



[Prof. Dr. Achim Streit //
**Verteilte und Parallele
Hochleistungssysteme**]

112

Achim Streit studierte von 1994 bis 1999 Ingenieurinformatik mit Nebenfach Elektrotechnik an der TU Dortmund. 2003 promovierte er über Job Scheduling Strategien für verteilte und parallele Computing-Systeme an der Universität Paderborn; während dieser Zeit war er wissenschaftlicher Mitarbeiter am Paderborn Center for Parallel Computing (PC²). Mitte 2005 wurde er Leiter der Abteilung „Verteilte Systeme und Grid Computing“ im Jülich Supercomputing Centre (JSC) am Forschungszentrum Jülich. Mitte 2010 wechselte er an das Karlsruher Institut für Technologie (KIT) als Mitglied des Direktoriums des Steinbuch Centre for Computing (SCC) und wurde zum Informatikprofessor ernannt.

Er ist KIT-Programmsprecher des Helmholtz Programms „Engineering Digital Futures“, zu dem die Informatik-Aktivitäten in Computational & Data Science, die IT-Sicherheitsforschung in KASTEL und die Technikfolgenabschätzungsforschung gehören. Als SCC-Direktor ist er verantwortlich für GridKa, das deutsche Tier-1 Daten- und Analysezentrum insb. für den LHC am CERN, die Large-Scale Data Facility (LSDF) sowie für die Forschungsaktivitäten zu Data Science, Datenmanagement, Metadaten, AAI, KI, Quantum Maschine Learning, Research Software Engineering; weiterhin für die SCC-Beteiligung in zahlreichen Drittmittelprojekten u. a. zur European Open Science Cloud (EOSC) und Nationalen Dateninfrastruktur (NFDI). Er ist Initiator der Helmholtz Information & Data Science Academy (HIDA) und Mitglied in den Graduiertenschulen KSETA und HIDSS4Health sowie den KIT-Zentren KCIST, KCETA, Klima und Umwelt sowie MathSEE.

// Überblick und Allgemeines

Die Professur Verteilte und Parallele Hochleistungssysteme ist eng in das SCC integriert und erlaubt so eine enge Verzahnung von Forschung, IT-Forschungs-großgerät und praxisorientierter Lehre.

Im **Forschungsdatenmanagement** werden Methoden und Technologien zum verteilten und sicheren Umgang mit Forschungsdaten, zu Metadaten, zu verteilten Authentifizierungs- und Autorisierung-Infrastrukturen (AAI) bis hin zur Datenarchivierung erforscht und entwickelt, die in Kooperation mit nationalen und internationalen Partnern in der Praxis eingesetzt werden.

Im **Data Intensive Computing** werden Methoden und Technologien zur Datenanalyse, zu skalierbaren Methoden der Künstlichen Intelligenz, zum Quantum Machine Learning, zur effizienten Nutzung verteilter Rechenressourcen sowie zum Job-Scheduling erforscht. Die entstehenden Verfahren und Anwendungen werden dabei stets unter Aspekten der Skalierbarkeit, der Recheneffizienz und eines möglichst geringen Energieverbrauchs auf heterogenen Rechensystemen einschließlich GPUs und Quantencomputern entwickelt.

Zahlreiche der oben genannten Forschungsaktivitäten erfolgen in **interdisziplinärer Kooperationen** mit anderen

Forschungsdisziplinen z. B. Teilchenphysik, Klima- und Umweltforschung, Materialforschung, Energieforschung bis hin zu den Geisteswissenschaften. Die dabei entstehenden Software-Lösungen und Datensätze werden im Sinne offener und transparenter Wissenschaft frei zugänglich veröffentlicht. In der Entwicklung von Software wird stark auf Methoden des Research Software Engineering gesetzt.

In der **Lehre** werden die genannten Themen u. a. in den Vorlesungen „Parallelrechner und Parallelprogrammierung“ und „Verteiltes Rechnen“ sowie in Seminaren und Praktika praxisnah vermittelt.

// Ergebnisse und Erfolge

Alle Forschungstätigkeiten finden im Rahmen des Helmholtz-Programms „Engineering Digital Futures“ sowie der Helmholtz-Plattformen zu Metadaten (HMC), zu künstlicher Intelligenz (Helmholtz.AI) sowie zu föderierten IT-Diensten (HIFIS) statt. Die Arbeiten werden durch erfolgreiche Einwerbung einer Vielzahl nationaler und internationaler Forschungsprojekte flankiert und ergänzt.

Das SCC koordiniert gemeinsam mit dem Forschungszentrum Jülich die „Helmholtz Platform for **Research Software Engineering** – Preparatory Study“ zur Etablierung von Research Software Engineering Aktivitäten im Helmholtz Forschungsbereich Information. Dies wird durch Community Software Infrastructure (CSI) Gruppen für ausgewählte Codes sowie die systematische Etablierung und Nutzung von Cx-Umgebungen für Continuous Integration, Testing, Deployment und Benchmarking realisiert.

Zur Etablierung einer **Nationalen Forschungsdateninfrastruktur (NFDI)** ist das SCC in den Fachkonsortien NFDI4Ing (Ingenieurwissenschaften), NFDI-MatWerk (Materialwissenschaft), PUNCH4NFDI (Teilchenphysik) sowie NFDI4Chem (Chemie) beteiligt und un-

terstützt NFDI4Earth (Erdsystemforschung). Thematisch im Fokus stehen die Themen Metadaten, AAI sowie Anbindung von Speicherdiensten.

In Europa engagiert sich das SCC beim Aufbau der **European Open Science Cloud (EOSC)** u. a. zur Etablierung der notwendigen Basisdienste. Jüngst wurden fünf neue EU-Projekte gefördert, u. a. zu Künstlicher Intelligenz (AI4EOSC), digitalen Zwillingen (interTwin) und Weiterbildung (Skills4EOSC).

Zahlreiche Vorträge und Veröffentlichungen der wissenschaftlichen Mitarbeitenden und Promovierenden ergänzen die Erfolge. Das Paper „Evolutionary Approach of Clustering to Optimize Hydrological Simulations“ wurde auf der International Conference on Computational Science and Its Applications (ICCSA) 2020 mit einem Best Paper Award ausgezeichnet. Für sein kontinuierliches Engagement im Helmholtz Inkubator „Information & Data Science“ wurde Achim Streit Mitte 2021 mit der Ehrennadel durch den Helmholtz Präsidenten ausgezeichnet.

// Ausgewählte Publikationen

D. Coquelin, C. Debus, M. Götz, F. von der Lehr, J. Kahn, M. Siggel, A. Streit: Accelerating neural network training with distributed asynchronous and selective optimization (DASO). *Journal of Big Data*, 9 (1), 14, 2022. DOI:10.1186/s40537-021-00556-1

M. Soysal, A. Streit: Collection of Job Scheduling Prediction Methods. *Proceedings of Job Scheduling Strategies for Parallel Processing (JSSPP 2021)*, Springer LNCS 12985, 2021. DOI:10.1007/978-3-030-88224-2_2

A. Streit, J. van Wezel: Deutschland in der European Open Science Cloud. *Praxishandbuch Forschungsdatenmanagement*. De Gruyter Saur, 2021. DOI:10.1515/9783110657807-003

// Mitarbeiterinnen und Mitarbeiter

Verwaltungspersonal

Anja Müller

Wissenschaftliches Personal

Dr. Rosella Aversa, Elnaz Azmi, Dr. Lisana Berberi, Dr. Isabelle Bierenbaum, Nicolas Blumenröhr, Dr. Ugur Çayoglu, Sabine Chelbi, Daniel Coquelin, Dr. Charlotte Debus, Borja Esteban Sanchis, Felix Ernst, Dr. Max Fischer, Anis Farshian Abbasi, Katharina Flügel, Laura Frank, Dr. Markus Götz, Germaine Götzelmann, Dr. Diana Gudu, Juan Pedro Gutiérrez Hermosillo Muriedas, Thomas Jejkal, Dr. Marcus Hardt, Volker Hartmann, Julian Herold, Maximilian Inckmann, Dr. Vandana Jha, Dr. Reetu Elza Joseph, Dr. Jörg Meyer, Dr. James Kahn, Dr. Valentin Kozlov, Peter Krauß, Dr. Eileen Kühn, Philipp Ost, Augustine Pane, Andreas Petzold, Andreas Pfeil, Dr. Nico Schlitter, Dr. Matthias Schnepf, Yusra Shakeel, Dr. Mehmet Soysal, Dr. Rainer Stotzka, Dr. Marcus Strobl, Melvin Strobl, Dr. Doris Ressmann, Oskar Taubert, Dr. Danah Tonne, Philipp Tögel, Dr. Pavel Weber, Dr. Marie Weil, Christoph Wendenius, Arvid Weyrauch, Jos van Wezel, Gabriel Zachmann

// Website

www.scc.kit.edu/personen/achim.streit.php