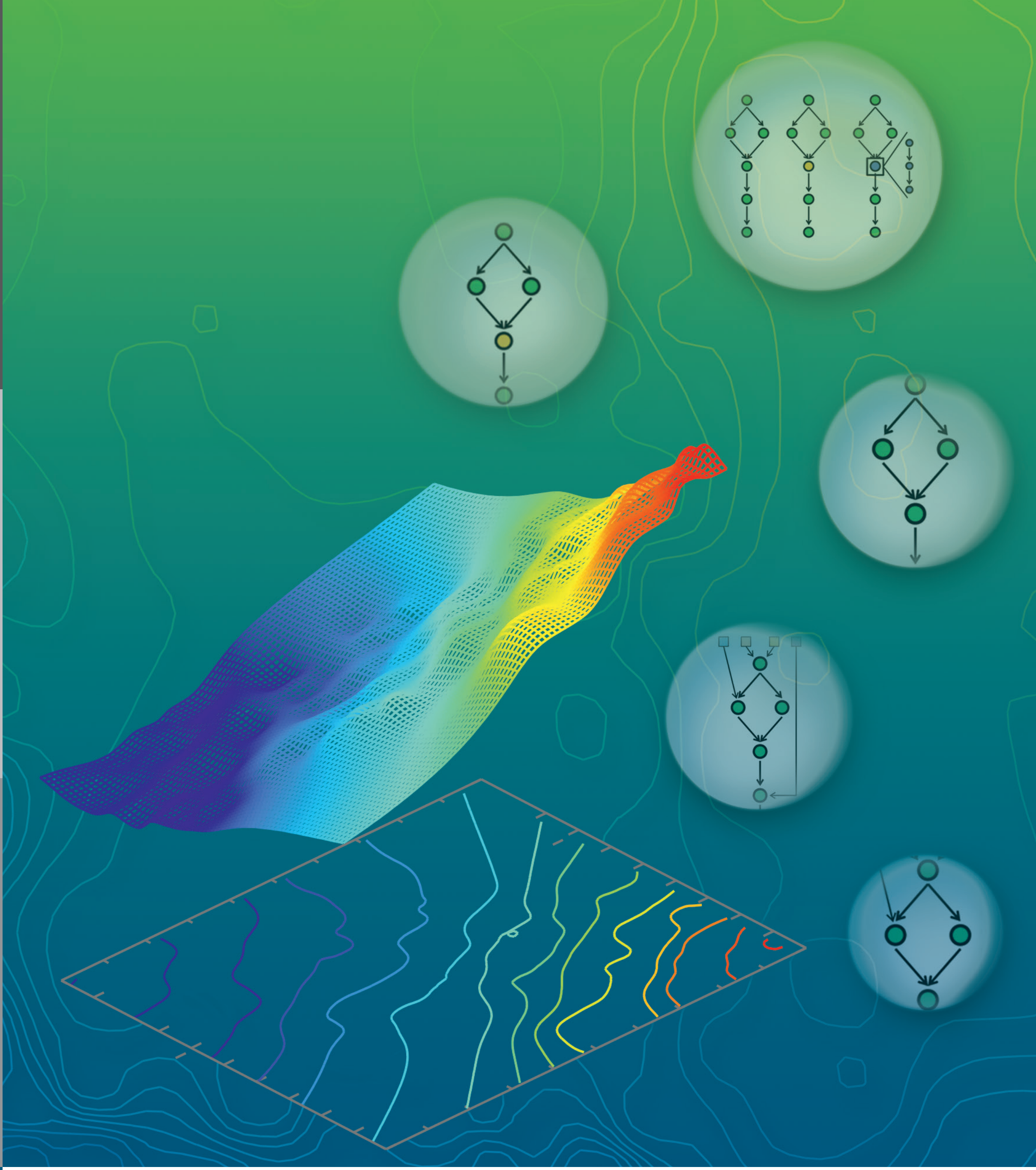

Automated Optimization Methods for Scientific Workflows in e-Science Infrastructures

Sonja Holl

JÜLICH
FORSCHUNGSZENTRUM

Member of the Helmholtz Association

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)

Automated Optimization Methods for Scientific Workflows in e-Science Infrastructures

Sonja Holl

Schriften des Forschungszentrums Jülich

IAS Series Volume 24

ISSN 1868-8489 ISBN 978-3-89336-949-2

Contents

Scientific workflows have emerged as a key technology that assists scientists with the design, management, execution, sharing and reuse of *in silico* experiments. Workflow management systems simplify the management of scientific workflows by providing graphical interfaces for their development, monitoring and analysis. Nowadays, e-Science combines such workflow management systems with large-scale data and computing resources into complex research infrastructures. For instance, e-Science allows the conveyance of best practice research in collaborations by providing workflow repositories, which facilitate the sharing and reuse of scientific workflows. However, scientists are still faced with different limitations while reusing workflows. One of the most common challenges they meet is the need to select appropriate applications and their individual execution parameters. If scientists do not want to rely on default or experience-based parameters, the best-effort option is to test different workflow set-ups using either trial and error approaches or parameter sweeps. Both methods may be inefficient or time consuming respectively, especially when tuning a large number of parameters. Therefore, scientists require an effective and efficient mechanism that automatically tests different workflow set-ups in an intelligent way and will help them to improve their scientific results.

This thesis addresses the limitation described above by defining and implementing an approach for the optimization of scientific workflows. In the course of this work, scientists' needs are investigated and requirements are formulated resulting in an appropriate optimization concept. This concept is prototypically implemented by extending a workflow management system with an optimization framework. This implementation and therewith the general approach of workflow optimization is experimentally verified by four use cases in the life science domain. Finally, a new collaboration-based approach is introduced that harnesses optimization provenance to make optimization faster and more robust in the future.

This publication was written at the Jülich Supercomputing Centre (JSC) which is an integral part of the Institute for Advanced Simulation (IAS). The IAS combines the Jülich simulation sciences and the supercomputer facility in one organizational unit. It includes those parts of the scientific institutes at Forschungszentrum Jülich which use simulation on supercomputers as their main research methodology.

IAS Series
Volume 24
ISBN 978-3-89336-949-2

JÜLICH
FORSCHUNGSZENTRUM