

Keynotes

EWA DEELMAN

University of Southern California
Information Sciences Institute
USA
e-mail: deelman@isi.edu
Web <http://deelman.isi.edu/>



Automating Science Workflows: Challenges and Opportunities

Science workflows help define the processes needed to understand our past, explain our world today, and predict how our planet will evolve tomorrow. They help to logically move from a hypothesis to its testing via appropriate methods, to the generation of findings and their publication. As in other areas of our lives, automation is increasing scientific productivity and is enabling researchers to analyze vast amounts of data (from remote sensors, instruments, etc.) and to conduct large-scale simulations of underlying physical phenomena. These applications comprise thousands of computational tasks and process large, heterogeneous datasets, which are often distributed across the globe. Computational workflows have emerged as a flexible representation to declaratively express the complexity of such applications with data and control dependencies. Automation technologies have enabled the execution of these workflows in an efficient and robust fashion. Up to now automation was based on a variety of algorithms and heuristics that transformed the workflows to optimize their performance and improve their fault tolerance. However, with the recent increased use of AI for automation, new solutions for workflow management systems can be explored. This talk describes some of the unsolved problems in workflow management and considers potential application of AI to address these challenges.

Biography

My main area of research is distributed computing. I research how to best support complex scientific applications on a variety of computational environments, including campus clusters, grids, and clouds. I have designed new algorithms for job scheduling, resource provisioning, and data storage optimization in the context of scientific workflows.

Since 2000, I have been conducting research in scientific workflows and have been leading the design and development of the [Pegasus](#) software that maps complex application workflows onto distributed resources. Pegasus is used by a broad community of researchers in astronomy, bioinformatics, earthquake science, gravitational-wave physics, limnology, and [others](#).

I am also the Principle Investigator for the CI CoE pilot, which provides leadership, expertise, and active support to cyberinfrastructure practitioners at NSF Major Facilities and throughout the research ecosystem in order to enable ongoing evolution of our technologies, our practices, and our field, ensuring the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depend.

In addition, I am interested in issues of distributed data management, high-level application monitoring, and resource provisioning in grids and clouds.

For the latest news, please check the [Pegasus blog](#).

GEOFFREY FOX

Indiana University, USA
e-mail: gcfo@indiana.edu
web: https://www.dsc.soic.indiana.edu/geoffrey_fox



Advancing Science with Deep Learning, HPC, Data Benchmarks and Data Engineering

- We describe how High-Performance Computing (HPC) can be used to enhance Big Data and Machine Learning (ML) systems (HPC for ML) but also how machine learning can be used to enhance system execution (ML for HPC) with promising deep learning surrogates
- We review the different aspects of data engineering needed to process largescale data and how it is implemented in the Cylon and Twister2 systems to support deep learning and Python notebooks. <https://cylondata.github.io/cylon/> and <https://twister2.org/>
- We give application examples from COVID-19 daily data, solutions of ordinary differential equations, and earthquakes
- We show how working with the industry consortium MLPerf, we may be able to establish a collection of science data benchmarks demonstrating best practices and motivating the next generation cyberinfrastructure.

Biography

Fox received a Ph.D. in Theoretical Physics from Cambridge University, where he was Senior Wrangler. He is now a distinguished professor of Engineering, Computing, and Physics at Indiana University, where he is the director of the Digital Science Center. He previously held positions at Caltech, Syracuse University, and Florida State University after being a postdoc at the Institute for Advanced Study at Princeton, Lawrence Berkeley Laboratory, and Peterhouse College Cambridge. He has supervised the Ph.D. of 73 students and published around 1500 papers (over 540 with at least ten citations) in physics and computing with a h-index of 82 and over 38000 citations. He is a Fellow of APS (Physics) and ACM (Computing) and works on the interdisciplinary interface between computing and applications. Particular interests are deep learning for observational science time series, cyberinfrastructure for data engineering, and benchmarking. He is involved in several projects to enhance the capabilities of Minority Serving Institutions. He has experience in online education and its use in MOOCs for areas like Data and Computational Science.



PIOTR SANKOWSKI

University of Warsaw,
Institute of Informatics
Poland
e-mail: P.Sankowski@mimuw.edu.pl
web: <https://www.mimuw.edu.pl/~sank/>



Breaking the PRAM $O(\log n)$ complexity bounds on MPC

For over a decade now we have been witnessing the success of massive parallel computation (MPC) frameworks, such as MapReduce, Hadoop, Dryad, or Spark. One of the reasons for their success is the fact that these frameworks are able to accurately capture the nature of large-scale computation. In particular, compared to the classic distributed algorithms or PRAM models, these frameworks allow for much more local computation. The fundamental question that arises in this context is though: can we leverage this additional power to obtain even faster parallel algorithms? In particular, graph connectivity seems to require $O(\log n)$ rounds under the 2 Cycle Conjecture. It is thus entirely possible that in this regime, which captures, in particular, the case of sparse graph computations, the best MPC round complexity matches what one can already get in the PRAM model, without the need to take advantage of the extra local computation power. In this talk, I will summarize our work on refuting that perplexing possibility. In particular, I will introduce the following MPC results that improve upon $O(\log n)$ time algorithms implied by PRAM results:

- an $(2 + \epsilon)$ approximation to maximum matching, for any fixed constant $\epsilon > 0$, in $O(\log^2 \log n)$ rounds in the case of slightly sublinear memory per machine regime [1],
- an $(1 + \epsilon)$ -approximation to PageRank in $O(\log^2 \log n + \log^{21}(\epsilon))$ rounds [2].

Based on joint work with Artur Czumaj, Jakub Lacki, Aleksander Madry, Slobodan Mitrovic, and Krzysztof Onak.

References

- Czumaj, A., Łącki, J., Mańdry, A., Mitrović, S., Onak, K., Sankowski, P.: Roundcompression for parallel matching algorithms. In: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing. p. 471–484. STOC2018, Association for Computing Machinery, New York, NY, USA (2018). doi.org/10.1145/3188745.3188764
- Łącki, J., Mitrović, S., Onak, K., Sankowski, P.: Walking randomly, massively, and efficiently. In: Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing. p. 364–377. STOC 2020, Association for Computing Machinery, New York, NY, USA (2020). doi.org/10.1145/3357713.338430

Biography

Piotr Sankowski is a Professor of Computer Science at Warsaw University. He has received habilitation in computer science in 2009 and a Ph.D. degree in computer science in 2005 both from the University of Warsaw. He has been working on combinatorial optimization problems with special emphasis on dynamic computations and stochastic properties of data. Moreover, he received a Ph.D. Degree in physics in 2009 from Polish Academy of Sciences - he has been working on a solid-state theory. In 2010 he has obtained the ERC Starting Independent Researcher Grant, in 2015 ERC Proof of Concept grant, and in 2017 ERC Consolidator Grant. He is a co-founder of MIM Solutions - a spin-off company working on the interplay of ML and algorithms.



Submission deadline for the Euro-Pas PhD Symposium has been extended to 27 May, 2023. Click here for more information - <https://t.co/wWxisCJSC>

17.05.2023 - 11:51

The Euro-Par PhD Symposium is a welcoming and supportive forum for PhD students to present their work. Click here for more information: <https://t.co/wWxisCJSC>

04.04.2023 - 09:25

Submit your paper for EURO-PAR 2023 Workshops and Minisymposia! [Click here for more information.](#) <https://t.co/UEseXWb3Dz>

07.03.2023 - 08:18

Abstract submission is due tomorrow 24 Feb, 2023 <https://t.co/eH2C9CRZA3>

23.02.2023 - 08:00



SPONSOR DETAILS

SHARE ON:

