The Ministry of Education and Science of the Russian Federation

Lobachevsky State University of Nizhni Novgorod

Computing Mathematics and Cybernetics faculty

The competitiveness enhancement program
of the Lobachevsky State University of Nizhni Novgorod
among the world's research and education centers

Strategic initiative
“Achieving leading positions in the field of supercomputer technology
and high-performance computing”

Introduction to MPI

Lecture 10. Parallel Computations for Systems with Distributed Memory.

Nizhni Novgorod

2014

Lecture_10_. Parallel Computations for Systems with Distributed Memory.

Citius, Altius, Fortius [1], the Olympic motto, is especially true for IT engineering. Implementing an empirical law first formulated in 1965 by Gordon Moore and still holding true is a matter of honour for any hardware manufacturer. The user's/consumer's point of view is best reflected by the following wording: computing system performance doubles every 18 months. We avoided the term "CPU" on purpose, as the end user is not interested in what improves the performance, be it CPU, accelerator or video card: the only thing that is important is better performance for the same money.

However, in the last few years it became clear that computer performance could no longer be improved by increasing CPU frequency, so the manufacturers, having opted for multicoreness as the main development path, had to ask software engineers for help. The existing sequential programs able to use only one core will not run faster on new CPU generation "for free", so parallel programming has to become ubiquitous.

In addition to the above, another Moore's law wording states that the computational performance available to mankind will double each 18 months. A tangible substantiation of this wording is the Top500 list [46] that ranks the world's most powerful computer systems updated twice a year. The 31st edition of Top500 in June 2008 evidenced passing the petaflop/s milestone by IBM Roadrunner [47] whose LINPACK benchmark performance reached 1,026 petaflop/s (the previous teraflop/s milestone was passed by Intel ASCI Red [48] in 1997, so we can see that within 11 years the performance peak grew three decades). The total performance of the systems constituting the 31st edition of Top500 reached 11.7 petaflop/s. Is this much or not? If the real performance of a good PC based on a quad-code CPU is about 20 Gflop/s, the total of Top500 will correspond to 500,000 such PCs. It is clear that this is only the top of the iceberg. According to Gartner, the total number of computers used worldwide exceeded 1 billion in 2008.

The Top500 data help identify typical trends of HPC development. The first version of Top500 was released in June 1993 and contained 249 shared memory multiprocessor systems and 97 single-processor supercomputers; more than 40% of solutions in this list were based on a platform developed by Cray. Only 4 years later, all single-processor supercomputers disappeared from Top500 giving way to the first cluster system with the performance of only 10 Gflop/s (which is 100 times less than that of ASCI Red topping the list); the then-new cluster systems now account for 80% of Top500 and are actually the main way to build supercomputers.

---

[1] Latin for "Faster, Higher, Stronger"

The main advantage of clusters that predetermined their omnipresence, has been the use of mass market hardware and software. Currently, 75% of the systems in the list are based on Intel processors, slightly more that 13% have IBM processors and 11% are AMD-based (each of the two remaining manufacturers, NEC and Cray, account for one system); 81% of the systems use only two types of communication network systems, i. e. Gigabit Ethernet or Infiniband; 85% of them are Linux-based. As we can see, the list shows little diversity, which is an obvious advantage from mass users' point of view.

However, users would be even more glad to own a desktop or, if worse comes to worst, a deskside supercomputer. So the clusters that brought the DIY supercomputer trend to the HPC industry meet this need in the best possible way. Now it is difficult to say what system can be called the world's first personal cluster. In any case, RenderCube [49] presented a namesake mini-cluster of 4 two-processor system in a compact 42-cm cubicle as early as in 2001.

The HPC personalization trend is developing more and more rapidly and has recently been welcomed by manufacturers of video cards whose performance improved so much that one may want to use them not only for graphical calculation, but as general purpose accelerators, too. The current solutions in this field are provided by NVIDIA (NVIDIA® Tesla™) and AMD (ATI FireStream™) and, being internal devices, demonstrate a stunning (compared to universal processors)peak performance exceeding 1 teraflop/s.

This lecture is devoted to computing system classification and discusses a number of characteristics of shared memory systems.

10.1. Parallelism as a basis of high performance computing

To say the least of it, computer system development is focused on speed and velocity of computation. If the performance of ENIAC, the first fully digital computer, was able to perform only several thousand operations per second, such top-ranking supercomputer as IBM RoadRunner can perform quadrillions (10^{15}) of operations per second The rate of computing hardware development is impressive: the computation speed has grown trillions (10^{12}) of times for the last 60 years! To get a better understanding of this extraordinary situation, one can see a spectacular example: if the automotive industry enjoyed the same growth, motor cars would weigh about 200 grams and require only a few litres of gasoline to run millions of kilometres!

The history of IT engineering is a fascinating description of notable sci-tech solutions, with the moment of joyful victory and bitter defeat. The problem of creating HPC systems is among today's most complex sci-tech challenges whose solution requires joint effort of numerous talented scientist and designers, use of state-of-the-art scientific and technical achievements and, at the same time, large-scale financial investments. Here, it is important to note that in spite of computational growth of 10^{12} times, response time of the hardware itself has grown only several million times. Such an additional effect is due to the introduction of *parallelism* on almost every stage of computation.

Without trying to speak about the history of parallelism in detail this time, we must mention assurance of independent operation of various hardware (CPUs and I/O devices), introduction of staged memory and CPU architecture improvement (superscalarity, pipelined nature, dynamic scheduling). For additional information on the history of parallelism see, for example, [30]; to sum the things up, let us make an important conclusion: many ways of CPU improvement have now been exhausted (indeed, the possibility of further CPU frequency growth is restricted by a number of technical challenges), so the most promising line of development consists in explicit support of multiprocessor computations.