Guido Schryen | Department of Management Information Systems | Paderborn University
Contact: guido.schryen@upb.de |www.misor.org
accepted for publication in Communications of the ACM (CACM)
Date: 21 Dec 2019

Integrating Management Science into the HPC Research Ecosystem

High Performance Computing (HPC) has been used in many academic disciplines for years, including systems medicine, biophysics, earth system sciences, material science, and astronomy. The relevance of HPC in research is leveraged by the insight that drawing on parallelism is the only known way to achieve substantial performance growth in the near future as sequential programs have stopped running faster on new generations of hardware. Usage statistics of academic supercomputing sites do not only reveal the power users in terms of disciplines; they implicitly uncover other scientific disciplines which have hardly used HPC to solve their research problems – despite the potential of HPC. One of these other areas is management science (MS).

Originally synonym for operations research, MS has become much broader in terms of problems studied and methods applied, including those related to econometrics, mathematical modeling, optimization, data mining and data analytics, engineering, and economics. Its multidisciplinary and often quantitative nature makes it a promising scientific field for HPC. Scientific use of HPC in MS has already covered the areas of computational economics, finance and risk analysis, operations research, and data analytics, among others. While it is not difficult to identify problems in MS that invite for applying HPC, it remains challenging to actually bring HPC to MS and to foster the position of MS in the HPC research ecosystem.

We argue that promising paths for integrating MS into this ecosystem include (1) the raise of awareness of the benefits that HPC might bring to MS, (2) the provision of a thorough understanding of both potential effiiciency gains and scalability limitations that may differ substanially depending on the particular type of MS problem to be solved, and (3) the need for computational thinking (including parallel design) beyond parallel programming in MS supported by the availability and usability of high-level languages at the application level.

The era of HPC in science

High Performance Computing (HPC) refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve problems in science, engineering, or business. HPC is usually realized by means of computer clusters or supercomputers. Interestingly, the June 2019 edition of the list of TOP500 supercomputer sites marks a milestone in its history because, for the first time, all 500 systems deliver a petaflop or more. But looking at the development of single core performance reveals that it has stopped growing due to heat dissipation and energy consumption issues. Concludingly, substantial performance growth has started to come only from parallelism, which, in turn, means that sequential programs will not run faster on successive generations of hardware.

Many academic disciplines have been using HPC for research. For example, HPC has become important in systems medicine for Alzheimer's research, in biophysics for HIV-1 antiviral drug development, in earth system sciences for weather simulations, in material science for discovering new materials for solar cells and LEDs, and in astronomy for exploring the universe. Usage statistics of academic supercomputing centers, which have started offering HPC as a commodity good for research, show a high diversity of scientific disciplines which make use of HPC in order to address unresolved scientific problems with computational resources that have been unavailable in the past. These statistics are also an indicator for other scientific disciplines which have hardly used HPC to solve their research problems. Undoubtly, several of these other research areas will hardly benefit from HPC, for example, because their research is not computationally intensive. But there are also areas where using massive computational resources can help solving scientific problems. One of these areas is management science, including its strong link to economics.

Management science benefits from HPC

Management science (MS) refers to any application of science to the study of management problems. Originally synonym for operations research, MS has become much broader in terms of problems studied and methods applied, including those related to econometrics, mathematical modeling, optimization, data mining and data analytics, engineering, and economics. Its multidisciplinary and often quantitative nature makes it a promising scientific field for HPC. The potential of applying HPC to MS includes the improvement of efficiency ("solving a problem faster"), effectiveness ("solving a problem of larger size and/or with enhanced quality") and robustness ("solving a problem in a way that makes the solution robust against changes").

Although MS very rarely occurs on HPC usage statistics, the potential of HPC for solving problems in MS has become being tapped in several of its subfields. A first example are fraud detection services in finance provided by Bertelsmann Arvato Financial Solutions [8]. Machine learning approaches are integrated into the real-time analysis of transaction data. These approaches are based on the development of self-learning analytical models from past fraud cases for early recognition of new fraudulent cases. From a technological perspective, the Hadoop framework and the Microsoft Azure cloud infrastructure are used.

A second example is the provision of the cloud-enabled software *PathWise®* (provided by Aon) that uses HPC to manage financial guarantee risk embedded in life insurance retirement products [2]. Applying HPC capabilities allows reducing time required to evaluate policies from hours and days to minutes through a variety of approaches, including the parallelization of Monte-Carlo simulations.

A third example of benefiting from HPC when solving problems in MS is the parallelization of algorithms for solving discrete optimization problems in operations research. In [9], the authors parallelize a branch-and-price algorithm for solving the "unrelated parallel machine scheduling problem with sequence-dependent setup times". Their computational experiments conducted on a large university cluster used MPI (Message Passing Interface) to connect 960 computing nodes. Results on efficiency show that their parallelization approach can even lead to superlinear speedup.

Further examples can be found in economics where dimensional decomposition for dynamic stochastic economic models has been implemented on a supercomputer [3] and equilibria in heterogeneous agent macro models have been solved on HPC platforms [7]. In data analytics, HPC has been applied not only to detect financial fraud but also to solve problems in social network analysis (e.g., [11]) and in global health management (e.g., [6]). Overall, the applicability of HPC to problems occurring in MS shows a large methodological diversity, with methods from machine learning and artificial intelligence, simulation, and optimization being included, and the identification of many computational problems in MS that may be solved through HPC is not difficult. However, what turned out to be difficult is bringing HPC to MS and fostering the position of MS in the HPC research ecosystem. This deployment essentially targets at exploiting technical HPC capabilities for solving managerial problems through i) raising awareness of this potential in the MS community, ii) implementing HPC-related education for MS researchers, and iii) providing software development support with frameworks, libraries, programming languages, etc., which focus on solving specific types of problems occurring in MS.

Raising Awareness of HPC Benefits for Management Science

Despite promising applications of HPC in different areas of MS, its deployment in MS is far away from being an established and well-known research approach. This subordinate role of HPC in MS is reflected in different phemomena, which, at the same time, point to opportunities for raising awareness of HPC benefits in MS and for informing researchers that HPC is not identical to "High Performance Technical Computing".

In the MS community, only very rarely (special issues of) journals, workshops, or conference tracks which are dedicated to HPC-based research can be found. A few examples exist (e.g., [10]) but much more of these efforts to identify and communicate the potential that HPC brings for MS and to foster corresponding research is needed. In addition, MS departments may profit from introducing HPC to (master and PhD) students and young scientists by offering HPC courses and HPC summer/winter schools, in close cooperation with HPC sites of universities. Such courses and schools are of particular benefit when MS students and researchers can bring their own problems, algorithmic blueprints, or codes, and learn how to think, design, and implement parallel. In short, we need a more thorough computational and HPC-oriented education of MS students, which are the MS scientists of tomorrow.

HPC sites at universities and research institutions today generally focus on their current "power users", which are scholars from the natural sciences, engineering disciplines, medical sciences, among others. Often, the expectation of these sites on users' expertise includes a clear understanding of how HPC works technically (e.g., shared vs. distributed memory), which parallel programming paradigms exist (threads, processes, etc.), which libraries and APIs are state-of-the-art (OpenMP, MPI, CUDA, etc.), and how parallel programming should be done (take care of data races, deadlocks, etc.). Unfortuntely, MS researchers often do not (need to) have this deep knowledge for understanding their research field. This gap between expected and existing knowledge of HPC finally prevents MS researchers from tapping the potential that HPC might bring to their research. HPC sites should contribute to closing this gap by providing high-educational courses dedicated to MS. It might be helpful to bridge the gap by establishing and financing (jointly with MS departments) positions of HPC-MS engineers.

Finally, funding programmes dedicated to computational and HPC research in MS are likely to foster the awareness of HPC benefits for MS and the attractiveness of HPC for MS researchers.

Scalability of parallel applications in Management Science can differ fundamentally

Depending on the specific type of MS problem to be solved, the parallelization of algorithms may scale substantially differently over the number of parallel processing units. We argue that it is important to thoroughly inform MS researchers on issues of efficiency and scalability so that they can assess what to expect when solving their particular problem types with HPC.

Some research problems in MS involve solving specific instances of an optimization problem. In such cases, often a fixed-size model (constant total workload, variable exection time) occurs and *strong scaling* applies: according to *Amdahl's law* [1], the speedup factor that can be be achieved from parallelization is upper-bounded by 1 divided by the serial fraction of code, which is always larger than zero in practice. Due to execution time required for coordination, this upper bound is usually not achieved, and speedup values even start dropping when a particular number of parallel processing units (we refer to these as "processors") is exceeded. Even when ignoring all coordination efforts, a serial fraction of code amounting to 20%, for example, would limit the maximum speedup by the factor of 5 regardless of the number of processors. Consequently, MS researchers need to be informed on speedup, efficiency, and scalability that can be expected and its' determining factors.

Other problems in MS, often occurring in data analytics and in a real-time decision making context, follow a scale-size model (variable total workload, constant workload per processor, constant exection time). Then, according to *Gustafsons's law* [5], *weak scaling* applies, which means that speedup can increase (almost) linearly with the number of processors (even when coordination efforts are considered).

Applications need not primarily to be rewritten but rethought by (re-)design

While some sequential applications can be parallelized straightforward, limiting parallelization efforts to the implementation phase is myopic. As noted in [4], "a*ttempts to extract parallelism from the serial implementations are unproductive exercises and likely to be misleading if they cause one to conclude that the original problem has an inherently sequential nature."* (p. 135) Thus, it is important to support MS researchers in designing parallel algorithms and to release them from parallel implementation and technical issues. While this approach is useful for scholars of all scientific disciplines, it is of particular importance in fields where researchers are not used to programming-intensive tasks, as this is often the case in the MS community.

Several frameworks that are applicable to MS have already been suggested (e.g., Apache Hadoop, Ubiquity Generator by ZIB, Branch-Cut-Price framework in COIN-OR) or are under development (e.g., the PASC project "Framework for computing equilibria in heterogeneous agent macro models"); however, we should strengthen our efforts to develop IT artifacts that support MS researchers in parallel design and parallel implementation . In particular, high-level languages at the application level rather than multi-purpose parallel languages at the programming level would need to be provided. The availability and usability of such application languages would allow MS reseachers to focus on parallel design issues and release them from writing parallel code at the programming level, which would be generated automatically by application language compilers. We deem such approaches appropriate for deploying HPC in MS at a large scale. Concluding, we see an auspicious time for integrating MS into the HPC research ecosystem, and the MS community can look forward to the promising developments to come.

References

[1] Amdahl. G.M. (1967). Validity of the Single-Processor Approach to Achieving Large Scale Computing Capabilities. In: AFIPS Conference Proceedings, Vol. 30, AFIPS Press, Reston. Va., 483-485.

[2] Aon (2019). PathWise®, https://www.aon.com/reinsurance/PathWise-(1)/default.jsp

[3] Eftekhari, A., Scheidegger, S., & Schenk, O. (2017). Parallelized Dimensional Decomposition for Large-Scale Dynamic Stochastic Economic Models. In: Proceedings of the Platform for Advanced Scientific Computing Conference.

[4] Fuller, S. H. & Millett, L. I. (2011). The Future of Computing Performance: Game Over or Next Level? National Academy Press.

[5] Gustafson, J. L. (1988). Reevaluating Amdahl's Law. Communications of the ACM, 31(5), 532-533.

[6] Juric, R., Kim, I., Panneerselvam, H., & Tesanovic, I. (2017). Analysis of ZIKA Virus Tweets: Could Hadoop Platform Help in Global Health Management? In: Proceedings of the 50th Hawaii International Conference on System Sciences.

[7] Kübler, F., Scheidegger, S., & Schenk, O. (2017). Computing Equilibria in Heterogeneous Agent Macro Models on Contemporary HPC Platforms, www.pasc-ch.org/projects/2017-2020/computing-equilibria-in-heterogeneous-agent-macro-models/

[8] Microsoft (2017). Big data: Improving fraud recognition with Microsoft Azure, https://customers.microsoft.com/en-us/story/arvato-azure-powerbi-germany-media-inovex

[9] Rauchecker G. & Schryen G. (2019). Using High Performance Computing for Unrelated Parallel Machine Scheduling with Sequence-Dependent Setup Times: Development and Computational Evaluation of a Parallel Branch-and-Price Algorithm. Computers & Operations Research (104), 338-357.

[10] Schryen, G., Kliewer, N., & Fink, A. (2018). Call for Papers Issue 1/2020 – High Performance Business Computing. Business & Information Systems Engineering, 60(5), 439-440.

[11] Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-Scale Network Analysis for Online Social Brand Advertising. MIS Quarterly, 40(4), 849-868.