures, will require a nuclear plant to operate. In this objective low energy consumption architectures need to be find. Power wall can be of two kind: the energy to power the machine itself and the many nodes, processors and accelerators but also, and not the least, the energy requires to handle the heat generated by the machine. For the second part many new technologies arise with direct water cooling in the supercomputer racks.

0.1.4 Computational wall

The computational wall is a conjugation of all the wall cited before. By increasing the memory wall, the energy consumption and the communications we can increase the overall computation of the supercomputer.

0.2 Benchmark

0.2.1 Irregular behavior

0.2.2 Our choices

0.3 Conclusion

Bibliography

[WM95] Wm A Wulf and Sally A McKee. Hitting the memory wall: implications of the obvious. *ACM SIGARCH computer architecture news*, 23(1):20–24, 1995.