Can parallelization save the (computing) world?

János Véghe, József Vásárhelyi and Dániel Drótos

Faculty of Mechanical Engineering and Informatics
University of Miskolc
Miskolc, Hungary
{jvegh,vajo,drdani}@mazsola.iit.uni-miskolc.hu

Index Terms—parallelization, performance, supercomputer, cloud, stalling, single-processor approach

EXTENDED ABSTRACT

As all other computing laws [1], the computing performance shows a "logistic curve"-like behavior, rather than unlimited growth. Since cca. 2000, the single-processor performance increased only marginally [2], unlike the need for more computing power in the everyday tasks. The stalling forced computer experts to look for alternative methods. The preferred way seems to be to continue the traditions of the single-processor approach: to assemble systems comprising several segregated processors, connected in various ways.

The Moore-observation has already been terminated in the sense that no more transistors can be added to a single processor, but persists in the sense that more transistors can be placed on a single chip in the form of several cores, complete systems or networks. In this way the nominal computing performance of processor keeps raising, but the task to produce some actual computing performance remains mostly for the software. Solving the task needs different parallelization methods, and the wide range of those distributed systems limits the computing performance in very different ways.

The today’s processors comprise a lot of different parallelization solutions [3]. Modern processors comprise several (hidden) processing units, but the efficacy of their utilization is rather poor: the theoretically achievable parallelism in most cases cannot even be approached.

The huge variety of the available parallelization methods, the extremely different technological solutions make it impossible to set up an accurate technological model. A (by intention strongly) simplified operating model, however, enables to provide a qualitative (or maybe: semi-quantitative) description of the operational characteristics of the parallelized systems.

One of the obvious fields where the computing performance is crucial and the number of the processors, their single-processor performance and the speed of their interconnection are critical resources, is supercomputing. Partly because of the growing size of computing tasks, partly because of the prestige value of having a higher number of supercomputers and/or higher performance supercomputer, a "gold rush" is experienced today with the goal to achieve the dream limit of 1 Eflop/s ($10^{18}$ flop/s). The well-documented, strictly controlled measurement database [4] enables to make both retrospective statistical conclusions on the logic of development behind the performance data, as well as to make predictions for the near future about the performance of supercomputers and its limitations. Applying the simple model to supercomputer operation it can be shown that supercomputers reached (or even exceeded) their technological bounds.

Some high performance clouds are offered as an alternative of supercomputers for certain computing tasks. The performance data [5] show that the relatively slow network connection limits the class of the tasks they can run effectively.

It is shown that the 70-years old computing paradigm itself limits the development of computing performance. For further developing the 50-years old idea [6] about making systems comprising cooperating processors must be renewed [7]. The need for cooperative computing is convincingly demonstrated by the success of the worlds first really cooperative processor [8]. An extension [9] to the computing paradigm, that considers both the technological state-of-the-art and the expectations agains computing, is also presented [10] and some examples of their advantageous features are demonstrated.

REFERENCES

[1] P. J. Denning and T. G. Lewis, “Exponential Laws of Computing Growth”, Comm. ACM, pp. 54-65, 2017.
[2] S. H. Fuller and L. I. Millett, “Computing Performance: Game Over or Next Level?”, Computer, vol 44, pp. 31-38, 2011.
[3] K. Hwang and N. Jotwani, “Advanced Computer Architecture: Parallelism, Scalability, Programmability” Mc Graw Hill, 2016.
[4] Supercomputing community, “The Top 500 supercomputers” https://www.top500.org/, 2016.
[5] E. Wustenhoff and T. S. E. Ng, “Cloud Computing Benchmark”, https://www.burstorm.com/price-performance-benchmark/1st-Continuous-Cloud-Price-Performance-Benchmarking.pdf, 2017
[6] G. M. Amdahl, “Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities”, AFIPS Conference Proceedings, pp. 483-485, 1967.
[7] J. Véghe, “Renewing computing paradigms for more efficient parallelization of single-threads”, IOS Press, Advances in Parallel Computing Vol. 29, pp. 305-330, 2018.
[8] F. Zheng et al, “Cooperative Computing Techniques for a Deeply Fused and Heterogeneous Many-Core Processor Architecture” J. Computer Science and Technology, Vol. 30, No 1, pp. 145–162, 2015.
[9] J. Véghe, “Introducing the Explicitly Many-Processor Approach”, Parallel Computing, Vol. 75, pp. 28–40, 2018.
[10] J. Véghe., “EMPAthY86: A cycle accurate simulator for Explicitly Many-Processor Approach (EMPA) computer.” https://github.com/jvegh/EMPAthY86, 2016.