The major event at the Zuse Institute (ZIB) in 2014 was ZIB's scientific eval-
uation on November 23–24 by an in-
ternational review board (headed by Professor Hans-Joachim Bungartz of TU Munich). This assessment, re-
quested by the Berlin ministries with which ZIB is associated, the Senate 
Department for Economics, Tech-
nology and Research and the Senate 
Department for Education, Youth and 
Science, was executed along the evalu-
ation guidelines for the institutes of 
the Leibniz Association. Based on a 
development plan for its future activ-
ities (ZiBzukunftskonzept), devised in 
close collaboration with ZIB's Scien-
tific Advisory Board, and approved by 
ZIB's Administrative Council on May 
23, 2014, the ZIB members spent the 
summer and fall preparing a compre-
prehensive report of ZIB's research and 
services, the latter in particular via the op-
eration report from January 16, 2015: 
"ZIB conducts very successful fund-
amental research and offers highly de-
manded consultancy and services in ap-
plication-driven mathematics and com-
puting. Its three divisions … contribute 
significantly to the regional, national, 
and international research landscape, in-
cluding the role of the supercomputer 
consortium HLRS, and (3) the appro-
priateness of ZIB's present internal or-
ganization and of its institutional frame. 
Here are a few sentences from the evalu-
atation report from January 16, 2015: 
"ZIB conducts very successful funda-
mental research and offers highly de-
manded consultancy and services in ap-
plication-driven mathematics and com-
puting. Its three divisions … contribute 
significantly to the regional, national, 
and international research landscape via both 
high level research and high quality ser-
vice, the latter in particular via the op-
eration of facilities of the North German 
Supercomputer Consortium (HLRS) … 
If the budget situation stays stable, ZIB seems very well prepared for the future. The 
present and the intended future position-
ing of ZIB are considered to be excellent … 
Dealing with the complete vertical stack 
from basic research to industry applica-
tions, keeping the balance between service provision and research, as well as plac-
ing algorithmic research directly around a 
tier-3 HPC facility (including rich software 
activities) altogether form a unique selling 
point of ZIB – at a local, regional, national, 
and international scale. All future activities 
should build upon and further strengthen 
this unique selling point … Overall, it shall 
be concluded that the review panel received 
an excellent impression of ZIB."
CONTENTS

28-39
DESIGN OF NANOPHOTONIC DEVICES
Photonics is a key enabling technology

52-59
A CURIOUS LOOK OVER THE EDGE
From Open Access to Open Science

08-27
ZUSE INSTITUTE BERLIN
Preface | Executive Summary | Organization | ZIB Structure | Zib in Numbers | 30 Years of ZIB Reviewed | Data Science | Research Service | Economic Situation | Spin-Offs | Number of Employees

40-51
IMAGE ANALYSIS FOR BIOLOGY
Solutions for advanced analysis questions

60-71
RESEARCH CAMPUS MODAL
Science and industry under one roof

72-83
FAIRNESS AND LOAD IN DISTRIBUTED FILE SYSTEMS
Utilizing shared heterogeneous storage resources efficiently

98-114
ZIB PUBLICATIONS REFERENCES IMPRINT

84-97
KEEPING MANY CORES BUSY
Getting the best out of tomorrow’s HPC processors
Mathematical Optimization and Scientific Information

The Research Campus MODAL has become fully operational. Its optimization labs support the long-distance passenger operations of Deutsche Bahn, and the gas transport of Open Grid Europe (which owns Germany’s largest gas pipeline network) and are developing the open-source solver SCIP for general discrete-continuous optimization problems. Flight-trajectory optimization, sustainable manufacturing, telecommunication-network and supply-chain management, and many more projects in infrastructure and logistics are in the focus of the department Mathematical Optimization as well.

The department Scientific Information plans and develops information systems and services for the sciences. It engages in a variety of projects such as digitization of cultural heritage, a software-information portal, library services, and digital libraries that support open access to scientific literature and data and open science in general.

Executive Summary

In accordance with ZIB’s development plan (Zukunftskonzept), the divisions of ZIB were restructured. The former division Numerical Mathematics changed its name to Mathematics for Life and Materials Sciences in order to emphasize the shift in priorities toward an integrated approach encompassing the whole spectrum of competences in Data Analysis, Visualization, and Modelling, Simulation, and Optimization (MSO) on hand at ZIB.

The roadmap of the division is to complement its core strengths by advancement of new groundbreaking concepts, such as uncertainty quantification and visualization or sparse approximation in high dimensions, while advancing our application projects toward more complex processes.

These changes are supported by the next generation of large-scale projects that were acquired in 2014 and for which the following four may serve as examples: the new Einstein Center for Mathematics has started with six projects in our division; we host the MedLab of the Research Campus MODAL; we are participating in the DFG-CRC 1114 Scaling Cascades in Complex Systems, and we are part of the new Berlin Big Data Center (BBDC) in close cooperation with ZIB’s division Parallel and Distributed Computing.

Parallel and Distributed Computing

A major system upgrade pushed the HLRS-III supercomputer at ZIB to rank 51 in the TOP-500 list of the most powerful supercomputers in the world. A total of 1,872 computing nodes with 44,928 cores and 117 terabytes of main memory are now available at ZIB. The Cray XC30/XC40 is supplemented by a test-and-development system with Intel many-core processors and nonvolatile memory. It helps to intensify our research on scalable software for exascale computers. We implemented the active message library HAM (see our feature article “Keeping Many Cores Busy”), which reduces the overhead of code off-loading on heterogeneous supercomputers. In the domain of high-performance data analysis, we participated in two new BMBF-funded research projects, the Berlin Big Data Center (BBDC) and the GeoMultiSens project. Both focus on the analysis of large-scale data in business and science. Their different access patterns require advanced mechanisms for fairness and load on distributed file systems as described in our corresponding feature article.

Life and Materials Sciences

In accordance with ZIB’s development plan (Zukunftskonzept), the divisions of ZIB were restructured. The former division Numerical Mathematics changed its name to Mathematics for Life and Materials Sciences in order to emphasize the shift in priorities toward an integrated approach encompassing the whole spectrum of competences in Data Analysis, Visualization, and Modelling, Simulation, and Optimization (MSO) on hand at ZIB.

The roadmap of the division is to complement its core strengths by advancement of new groundbreaking concepts, such as uncertainty quantification and visualization or sparse approximation in high dimensions, while advancing our application projects toward more complex processes.

These changes are supported by the next generation of large-scale projects that were acquired in 2014 and for which the following four may serve as examples: the new Einstein Center for Mathematics has started with six projects in our division; we host the MedLab of the Research Campus MODAL; we are participating in the DFG-CRC 1114 Scaling Cascades in Complex Systems, and we are part of the new Berlin Big Data Center (BBDC) in close cooperation with ZIB’s division Parallel and Distributed Computing.
SCIENTIFIC ADVISORY BOARD

The Scientific Advisory Board advises ZIB on scientific and technical issues, supports ZIB’s work, and facilitates ZIB’s cooperation and partnership with universities, research institutions, and industry.

The Board of Directors appointed the following members to the Scientific Advisory Board:

- PROF. DR. JÖRG-RÜDIGER SACK
  Carleton University, Ottawa, Canada
- PROF. DR. ALFRED K. LOUIS
  Universität des Saarlandes, Saarbrücken
- PROF. DR. RAINER E. BURKARD
  Technische Universität Graz, Austria
- PROF. DR. MICHAEL DELLNITZ
  Universität Paderborn
- DIPL.-MATH. LUDGER D. SAX
  Grid Optimization Europe GmbH
- DR. ANNA SCHREIECK
  BASF SE, Ludwigshafen
- DR. HORST D. SIMON
  NERSC, Lawrence Berkeley National Laboratory, USA
- DR. REINHARD UPPENKAMP
  Berlin Chemie AG, Berlin
- DR. KERSTIN WAAS
  Deutsche Bahn AG, Frankfurt am Main

The Scientific Advisory Board met on July 7 and 8, 2014, at ZIB.
ZIB is structured into four divisions, three scientific divisions, and ZIB’s administration. Each of the scientific divisions is composed of two departments that are further subdivided into research groups. For further information on the broad spectrum of thematic and application focuses of the research groups please visit www.zib.de.

ZIB Structure

LEGEND

- SCIENTIFIC DIVISIONS AND DEPARTMENTS
- RESEARCH GROUPS
- RESEARCH SERVICE GROUPS
ON NOVEMBER 22, 2014:
ZIB SUPERCOMPUTER IS NO. 51 IN TOP-500 LIST

PEER-REVIEWED PUBLICATIONS IN INTERNATIONAL SCIENTIFIC JOURNALS
111

PROMOTION OF YOUNG SCIENTISTS:
DISSERTATIONS
DIPLOMA
MASTER

VISITORS
LONG NIGHT OF THE SCIENCES
1,283

ZIB IN NUMBERS

SEMINARS GIVEN BY ZIB SCIENTISTS AT UNIVERSITIES
13

LECTURES GIVEN BY ZIB SCIENTISTS AT UNIVERSITIES
10

SCIP
DOWNLOADS OF MIP SOLVER SCIP
11,852

59 OUTREACH EVENTS FOR SCHOOL CLASSES AND THE GENERAL PUBLIC

10 AWARDS

PROFESSORSHIPS OFFERED TO ZIB RESEARCHERS
4

SOFTWARE ACQUISITIONS OF AMIRA AND AVIZO INSTALLED WORLDWIDE IN 2014
6,000

SCIP
DOWNLOADS OF MIP SOLVER SCIP
11,852

INDUSTRY FUNDING INCREASED BY MORE THAN 55%!

€ 7,257,832
EUROPEAN RESEARCH AND DEVELOPMENT (INDUSTRY)

€ 2,919,279

ZIB ACQUISITION OF THIRD-PARTY FUNDS

€ 4,338,553
PUBLIC PROJECT REVENUES

6,200 INTERNATIONAL GUESTS AT ZIB IN 2013

115 CONFERENCES AND WORKSHOPS AT ZIB

197 SCIENTIFIC TALKS
37 DISTINGUISHED
160 INVITED

6,000 LICENSES OF AMIRA AND AVIZO INSTALLED WORLDWIDE IN 2014
The preface of this annual report contains a few quotes from the evaluation report of the review panel, and there is no need to repeat the excellent overall rating of our research and service activities as well as our unique position on the regional, national, and international research landscape.

According to the Zukunftskonzept ZIB now presents itself as the Berlin Competence Center for Computing and Data Science which seems the logical next step in the continued development of ZIB's original focus in view of the major global research trends. The illustration shows that the organization of ZIB reflects this focus.

One of the main new elements of the Zukunftskonzept is the aggregation of research activities at ZIB into Research Focus Areas (RFA). The RFA were selected based on the following criteria: ZIB shall have attained or shall have the potential to attain high impact in each of these areas, they shall cluster activities at ZIB and shall offer high potential for fostering in-house cooperation, and they shall improve visibility and attractiveness of ZIB for potential collaboration partners. Moreover, the RFA may well overlap to foster cross-fertilization.

The present Research Focus Areas are as follows:

- Molecular and biological processes
- Solutions for individualized medicine
- Robust solution under uncertainty
- Materials and optical processes
- Combinatorics and MINLP
- Digital humanities
- Scalable and fault tolerant algorithms
- Modeling, simulation, and optimization for multiscale systems
- Large-scale data management, curation, and analysis
- Highly parallel computing

For further details on the RFA, please visit our Web site: www.zib.de
One of the cornerstones of ZIB’s research services has always been high-performance computing (HPC). Since its foundation in 1984, ZIB has operated high-performance computers that are among the fastest in Germany. Currently, ZIB operates a Cray XC40 supercomputer as one of the two HPC sites for the North German Supercomputer Alliance (HLRN). But the provision of supercomputer time is just one aspect of ZIB’s HPC activities; another important research service is the organization of courses, workshop, and consulting for the HLRN users. Consultants at ZIB provide valuable advice for HPC users throughout Germany in fields such as computational chemistry, fluid dynamics, structural mechanics, earth and environmental sciences, material sciences, bioinformatics, and the life sciences.

ZIB’s HPC activities have always been accompanied by data-archive services: ZIB operates large-scale data archives for in-house use and as a service for bitstream preservation to its partners in the Berlin area. Our high-speed and high-capacity data archive is protected against various environmental disasters such as water, fire, smoke, and theft. High standards in data security are necessary because the vast majority of the stored data is irreplaceable, such as raw measurement data from scientific experiments, long-term climate data, or digitized views of archaeological artifacts and documents. Consequently, we always store two copies of the data, periodically testing the readability by verifying checksums, rewriting the data when necessary, and migrating them to new storage technology when it becomes available.

RESEARCH SERVICES

However, ZIB’s research services are not limited to the more classic aspects of information technology. Rather, they heavily extend into the field of Scientific Information. For example, ZIB hosts the Cooperative Library Network Berlin-Brandenburg (KOBV), the common platform for all university libraries and public libraries and numerous research, special, and government libraries in Berlin and Brandenburg, and the Service Center for Digitization Berlin (dggB), a project to coordinate and support digitization efforts of cultural assets in the city-state of Berlin.

Research Services

ZIB provides high-quality research services in Computing, Data Science, and Scientific Information. These services are closely linked to our research activities to their mutual benefit.

HLRN-III complex at ZIB
The extraction of knowledge from data has always been of key relevance for science, economics, and society. In the age of the digital revolution, this process is gaining importance with dramatic consequences in science: with the increasing use of high-volume data from, for example, next-generation sequencers in the biosciences, medical imaging in health care, sensor data in geo- and environmental sciences, or text, audio, and video streams in social networks, Data Science became a newly emerging field.

It encompasses an interdisciplinary spectrum of aspects ranging from large-scale data management, via data curation and quality control to data analysis and visualization. The challenges of handling big data are often described by the four Vs, volume, variety, velocity, and veracity: very large data volumes are streamed with high velocity from various heterogeneous and potentially unreliable data sources and need to be analyzed for scientific results and economic benefit.

ZIB is active in various aspects of Data Science: the distributed management of structured and unstructured high-volume data sets, long-term bitstream preservation, highly parallel, deep, and sparse data analytics, data curation and fusion, and scientific visualization and visual data analysis. With ZIB’s development plan (Zukunftskonzept), we have started to reorganize these activities into one of the main cross-divisional research pillars of ZIB.

One first important step in this direction was the acquisition of the Berlin Big Data Center (BBDC). The BBDC is one of only two competence centers for big data in Germany. It was co-founded in 2014 by TU Berlin (coordinator), ZIB, DFKI, Beuth Hochschule, and the FZI Research Institute. ZIB participates in BBDC through concerted efforts of its divisions Parallel and Distributed Computing and Mathematics in Life and Materials Sciences. They contribute a scalable, declarative, open-source system that enables the design, optimization, parallelization, and fault-tolerant execution of advanced data analytics pipelines, and they are responsible for one of the key application pilot projects at the BBDC – the automated analysis of large-scale medical data. Fast and reliable access to distributed data is a necessary precondition for Data Science.

One of the major open-source software packages the distributed cloud file system XtreemFS, the fast in-memory database BabuDB, and the transactional key-value store Scalaris. All three software packages are internationally recognized and are being adopted by a growing developer community. The focus of these developments was the design and proof of concept for novel protocols on scalability and concurrency control in distributed data management.

Long-term data storage is another field of expertise at ZIB. We work with several institutions in the Berlin area on the secure storage of large data volumes. Advanced techniques, like predictive methods for anticipated bit errors on tapes, have been developed jointly with our industry partners and implemented on the SL8500 data robots for more reliable long-term data storage. These activities are complemented by research in data analysis that combine method, algorithm, and software development in visualization, machine learning, uncertainty quantification and visualization, data sparsity as well as model-based approaches to the extraction of knowledge from data such as data assimilation or parameter estimation.
ECONOMIC SITUATION IN 2014

ZIB THIRD-PARTY FUNDS BY SOURCE

- 8% EU
- 16% BMBF/BMWi
- 17% DFG (incl. MATHEON)
- 19% OTHER PUBLIC FUNDS
- 40% INDUSTRY

€ 2,919,279
INDUSTRY

€ 1,269,374
DFG (incl. MATHEON)

€ 1,166,042
BMBF/BMWi

€ 544,944
EU

€ 1,358,193
OTHER PUBLIC FUNDS

ECONOMIC SITUATION IN 2014
The Zuse Institute Berlin (ZIB) finances its scientific work via three main sources: the basic financial stock of the Federal State of Berlin, and third-party funds of public sponsors, and industrial cooperation contracts.

In 2014, ZIB raised third-party funding through a large number of projects. Unfortunately, project-related public third-party funds declined from €4,999 in 2013 to €4,339 in 2014, but industrial third-party projects more than compensated the losses, and funding rose significantly compared to last year from €1,803 to €2,919. Altogether, €7,258 in third-party funding marked a new record in ZIB’s history.

Despite the economically difficult times, ZIB was once again able to acquire interesting research projects and remains a very attractive cooperation partner for medium-sized, innovative companies. In particular, the Research Campus MODAL strongly supports the cooperation with industry, as well as with small and medium-sized enterprises. ZIB’s research-and-development collaborations were able to gain new, long-term prospects.

Based on these success stories, ZIB strengthened its position as an attractive place for innovative mathematicians and computer scientists such that the number of employees at ZIB stayed almost constant despite the fact that many highly qualified scientists were offered careers in academia or industrial research elsewhere.

The new HLRN-III high-performance computer at ZIB started operating in September 2013 and performed well in 2014. Again, the institute’s long-term data storage was further expanded. As a consequence, a large number of new research collaborations with Berlin-based, national, and international institutions in this area could be initiated. With these developments, ZIB strengthened its position as a competence center for computing and Data Science.
SPIN-OFFS

COMPUTING IN TECHNOLOGY GMBH (CIT) 1992 | www.cit-wulkow.de
Mathematical modeling and development of numerical software for technical chemistry

RISK-CONSULTING PROF. DR. WIEBER GMBH 1994 | www.risk-consulting.de
Database marketing for insurance companies

INTRANETZ GMBH 1996 | www.intranetz.de
Software developments for logistics, database publishing and e-Government

AKTUARDATA GMBH 1998 | www.aktuardata.de
Development and distribution of risk-evaluation systems in health insurance

Visualization and data analysis (Amira etc.), especially medical visualization

ATESIO GMBH 2000 | www.atesio.de
Development of software and consulting for planning, configuration, and optimization of telecommunication networks

BIT-SIDE GMBH 2000 | www.bit-side.com
Telecommunication applications and visualization

DRES. LÖBEL, BORNDÖRFER & WEIDER GBR 2000 | www.lbw-berlin.de
Optimization and consulting in public transport

LENNÉ3D GMBH 2005 | www.lenne3d.com
3-D landscape visualization, software development, and services

JCMWAVE GMBH 2006 | www.jcmwave.com
Simulation software for optical components

ONSACLE SOLUTIONS GMBH 2006 | www.onscale.de
Software development, consulting, and services for parallel and distributed storage and computing systems

XTREEM STORAGE SYSTEMS GMBH 2013 | www.quobyte.com
Xtreem Storage Systems develops software-only storage systems that run as software on COTS hardware; it offers carrier-grade storage software for enterprises that runs on off-the-shelf hardware; Xtreem Storage Systems GmbH was founded in 2013

LAUBWERK GMBH 2008 | www.laubwerk.com
Construction of digital plant models

1000SHAPES GMBH 2010 | www.1000shapes.com
Statistical shape analysis

TASK - KURT-HERTHOLD GLEIXNER HEINZ KOCH GBR 2010
Distribution, services, and consulting for ZIB’s optimization suite

QUOBYTE INC. 2013 | www.quobyte.com
Quobyte develops carrier-grade storage software that runs on off-the-shelf hardware

ATESIO GMBH 2000 | www.atesio.de
Development of software and consulting for planning, configuration, and optimization of telecommunication networks

BIT-SIDE GMBH 2000 | www.bit-side.com
Telecommunication applications and visualization

DRES. LÖBEL, BORNDÖRFER & WEIDER GBR 2000 | www.lbw-berlin.de
Optimization and consulting in public transport

LENNÉ3D GMBH 2005 | www.lenne3d.com
3-D landscape visualization, software development, and services

JCMWAVE GMBH 2006 | www.jcmwave.com
Simulation software for optical components

ONSACLE SOLUTIONS GMBH 2006 | www.onscale.de
Software development, consulting, and services for parallel and distributed storage and computing systems

XTREEM STORAGE SYSTEMS GMBH 2013 | www.quobyte.com
Xtreem Storage Systems develops software-only storage systems that run as software on COTS hardware; it offers carrier-grade storage software for enterprises that runs on off-the-shelf hardware; Xtreem Storage Systems GmbH was founded in 2013

LAUBWERK GMBH 2008 | www.laubwerk.com
Construction of digital plant models

1000SHAPES GMBH 2010 | www.1000shapes.com
Statistical shape analysis

TASK - KURT-HERTHOLD GLEIXNER HEINZ KOCH GBR 2010
Distribution, services, and consulting for ZIB’s optimization suite

QUOBYTE INC. 2013 | www.quobyte.com
Quobyte develops carrier-grade storage software that runs on off-the-shelf hardware

In the year 2014, 208 people were employed at the ZIB, of these, 145 positions were financed by third-party funds.

NUMBER OF EMPLOYEES

<table>
<thead>
<tr>
<th>SPIN-OFFS</th>
<th>MANAGEMENT</th>
<th>SCIENTISTS</th>
<th>SERVICE PERSONNEL</th>
<th>KOBV HEADQUARTERS</th>
<th>STUDENTS</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2014</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1.1.2015</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>18</td>
<td>85</td>
<td>103</td>
</tr>
<tr>
<td>18</td>
<td>81</td>
<td>99</td>
<td>34</td>
<td>14</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>49</td>
<td>49</td>
<td>0</td>
<td>38</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>149</td>
<td>213</td>
<td>63</td>
<td>145</td>
<td>208</td>
<td></td>
</tr>
</tbody>
</table>
DESIGN OF NANOPHOTONIC DEVICES

Photonics is a key enabling technology

The International Year of Light, 2015, established under the auspices of the United Nations, aims to strengthen the global awareness of how light-based technologies enforce sustainable development and provide solutions to global challenges in energy, information technologies, education, agriculture, and health. ZIB contributes in particular to nanophotonics research in areas such as solar-energy harvesting and advanced quantum devices, which are key to progress with respect to the global challenges.
A path with great promise on the road to a clean and sustainable world energy supply is the use of solar energy. Our research follows two paths.

The first one is photovoltaic (PV) energy conversion; the second and new one is artificial photosynthesis. The technology roadmap of the International Energy Agency predicts an annual market of 105 GW for photovoltaic systems in 2030 and a total installed capacity of 900 GW that contributes 5% to the total electricity generation. In view of a sustainable energy supply, these technologies should also not be limited by the abundance of the required materials. This research requires the close collaboration of partners from different fields, and the highlight of 2014’s activities was the opening of the cross-institutional Berlin Joint Lab for Optical Simulations for Energy Research (BerOSE).

BERLIN JOINT LAB FOR OPTICAL SIMULATION FOR ENERGY RESEARCH (BEROSE)

The lab assists in the research and development of materials and devices for photocatalysis, photovoltaics, and photonics with structure sizes in the nanometer range. The main focus will be the 3-D simulation of both single structures and entire functional systems. The partners of the joint lab are the Helmholtz Center Berlin, the Freie Universität Berlin, and the Zuse Institute, whose specific experience combines experimental and theoretical work and scientific computing. The joint lab is based on a long-term and very successful collaboration of the partners in the field of thin-film photovoltaics. But the lab includes also new research fields: Nanoplasmonics in solar-cells, artificial photosynthesis, and nanophotonic concepts for sensors and functional optical components.

© Andreas Kubatzki, Helmholtz-Zentrum Berlin
THIN-FILM SOLAR CELLS: IMPROVING EFFICIENCY WITH LIGHT MANAGEMENT

Silicon thin-film technology has the potential to meet all economic and ecological requirements if the efficiency of solar cells incorporating silicon thin films (today about 12%) can be brought into the range of silicon wafer devices (20–25%). Recent years saw a dramatic breakdown in the industrial production of thin-film cells due to the cheap availability of thicker wafer-based modules. Apart from current industrial considerations, the optimization of thin-film techniques, which are gentle on resources, remains a scientific goal as the gained insight is not limited to solar-cell applications, as it can possibly also be applied, for example, in sensing applications.

Thin-film devices suffer from insufficient absorption of sunlight in the semiconductor layer. A promising way of increasing the absorption and thus the efficiency of thin-film solar cells is the use of light-management structures. These structures act as a scattering layer that deflects more of the incident solar irradiation into the absorber layer than a planar material interface. Through total internal reflection, light is kept for as long as possible inside the active semiconductor region of the solar cell, thus enhancing the conversion of photons into electrons. This process is also referred to as “light trapping.” Figure 2 shows snapshots from different technological approaches to improve light trapping. All of these different concepts have been investigated and optimized with respect to their light-trapping properties through means of optical simulation.

Another commonly investigated approach is a periodic arrangement of scattering structures on the metallic back reflector of a solar cell, such as the plasmonic silver nanopyramids shown in Figure 2b. For this simulation, the deposition process of silicon is modeled as well and results in a smoother texture. Novel texturing efforts such as nanoimprint lithography can be used to cheaply texture large areas of high-quality silicon. The resulting periodic structures are highly efficient in trapping light and exciting resonances (Figure 2c).

ARTIFICIAL PHOTOSYNTHESIS

BerOSe offers new and exciting research topic opportunities in addition to solar-cell light management. Photovoltaic energy conversion relies on sunlight, and the produced electricity cannot be efficiently stored in large quantities today. The most efficient way to store large amounts of energy is in chemical fuels such as gasoline or hydrogen. A novel idea is thus to store solar energy within solar fuels produced through artificial photosynthesis. The goal of artificial photosynthesis is to mimic the green plants in using sunlight to make high-energy chemical solar fuels like hydrogen or ethanol. In close collaboration with our partners Prof. van de Krol and Dr. Abdi from the Institute of Solar Fuels at the Helmholtz Center Berlin, we investigate monolithically integrated water-splitting devices. We intend to improve optical absorption within the metal-oxide semiconductor layer with plasmonic effects that also might enhance the catalytic conversion of water to hydrogen and oxygen.
In addition to solar energy, another focus of our research activities lies in the development of advanced tools for the analysis and design of nano-photonic devices.

**Vertical-Cavity Surface-Emitting Lasers (VCSELS)**

VCSELS are key components for optical interconnects and are widely applied in high-performance computers and data centers. Single transverse-mode VCSELS are used in sensing, illumination, and display applications. Within this collaboration project, we study designs aiming at large numerical apertures. An example of an excited whispering gallery single-transverse TEM$_{0,49}$ mode of the cylinder symmetric structure is shown in Figure 2. The field intensity is concentrated along a small circular ring and vanishes in the center. The designed lasers have been fabricated. The 850 nm GaAlAs-based VCSELS demonstrated single-mode lasing and a confirmed operation speed up to 25 Gb/s [3].

**Optical Metrology: FinFETs and VNANDs**

Optical metrology makes use of the interaction of light with nanostructured objects to measure unknown quantities—most importantly, geometric quantities. Reliable measurements are an essential precondition for any future lithographic manufacturing processes and need to be reliable down to a subnanometer level. In the semiconductor industry, it is used in process control and in mask quality control. Numerical modeling is an important part of optical metrology setups in this field; measurement results are compared to simulation results of a parameterized model in order to quantitatively determine dimensions of the measured sample. With increasing complexity and decreasing feature sizes, the need for accurate optical metrology methods for complex 3-D shapes is increasing [2].
QUANTUM DOTS WITH LENSES

Due to the increasing use of telecommunications in all areas of society, be it political, industrial, or private, potential threats by unwarranted access to data by third parties has become a major source of concern for users as well as providers of communication channels. Great efforts have been made in guarding communication against attacks and eavesdropping; however, 100% security cannot be guaranteed in classic transmission systems as the risk of wiretapping or other attacks is always present. Quantum communication, on the other hand, has the potential of achieving that long-sought-after goal of absolute security, as at its core it relies on the quantum nature of single particles, which may not be copied or altered without being noticed by the communication partners.

To enable quantum communication, novel, nonclassic light sources are needed which are capable of reliably generating single quanta of light, a.k.a. photons. In contrast to classic light sources, which are designed to emit as many quanta as possible, light sources suitable for quantum communication may only generate one photon at a time to ensure that only one copy of the information the photon carries is in existence.

Our group in collaboration with the Institute for Solid-State Physics of TU Berlin is involved in the optimization of such nonclassic light sources based on quantum dots in GaAs semiconductors. Simulating the optical field in a quantum-dot-based single-photon emitter is a fast, inexpensive, and reliable way of optimizing such devices and gives rise to higher efficiency, better thermal properties, a smaller footprint, and many more technological and economic properties [4].

OPTICAL CHIRALITY

An object is called chiral if it cannot be identified with its mirror image. Chiral objects react differently to left or right circular polarized light. This is called optical activity and is used in many applications. A chiral structure, which can be dispersed in a liquid to investigate interactions with large chiral molecules, can, for example, be achieved by etching a surface that is not symmetric in the sense that it has a crystalline structure with nonstandard orientation. The etching causes chiral cavities in the surface from which negative forms can be extracted and dispersed (see Figure 4). Recently, there has been an interest in studying this kind of artificially manufactured chiral nano-objects to study qualitatively new physical effects on molecules for biological or pharmaceutical applications [5].
ADVANCES IN ADAPTIVE FINITE ELEMENT METHODS

The focus of this algorithmic-oriented research is the generalization and performance improvement of the adaptive finite-elements framework for the solution of Maxwell’s equations. In recent years, we have added a number of new key features.

UNBOUNDED DOMAINS

A key to a rigorous and efficient solution for optical scattering problems is a suitable treatment of wave propagation on unbounded, possibly heterogeneous, exterior domains. During the entire period of our research, we developed and continuously improved a new approach called Pole Condition, and extended a given approach, the Perfectly Matched Layer technique, to cope efficiently with such problem types. Besides Maxwell equations, we extended the concepts to the Schrödinger equation, Klein-Gordon equation, and drift-diffusion equation.

REDUCED BASIS METHODS

Based on the finite element package, adaptive, error-controlled reduced basis methods for solving parameterized 3-D optical scattering problems have been developed and applied in manifold multi-query and real-time contexts. The main application fields are the optimization of solar-cell structures as well as inverse problems for parameter reconstructions as they occur, for example, in optical metrology.

HIERARCHICAL HIGH-ORDER FINITE ELEMENTS

We completed our software package with hierarchical high-order finite elements for different types of mesh cells (tetrahedrons, hexahedrons, prisms, and pyramids) that can be combined in hybrid meshes. Such hybrid meshes with adaptively chosen mesh size and order are superior to traditional meshes of uniform type and uniform in size and order.

HYDRODYNAMIC MODELS IN PLASMONICS

The light-matter interaction, even in classical or semiclassical formulations, gives rise to new physical effects when the structure sizes approach the subwavelength scale. In the range of a very few nanometers, the classical local Drude model has to be replaced by models closer to quantum mechanics. We developed a numerically consistent nonlocal hydrodynamic model that was able to predict the right resonances in nanoparticles.

PERIODIC UNBOUNDED DOMAINS

A very recent result involving unbounded domains is the optimization of organic light-emitting diodes (OLED) where we developed novel numerical techniques allowing fast computation of the light out-coupling efficiency. The special situation is that the unbounded simulation domain consists of periodic or quasiperiodic scatterers. To account for this horizontal extension of the OLED, we applied a recently proposed technique based on a Floquet transform. This allows for a restriction of the calculations to a unit cell of the periodic structure and results in an enormous gain in the numerical efficiency.
IMAGE ANALYSIS FOR BIOLOGY

Solutions for advanced analysis questions

Microscopy techniques are becoming an increasingly important tool for research in biology. They are key to analyzing biological systems on various scales, from the structure of biomolecules, using electron microscopy, up to whole organs, using optical microscopy. ZIB helps to create an understanding of biological systems through image-analysis solutions. We develop analysis methods and software that biologists apply to answer advanced analysis questions.
Microtubules are tube-like polymers that play an important role in cell division. They form the spindle - a microtubule assembly that segregates the chromosomes. One important method for investigating microtubules in cells is electron tomography: Thick (300 nm) serial sections of cells are prepared for electron microscopy and imaged on a CCD camera. With the application of image-processing techniques, we are able to automatically extract the centerlines of microtubules within a section [7][8][9]. Furthermore, we have been developing reconstruction methods for a quantitative analysis of full mitotic spindles. These methods allow joining electron tomograms and extracted microtubule centerlines across physical sections to create a geometric model of the entire spindle [10]. The development started in collaboration with the MPI-CBG Dresden and is now being used in several biology laboratories worldwide. With the achieved level of automation, the analysis of microtubule organization for large parts of the cell from serial-section electron tomography now seems feasible for the first time. The results will be visually and quantitatively analyzed.
Neuroscientists would like to understand how the brain processes sensory information, and how this ultimately leads to certain behavior. To investigate this, our collaborators (Oberlaender Group, MPI for Biological Cybernetics, Tübingen) used the rat somatosensory (barrel) cortex, which processes information from the whiskers, as a model system. The brain is comprised of a large number of interconnected neurons, which have a complex tree-shaped form, allowing them to communicate with many other nerve cells over potentially long distances. Together with our project partners, we developed methods for the detailed 3-D reconstruction of the individual neuron’s shape from microscopic images, including its complex and long-ranging axonal arbor [11]. Based on such carefully reconstructed anatomical data, we created a 3-D digital model representing the rat barrel cortex [12] consisting of approximately 0.5 million neurons and their synaptic connections. The model is able to reproduce important network properties that have been reported in the literature (for example convergence) and could therefore be used to predict other network properties that are difficult to access through laboratory experiments. This model provides insight into the neural circuits that process touch information and can, for example, serve as the anatomical basis for simulating the flow of information through parts of the brain.

Modeling synaptic connectivity in anatomically realistic neural networks

Cell-type-specific synaptic connectivity in the rat barrel cortex. A 3-D model of the rat barrel cortex was created by placing a measured number of neuron reconstructions (left) in a reference frame defined by the pia (top) and white-matter (bottom) surfaces and 24 cortical columns (illustrated by the transparent cylinder and the contours at the top). From this dense model, we can estimate synaptic connectivity between individual neurons or at the level of cell types. Shown are the complex dendritic (red) and axonal (black) trees of a single-layer 5 thick-tufted cell. Through their long and complex axonal branches, neurons of this type make many synaptic contacts in the surrounding cortical columns, mainly at the top of the barrel cortex (orange).
REGISTRATION OF 2-D IMAGE DATA INTO 3-D VOLUMETRIC DATA

The first project, a collaboration with Wolfgang Wagermaier and Rebecca Hoerth is aimed at understanding bone healing after fracturing. For different fracture gaps of rat bones, the healing process was studied. Of particular interest here was the orientation of the mineral platelets in the fractured regions during the healing, which was studied using SAXS imaging, a 2-D imaging technique with very high resolution. In order to better understand the results of this imaging technique, the 2-D SAXS images were integrated into the 3-D volume of the bone obtained using micro-CT. We developed a multistep semiautomatic registration method using 2-D EM images as a methodic link [14]. In a first step, the SAXS images were registered to the 2-D EM image, which was then registered to the 3-D micro-CT image. We developed an approach using a combination of the generalized Hough transform and image-based optimization to register the 2-D EM images into the 3-D volumes with only minor user interaction. We could register all data sets within less than one hour while spending only 10-15 minutes of manual labor per slice [14].

UNDERSTANDING BIOMATERIALS

Many biological materials are composites with a hierarchical structure [13]. In order to understand the hierarchical nature of materials, they are investigated on different length scales using a variety of imaging techniques, including medical CT, micro-CT, electron microscopy (EM), nanoindentation (NI), and small-angle X-ray scattering (SAXS). At ZIB, we develop automated techniques for image registration, segmentation, and analysis of data from such imaging techniques. In collaboration with the MPI of Colloids and Interfaces in Potsdam-Golm, we have shown in two projects how automated image-analysis techniques help in understanding the structure of biological materials.
ANALYSIS OF SKELETAL ELEMENTS OF CARTILAGINOUS FISH

In a second project, we have been developing methods for image segmentation and shape analysis in collaboration with Mason Dean and Ronald Seidel. The project focus is the skeletons of cartilaginous fishes (sharks and rays), which are covered by a mineralized layer composed of a network of tiles, called tesserae. This mineralized layer is known to stabilize the skeletal elements. But many questions concerning the properties, the distribution, and the growth of these tesserae remain unanswered [15], in part because no methods were available for quantifying the shapes and arrangements of the massive number of tesserae on each skeletal element. The methods we have been developing allow us to investigate the anatomy of the mineralized layer at various levels of structural hierarchy, working from micro-CT and synchrotron-CT scans of tessellated cartilage. In his master’s thesis [16], David Knoetel developed an automated image-analysis pipeline to segment tesserae at different levels of detail and then reduce the complex network to a series of data visualizations to facilitate analysis. On a different length scale, bachelor student Merlind Schotte works on the automatic extraction of shape parameters from entire skeletal elements based on principal component analysis, enabling the comparison of shapes across different species of sharks and rays. These analyses of structural complexity on different size scales provide a unique window into the organization, growth, and mechanics of this interesting biological tissue composite and are invaluable input for bio-inspired modeling projects.

Amira is commercial software for visualization and data analysis with approximately 7,000 users worldwide. The initial version of Amira was developed at Zuse Institute Berlin (ZIB) in 1994 in response to the lack of an interactive, simple-to-use, and extensible software solution for non-computer scientists that wanted to apply visualization. Today, Amira is commercially developed and distributed by FEI Visualization Sciences Group. ZIB still contributes to the development of Amira through technology transfer, based on a long-term cooperation with FEI that funds the development of novel algorithms at ZIB. The transfer allows ZIB to distribute its methods to a large audience while benefiting from the maintenance and support of the company. Support for several operating systems, code maintenance, and bug fixes are provided over the long term, which helps to ensure that Amira is a reliable framework for biological image analysis.

In recent years, ZIB contributed several improvements for image analysis to Amira, such as tools for interactive image segmentation and for deriving quantitative information from reconstructed geometry. One highlight is a workflow for automatic and semiautomatic reconstruction of filamentous structures from electron tomograms. With this new workflow, Amira provides a unique solution that is used by laboratories worldwide for analyzing microtubule organization during cell division.

Amira is also used as the basis for many research projects at ZIB and is freely available to ZIB cooperation partners. ZIB offers selected image-analysis solutions for download on the Web site, as extensions to Amira to allow researchers to reproduce and apply methods that have been developed and published by ZIB. Sharing our solutions provides direct value to the research community and helps us to test and verify our algorithms. Examples are the automated segmentation of actin filament networks, tracing of microtubules, and the 3-D reconstruction of neurons.
From Open Access to Open Science

Today, 12 years after the Berlin Declaration, the characteristic adjective “open” can be found in the most diverse scientific and social discourses [25]. In science and research, the subject Open Access increasingly refers to research data and research processes. Mass digitization of cultural artifacts and the application of information technology not only technically open up new framework conditions for knowledge production; they also enable a view beyond the horizon and confront traditional academic discourse with changed epistemic aspects.
ON OPEN ACCESS, OPEN SCIENCE, OPEN DATA

For the Open Access Movement, 2014 was definitely a good year. Following decades in which this topic was raised by committed scientists (among them numerous mathematicians) again and again, it has now made its way into the center of society.

The so-called “Green Road” has become law for a certain area, since in January 2014 the self-archiving law (Zweitveröffentlichungsrecht) came into force [33]. It gives scientists the right to publish their own articles on the Internet after a period of one year. However, this is not an “indispensable” right; the focus is on results from publicly funded projects, not on publications from university research itself [34]. Critics also complain that especially in scientific disciplines an embargo of six months would have been advisable. Nevertheless, another step has been made.

The resolution of the Scientific Committee of the Berlin House of Representatives in May 2014 in favor of an “Open Access Strategy for Berlin: Making Scientific Publications Accessible and Usable for Everyone” was aimed at universities and scientific institutions in Berlin. They are requested to develop “their own Open-Access strategies” comprising publications, research data, and scientific software. Furthermore, the Senate has been asked to examine to what extent the Open Access concept may be embedded in the (future) agreements with Berlin universities [35].

Before the Open Access concept finally reached the legislator, it required a large number of scientific-political statements of intent, most notably the “Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities” published back in 2003. It has been signed by more than 500 universities, academies, foundations, and associations worldwide [36]. Already in its preamble, the Berlin Declaration explicitly includes cultural heritage in its definition of goals for Open Access: “The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access” [37]. The entire spectrum of potentially available material is explored: “Open-access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials, and scholarly multimedia material” [38].

One of the latest contributions of ZIB to the mindset of Open Access stems from the DFG-funded research project EWIG. Recently completed, the project looked at the various obstacles in the workflow from scientific research data into the long term archive or the permanent domain [39]. There are undoubtedly a few technical hindrances but the main problem seems to be the attitude towards and knowledge on the concept of Open Access to research data. To lower the barriers and pave the way for the idea of the open sharing of knowledge as early as possible, a brochure was published targeting junior scientists and teachers alike. It covers the main components of research data management and can serve as a component for developing a curriculum dealing with this topic as well as a reference guide providing orientation to young scientists [40].
FROM OPEN ACCESS TO OPEN SCIENCE

Today, 12 years after the Berlin Declaration, the characteristic adjective “open” can be found in the most diverse scientific and social discourses [25]. In science and research, the subject Open Access increasingly refers to research data and research processes. Mass digitization of cultural artifacts and the application of information technology not only technically open up new framework conditions for knowledge production, they also enable a view beyond the horizon and confront traditional academic discourse with changed epistemic aspects.

For instance, the “Digital Humanities” see themselves confronted with new questions in terms of theory and practice of cognition. “We are (…) in the midst of a transformation of the institutional and conceptual conditions of possibility for the generation, transmission, accessibility, and preservation of knowledge” [26]. The great challenge arises from the altered quality of the technological change. Not only large amounts of data, but also extremely heterogeneous data such as digital representations of cultural heritage objects have to be standardized, quality-assured and made available [27]. “Research data at your fingertips” [28] is formulated as a vision by the working group for research data of the German science organizations’ priority initiative “Digital Information”. Here, topics such as licensing, sustainable digitization, and long-term availability of data as well as Open Access are parts of comprehensive research data management [29], which, in the spirit of Open Science, must be more than the sum of all these parts.

Integration of the Open Data concept into the Open Science Movement should be understood in this spirit, too. Open Data is based on the fundamental assumption that only open knowledge in the form of freely accessible and reusable data is able to promote innovation in the development of knowledge. Distribution and reuse of open knowledge resources is the explicit objective of Open Data and desirable without any restrictions. According to the “open definition” of the Open Knowledge Foundation, open knowledge must not be subject to legal, technical, political, or social restrictions [30].

CULTURAL HERITAGE IN THE DEPARTMENT FOR SCIENTIFIC INFORMATION AT ZIB

Today, ZIB is operating as the Berlin Competence Center for Computing and Data Science combining internationally visible research competences with first-rate scientific services and appropriate infrastructure. At present, the multi-disciplinary units in the department for Scientific Information act as enablers for memory institutions (libraries, museums, archives) by supporting their rapidly growing activities in the digital humanities.

OPEN ACCESS TO OPEN SCIENCE

The task of digiS – Servicestelle Digitalisierung Berlin (service center for digitization, Berlin) – is to make cultural data available. Cultural heritage institutions frequently lack both organizational and technical prerequisites to make their digital content permanently accessible. Data that is to be used in the future, too, requires high quality in terms of their representations and descriptions so that users are able to reconstruct the informational content of an object [31].

Therefore, digiS has been commissioned by the Federal State of Berlin to develop a forward-looking plan for the long-term preservation and availability of the digital cultural heritage. In the future, cultural heritage institutions should have transparent access to their own data via the digital preservation system currently under development as a joint endeavor of digiS and KOBV. At the same time, these cultural data resources should be available to the general public, as far as this is legally permissible. The goal is to maintain the reuse and re- contextualization of cultural data, for both today’s and future users [32].

How users’ expectations regarding (cultural) data have changed, which opportunities open cultural data can make available, and what multitude of uses are possible with them have been demonstrated impressively in 2014 by means of Coding da Vinci [33], the first hackathon using cultural data in Germany.

OPENS

OPEN ACCESS WITH OPUS

The “Green Road” to self-archiving often passes through an institutional repository, which usually is organizationally assigned to a library. In this context, the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), part of ZIB since 1997, provides services for its repository OPUS 4. Following its initial DFG funding, the open-source software has been further developed at ZIB and offered as a national-wide hosting service since 2010 [19]. OPUS 4 supports the relevant technical open standards (OA/PMH, OpenAire, URN, XMetaDissPlus). Hosting also implies quality-tested, trustworthy repositories for which the technical infrastructure was certified at the end of 2014 (DINI Certificate 2013) [20].

In particular, the combination of development and hosting can be considered a success. This becomes evident from mere numbers alone: of the more than 60 known OPUS 4 instances, more than half are hosted at ZIB. Among these are not only numerous institutional repositories but also specialized preprint servers – for example, that of the research center MATHICEN [21].

CULTURAL HERITAGE IN THE DEPARTMENT FOR SCIENTIFIC INFORMATION AT ZIB

Today, ZIB is operating as the Berlin Competence Center for Computing and Data Science combining internationally visible research competences with first-rate scientific services and appropriate infrastructure. At present, the multi-disciplinary units in the department for Scientific Information act as enablers for memory institutions (libraries, museums, archives) by supporting their rapidly growing activities in the digital humanities.

OPEN ACCESS TO OPEN SCIENCE

The task of digiS – Servicestelle Digitalisierung Berlin (service center for digitization, Berlin) – is to make cultural data available. Cultural heritage institutions frequently lack both organizational and technical prerequisites to make their digital content permanently accessible. Data that is to be used in the future, too, requires high quality in terms of their representations and descriptions so that users are able to reconstruct the informational content of an object [31].

Therefore, digiS has been commissioned by the Federal State of Berlin to develop a forward-looking plan for the long-term preservation and availability of the digital cultural heritage. In the future, cultural heritage institutions should have transparent access to their own data via the digital preservation system currently under development as a joint endeavor of digiS and KOBV. At the same time, these cultural data resources should be available to the general public, as far as this is legally permissible. The goal is to maintain the reuse and re-contextualization of cultural data, for both today’s and future users [32].

How users’ expectations regarding (cultural) data have changed, which opportunities open cultural data can make available, and what multitude of uses are possible with them have been demonstrated impressively in 2014 by means of Coding da Vinci [33], the first hackathon using cultural data in Germany.

OPEN ACCESS WITH OPUS

The “Green Road” to self-archiving often passes through an institutional repository, which usually is organizationally assigned to a library. In this context, the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), part of ZIB since 1997, provides services for its repository OPUS 4. Following its initial DFG funding, the open-source software has been further developed at ZIB and offered as a national-wide hosting service since 2010 [19]. OPUS 4 supports the relevant technical open standards (OA/PMH, OpenAire, URN, XMetaDissPlus). Hosting also implies quality-tested, trustworthy repositories for which the technical infrastructure was certified at the end of 2014 (DINI Certificate 2013) [20].

In particular, the combination of development and hosting can be considered a success. This becomes evident from mere numbers alone: of the more than 60 known OPUS 4 instances, more than half are hosted at ZIB. Among these are not only numerous institutional repositories but also specialized preprint servers – for example, that of the research center MATHICEN [21].

OPEN ACCESS WITH OPUS

The “Green Road” to self-archiving often passes through an institutional repository, which usually is organizationally assigned to a library. In this context, the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), part of ZIB since 1997, provides services for its repository OPUS 4. Following its initial DFG funding, the open-source software has been further developed at ZIB and offered as a national-wide hosting service since 2010 [19]. OPUS 4 supports the relevant technical open standards (OA/PMH, OpenAire, URN, XMetaDissPlus). Hosting also implies quality-tested, trustworthy repositories for which the technical infrastructure was certified at the end of 2014 (DINI Certificate 2013) [20].

In particular, the combination of development and hosting can be considered a success. This becomes evident from mere numbers alone: of the more than 60 known OPUS 4 instances, more than half are hosted at ZIB. Among these are not only numerous institutional repositories but also specialized preprint servers – for example, that of the research center MATHICEN [21].

OPEN ACCESS WITH OPUS

The “Green Road” to self-archiving often passes through an institutional repository, which usually is organizationally assigned to a library. In this context, the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), part of ZIB since 1997, provides services for its repository OPUS 4. Following its initial DFG funding, the open-source software has been further developed at ZIB and offered as a national-wide hosting service since 2010 [19]. OPUS 4 supports the relevant technical open standards (OA/PMH, OpenAire, URN, XMetaDissPlus). Hosting also implies quality-tested, trustworthy repositories for which the technical infrastructure was certified at the end of 2014 (DINI Certificate 2013) [20].

In particular, the combination of development and hosting can be considered a success. This becomes evident from mere numbers alone: of the more than 60 known OPUS 4 instances, more than half are hosted at ZIB. Among these are not only numerous institutional repositories but also specialized preprint servers – for example, that of the research center MATHICEN [21].
The academic publishing industry for Science, Technology and Media (STM) is dominated by a small oligopoly of international specialist publishers. Here, a precarious imbalance has developed in which scientists write, are members of editorial boards conducting elaborate peer reviews, and finally buy back the publishing products via their library at a high price. The unmistakable annoyance about this [22] has also led to the situation that publishing houses now feel compelled to offer Open Access options – though seemingly reluctant.

An in-house example is the online version of the Journal Mathematical Programming Computation (MPC). The journal is produced using the Open Journal System OJS and has been hosted by ZIB since 2008. Every year about four issues with three to four articles each are published. The issues are both available on the Springer server and on mpc.zib.de. There also is a printed edition available for a fee.

MPC publishes original research articles covering computational issues in mathematical programming. Articles report on innovative software, comparative tests, modelling environments, libraries of data, and/or applications. A main feature of the journal is the inclusion of accompanying software and data with submitted manuscripts. The journal’s review process includes the evaluation and testing of the accompanying software. Where possible, the review will aim for verification of reported computational results.

MPC supports the creation and distribution of software and data that foster further computational research. The opinion of the reviewers concerning this aspect of the provided material is a considerable factor in the editorial decision process. Another factor is the extent to which the reviewers are able to verify the reported computational results. To this aim, authors are highly encouraged to provide the source code of their software. Submitted software is archived with the corresponding research articles.

Since mathematical software could be considered research data in mathematics compared to the publication itself, it is of great interest to keep that research data for future reference. To this end, an open repository for mathematical software has been established at ZIB – SW-Math.

Mathematical software is a new and emerging component of mathematical knowledge and a bridge between mathematical research and applications of mathematics in science, service, and administration. But visibility and accessibility of mathematical software should be improved. Searching and identifying mathematical software for a given problem are different for a variety reasons:

- Software packages are dynamic (typically, there are a lot of different versions).
- Software is dependent on hardware, operating systems, and other software.
- There are a lot of different licenses and usage conditions.
- Documentation of software is often sparse.
- There exists no standard for software citation.

The mathematical community has provided some special services for mathematical software, but a comprehensive portal for mathematical software is still missing. The portal swMATH (www.swmath.org) is a new approach for a special search engine and information service for mathematical software. A unique feature of this approach is its publication-based approach. The concept of swMATH exploits mathematical publications to get information about software. This is an efficient approach that can be done semi-automatically. As a relevant source of mathematical literature, the bibliographic database zbMATH, the most comprehensive database of mathematical literature, is used.

Currently, the portal swMATH provides free access to information on more than 5,000 mathematical software packages. The further quantitative and qualitative development of information about mathematical software, especially the portal swMATH but also the development of standards, is a common activity of ZIB and Zentralblatt für Mathematik/FIZ Karlsruhe within the Research Campus MODAL, which aim to strengthen the role of mathematical software development as a research area of their own and to improve the visibility and usability of mathematical software.

For years, mathematics has been dreaming vibrantly of a World Digital Mathematics Library [39]. What has to be overcome are the usual legal hurdles and organizational/financial challenges of long-term operation, and ZIB forms part of the movement to realise this vision. In this context, prerequisites are much better than in other disciplines the international mathematical community principally agrees in this matter. Mathematical knowledge is known and classified, much of it has already been retrodigitized, and technical specifications are available. In planning, this world library of mathematics is far more than an ordered collection of publications. The working group, that has been entrusted with the planning recommended in 2014: “The Digital Mathematical Library should serve as a nexus for the coordination of research and research outcomes, including community endorsements, and encourage best practices to facilitate knowledge management in research” [40]. In this recommendation, considerations in terms of Open Access and Open Data come together to form the concept of Open Science. It is a concept that is dependent on broad exchange and sharing, or, to put it in the words of Karl Popper: “The game of science is, in principle, without end. He who decides one day that scientific statements do not call for any further test, and that they can be regarded as finally verified, retire from the game” [41].
Science and industry under one roof

The BMBF Research Campus MODAL starts its first main phase. In four mathematical labs, academia and industry join forces to work on data-driven modeling, simulation, and optimization of complex real-world processes from transportation, energy, medicine, and the development of general solver technology.
This is what the acronym MODAL stands for. MODAL is one out of nine Research Campuses supported by the Federal Ministry of Research and Education as part of the Hightech Strategy of the German government.

Located at and operated by the Zuse Institute Berlin in cooperation with Freie Universität Berlin, MODAL fosters public-private research partnerships. Currently, twelve companies participate, ranging from small spin-offs to established global players. With its 15-year horizon and a research budget of more than two million euros per year, the more than 40 researchers of the campus will advance the data-driven development of pioneering modeling, simulation, and optimization methods. Within the excellence platform of the Einstein Center for Mathematics Berlin, MODAL provides the link to industrial mathematics. A unique feature of the campus is MODAL AG (MAG), a nonprofit-oriented company that offers maintenance, support, and system-integration services for MODAL partners and other interested parties that the participating research institutions cannot provide. MAG ensures the long-term utilization of research results.

The Research Campus MODAL consists of four interdisciplinary labs whose basic application-oriented research is focused on specific innovation problems while at the same time concrete industrial problems are tackled.

The RailLab works on the optimization of train rotations using algorithmic hypergraph theory. Industry partner DB Fernverkehr, the long-distance passenger transport division of Deutsche Bahn, uses the optimization core ROTOR to schedule the German ICE fleet, while Berlin companies LBF and IVU Traffic Technologies market the rotation optimizer VRT-OPTrail in the scheduling system IVU.rail.

The MedLab researches the efficient storage, classification, and pattern recognition of mass spectrometric omics data for cancer diagnosis. Cooperation partner SAP provides high-performance in-memory storage technologies, the SMEs companies 1000shapes and CIT develop professional software for model-based data analysis, and the associated partner Inbion helps improve standard operating procedures.

The GasLab develops mixed-integer nonlinear constraint programming methods to control the gas transport network of Germany. Operator Open Grid Europe, formerly known as E.ON Gas Transport, must ensure supply of quality, non-discriminatory access, efficiency, and flexibility to make the energy revolution in Germany a success.

The SynLab investigates general discrete-continuous-stochastic optimization problems and provides the open-source solver SCIP as an academic research and development platform. The manufacturers of the leading industrial solvers CPLEX, Gurobi, and Xpress, the software integrator GAMS, and SAP and Siemens cooperate within SynLab on the development of new optimization methods.

In addition to the activities of the labs, MODAL bundles respective legal competence, international cooperation, and capacity building. For example, MODAL runs several programs for attracting and educating the next generation of researchers and practitioners. In this respect, MODAL runs the Graduate Research in Industrial Projects for Students (G-RIPS) program in cooperation with the NSF Institute for Pure and Applied Mathematics (IPAM) of the University of California, Los Angeles, and Freie Universität Berlin. Every year, G-RIPS provides an opportunity for high-achieving graduate-level students to join the labs. Working in international teams (two students from the US and two from European universities), an academic mentor, and an industrial sponsor, a real-world research project is addressed. Successful participation is rewarded with ten 1ECTS credits.
RAILLAB: OPTIMAL TRAIN ROTATIONS

The German railway system consists of 340,000 kilometers of tracks. A total of 5,300 cargo and 27,000 passenger trains per day transport more than one million tons of goods and 5.2 million people. The high-speed fleet alone operates about 1,400 trips per day with 250 trains of five different types, conveying about 340,000 people. A single ICE3 train costs about 40 million euros. This rolling stock should be put to best possible use.

Despite the complexity of the system, railway scheduling is still largely done by hand. This is in sharp contrast to the related airline and public-transport sectors, where mathematical fleet and crew optimizers are today established as an industry standard [42]. These tools have, in particular, played an important role in revolutionizing the way airlines have operated since the 1990s. Clearly, the railways want similar tools. The key problem here is vehicle rotation scheduling, the topic of the MODAL RailLab.

RAILWAY SYSTEMS ARE HYPERGRAPHS

The main obstacle in railway optimization is the technical constraints of the wheel-rail system that prevent straightforward attempts to decompose the planning process. In contrast to buses and aircraft, where individual vehicles are considered, trains are composed of multiple units that operate in sequence and orientation, such that classic network flow methods do not apply. Instead, an algorithmic theory of flows in hypergraphs is needed to deal with vehicle rotations and train compositions at the same time [43]. Such methods are also useful for scheduling crews in teams.

Vehicle-rotation-planning problems are large scale, and their solution requires high-performance algorithms. The RailLab has developed a new coarse-to-fine method, which allows them to deal with most of the problems on a coarse train or vehicle level and to focus on the important parts in a finer sequence and orientation level where necessary, guided by bounds that control the approximation error [44]. On this basis, high-quality vehicle rotations can be constructed using novel Lin-Kernighan-type algorithms.

RAILWAY SYSTEMS ARE HYPERGRAPHS

Despite the complexity of the system, railway scheduling is still largely done by hand. This is in sharp contrast to the related airline and public-transport sectors, where mathematical fleet and crew optimizers are today established as an industry standard [42]. These tools have, in particular, played an important role in revolutionizing the way airlines have operated since the 1990s. Clearly, the railways want similar tools. The key problem here is vehicle rotation scheduling, the topic of the MODAL RailLab.

THE ROTOR AND VS-OPT-RAIL OPTIMIZERS

These methods are at the heart of RailLab’s ROTOR and VS-OPT-rail optimization cores that are integrated as solvers in the FEO (Fahr- und Einsatzoptimierung) and ivu.rail systems of the industry partners DB Fernverkehr AG and LBW GmbH and IVU Traffic Technologies AG, respectively. These tools are already in daily use at several railways in Germany and elsewhere. Currently, cyclic strategic planning problems for the “standard week” are solved; this is important for service design and vehicle procurement. To achieve a more direct impact on costs and quality, detailed fully dated problems for a set calendar period have to be solved. This leads to acyclic hyperflow problems with boundary conditions. Ultimately, RailLab aims at providing mathematical optimization support for the complete range of railway planning problems involving rolling stock.
In recent years, the development of high-throughput technologies has enabled the generation of vast amounts of information from multiple biological data sources such as genomics, transcriptomics, proteomics, and metabolomics — the "omics" sources.

Using these data sources has increased our understanding of many molecular and cellular mechanisms in biological systems like the processes underlying severe diseases in the human body. In contrast, this explosive growth of data poses a challenge in our understanding of medicine: current analysis approaches can use the huge amount of collected mass data only minimally; medically important relationships are lost in this data tsunami.

While the potential opportunities are not questioned in principle, many high expectations from the early 2000s have been disappointing so far. For example, proteomics-based biomarkers that have been hyped for quite some time never met the expectations that were promised, such as new diagnostic tools for fighting cancer etc. The main reasons for this were technical difficulties with data analysis and problems with data acquisition, which led to studies that could not be reproduced across laboratories. These problems are usually not due to principle impossibility but mainly exist because of a lack in data quality and analysis methods.

**MEDLAB: DIAGNOSING CANCER USING OMICS-DATA**

**TAMING THE DATA TSUNAMI**

While it has been known for a long time that changes in cells while they are undergoing transformation from "normal" to malignant cells (e.g. during infections) happen on many biological levels, such as genome, transcriptome, proteome, and metabolome, traditional analysis approaches still depend on only one data source — for example, transcriptomics. This is in contrast to the central dogma of molecular biology that states that these levels are actually highly interconnected and depend on each other.
An important part of the energy consumed in Germany is produced from natural gas. Therefore, in optimizing the allocation and usage of the gas grid, one could save a sufficient amount of energy. However, the liberalization of the German gas market and the associated divestiture of gas companies have led to new challenges in the planning and operation of gas-transport grids over the last few years.

The MODAL GasLab addresses these challenges and aims to develop new methods for gas-grid planning and operation, combining the most modern mathematical algorithms and up-to-date information technology. The tools developed in the GasLab will assist gas dispatchers and planners in their work and put them in a position to make better decisions based on foresighted and comprehensive information. The GasLab brings together the main areas of expertise of the scientific partners; that is, modeling, simulation, and optimization to advance the state of the art in gas-grid management and facilitate innovations. This is supported by a detailed understanding of the issues and comprehensive data, which will be ensured by collaboration with and the know-how of the industrial partner Open Grid Europe GmbH.

The main task of the project is the scientific research and development of solution procedures to design efficient algorithms and construct a novel software tool, the Navi, that provides a plan for managing the gas grid in advance for a period of at least twenty-four hours. Similarly to a navigation route guidance device for cars, the tool should provide a safe and cost-minimizing sequence of operating decisions to the gas dispatcher to assist in actual decision making with regard to gas-grid management measures, such as changing the mode of compressor stations or closing or opening valves.

There are several aspects that make the corresponding problem challenging. For example, due to the compressibility of gas, the gas grid itself acts as storage. The appropriate description of the gas behavior involves a set of partial differential equations for each pipeline, and the feasible operating ranges of the gas compressors form a nonconvex set. The model should be able to handle the transient case, which allows a dynamic change of the variables in time. The model involves both discrete and continuous decisions, like opening a valve or deciding to run a compressor on a certain pressure adjustment level, which further complicates the modeling. In addition to the challenges in modeling, a good prediction of transport demands and an estimate of the associated uncertainty is crucial for detecting critical future network situations early to react to these properly. Finally, the navigation tool has to operate a real-world gas grid without interruption.

In addition to technical measures, the gas-grid operators have regulatory options to ensure the secure operation of the grid while satisfying the transport demands of the customers. One option is to restrict a gas power station in a critical situation to obtaining gas from one specific entry point instead of ordering gas from any entry point of the network.

Given a set of power stations at exit points with their corresponding fixed restriction entry point as well as the current network situation, the project aim is the development of a tool, the KWP-Tool, that calculates a foresighted decision suggestion for the gas-grid operator as to which power stations this option will need to be used on the upcoming gas day. Therefore, the tool has to identify critical network situations in advance for the next forty hours and that cannot be balanced by technical measures only. Moreover, since the product may be applied for several gas power stations, the decisions suggested by the KWP-Tool need to be nondiscriminatory. These challenges will be addressed by research on game theory, modeling of gas flow, and statistical forecasting and evaluation methods; enhanced by heuristics derived from expert knowledge from our industrial partner; facilitating both scientific progress and practical use of the novel methods.
SYNLAB: ADVANCING OPTIMIZATION TECHNOLOGY

The development of state-of-the-art optimization software at ZIB started more than 25 years ago. It has been driven by various industrial applications coupled with a strong academic focus. The optimization software developed and maintained at ZIB, forming the SCIP Optimization Suite, is globally recognized and stands as one of the most versatile and best performing open-source packages for mixed-integer linear and nonlinear optimization.

THE POWER OF COOPERATION

In the SynLab of the Research Campus MODAL, the world’s leading researchers and developers from companies like Gurobi [45], FICO Xpress [46], and GAMS [47] cooperate with ZIB researchers to develop the mathematical methods behind the next generation of mathematical optimization tools. In addition, long running collaborations with SAP [48] and Siemens [49] have also been integrated into the SynLab, supplying challenging industrial problems. As a result of the Siemens cooperation, the linear programming (LP) solver SoPlex has become the fastest exact LP solver on the market. Finally, in 2014 Google started to use SCIP in their combinatorial optimization library or-tools [50].

STRIVING FOR THE BEST

The most recent release of SCIP, version 3.1.1, is currently the fastest non-commercial mixed-integer programming solver. SCIP integrates the techniques of mixed integer programming, constraint programming, and satisfiability testing to solve complex problems. In recent years, parallel extensions of SCIP have been developed as part of the UG [51] framework. The result of this development has seen SCIP solving mixed-integer programs on supercomputers utilizing up to 80,000 cores in parallel. Advancements achieved with SCIP have led to the successful completion of numerous outstanding collaborative research projects and underlines many of the projects within the Research Campus MODAL.

CONTINUED, SUCCESSFUL, AND OPENLY SHARED DEVELOPMENT

The projects of SynLab successfully promote the further development of the SCIP Optimization Suite. A focus of SynLab is the development of software to meet the growing number of everyday data-intensive processes. A Google Faculty Award was awarded for the project Mixed Integer Optimization as a Service, which aims to advance SCIP to address modern technological requirements. Furthermore, the SynLab strives to advance mixed-integer programming techniques by improving the cutting plane and branching approaches of SCIP.

Keeping with the “shared development” theme of the SynLab, an extension of SCIP called SCIP-Jack [52] has recently been developed. It is a solver, available in source, for the general Steiner Tree Problem in Graphs – a classic combinatorial optimization problem – and won in multiple categories at the recent DIMACS challenge [53]. SCIP-Jack can solve ten different variants of the Steiner Tree Problem, including the prize-winning Steiner Tree, maximum weight connected subgraph, and the group Steiner Tree Problems. The promotion of the SCIP Optimization Suite to the wider academic community is an important focus of the SynLab. This is achieved through various visits by researchers – including Domenico Salvagnin, Zonghao Gu, Hans Mittelmann, Christina Burt, Atsuko Ikegami, and Andrea Lodi – and the successful hosting of an intensive three-day SCIP workshop at ZIB with more than 50 participants.
FAIRNESS AND LOAD IN DISTRIBUTED FILE SYSTEMS

Utilizing shared heterogeneous storage resources efficiently

Data volume, storage capacity, data exchange, and digital collaboration over the internet have grown rapidly over the last few years. Distributed storage "in the cloud" became common day-to-day practice for many end users, and ‘Big Data’ is the new must-have in business and science. But what does the corresponding storage infrastructure look like and what can be done to provide predictable performance and efficient use of the costly resources – a mix of fast, but expensive, flash memory, spinning, magnetic discs, and classic tape storage? ZIB addresses these questions and contributes advanced solutions.
SCALABLE STORAGE SYSTEMS

Scientific applications as well as commercial cloud services generate large amounts of data that may range in the area of multiple petabytes or even beyond. Storing such volumes of data requires thousands of disks and servers. Managing such infrastructure effectively, where the failure of storage devices, whole servers, or network links is the norm rather than the exception, is challenging.

Various large-scale distributed storage systems exist and are accessible through different interfaces that vary from the traditional POSIX-compliant file system interface to simplified object-based access like Amazon’s S3 and databases that offer ACID semantics, such as Amazon’s Relational Database Service (RDS). In the following, we focus on storage systems that can be accessed via the POSIX file-system interface.

Various large-scale distributed storage systems have in common that they are usually accessed by multiple users concurrently. Users may influence the performance of each other’s applications significantly. They may share network links, read or write request queues, and may delay each other due to limited concurrency abilities of the underlying storage hardware.

DISTRIBUTED FILE SYSTEMS

Distributed file systems have been developed for many years. The Network File System (NFS) was one of the earliest implementations, typically relying on a single server with direct attached storage or remote disks and connected via a storage area network (SAN). The storage devices themselves provide a block interface where fixed-size data can be accessed at defined block addresses. Although NFS supports concurrent access to files over the network and can be scaled to huge capacities by using a SAN, the single server remains a potential performance bottleneck and a single point of failure.

Cluster file systems like GFS2, OCFS2, and SanFS are more scalable and allow clients to access data directly on shared block devices, such as a SAN, while concurrent accesses are coordinated by a lock manager or a central metadata server. This file-system architecture often requires expensive SAN hardware and trusted clients due to limited permission management on the block devices.

OBJECT-BASED STORAGE

The next advancement of distributed file systems is object-based file systems, which separate the handling of the file content and metadata. File content is stored on object storage devices (OSDs), while file metadata is stored and managed by dedicated metadata servers. In contrast to a block device, object storage devices can store data chunks of arbitrary size and are addressed by a user-given object identifier. Object storage devices themselves use block devices internally – for example, HDDs or SSDs – while the low-level block allocation is handled locally by the machines running the OSD servers.

File access in an object-based file system is usually initiated by an open call sent to a metadata server. The metadata server checks the access permissions and returns which OSDs serve the file, typically including a capability object that allows the client to authenticate to the OSDs and authorizes it to read or write the file.

© Travis Wise

UBC Library © by-nc-nd

UBC Library © by-nc-nd
THE DISTRIBUTED FILE SYSTEM XTREEMF

Our distributed and object-based file system XtreeFS was developed at ZIB since 2006. XtreeFS consists of metadata and replica catalogs (MRCs), object storage devices (OSDs), and a directory service (DIR).

DIRECTORY SERVICE

The directory service allows clients to discover the other server components of an XtreeFS deployment. MRCs and OSDs register themselves with their addresses in the directory service.

METADATA AND REPLICA CATALOG

XtreeFS MRCs store all file metadata, that is, the names and hierarchy of files and directories and their attributes like permissions, file sizes, or time stamps. The MRC is responsible for permission checking while opening files. The DIR and MRC services store their data in the key-value store BabuDB, which has been designed with the aim of storing file-system metadata. BabuDB uses a Log-Structured Merge-Tree (LSM-Tree) data structure that holds large amounts of the database in memory and has support for asynchronous checkpoints. The DIR and MRC services can be made fault-tolerant by utilizing BabuDB’s replication.

OBJECT STORAGE DEVICES

File content is stored on XtreeFS OSDs. The OSD service splits files into chunks of a configurable size and stores each chunk in an own file in a local file system. By utilizing a local file system, a wide range of storage devices like spinning disks, SSDs, or RAM disks are supported without dealing with low-level properties of these devices or worrying about block allocation.

Files can be striped similarly to a RAID 0 over multiple OSDs to utilize the capacity and throughput of multiple OSDs for one file. Furthermore, files can be replicated to multiple OSDs. XtreeFS offers different replication policies to cover multiple use cases. The most common replication policy is a quorum based protocol that allows concurrent read and write access to the same file while preserving POSIX semantics.

XtreeFS architecture consisting of a directory service, metadata and replica catalogs, and object storage devices.
ACCESS PATTERNS AND STORAGE-PERFORMANCE CHARACTERISTICS

APPLICATION ACCESS PATTERNS

Storage-access patterns vary for different applications. Many parallel applications like MapReduce jobs or HPC applications that write checkpoints have purely sequential data access. Other applications like relational databases access files at random offsets. Also applications with a fraction of sequential and random access are possible – for example, a relational database that also writes a transaction log. We focus on the cases at both ends of the spectrum. While sequential throughput is measured in MB/s, random throughput is usually measured in I/O operations per second with a size of 4 kB (IOPS).

SPINNING DISKS

The mechanics of spinning disks require disk head movements while accessing blocks in random order. A disk loses much of its peak performance in this phase. Spinning disks reach their peak performance for sequential access, that is, blocks are read or written according to the order of their block addresses. If multiple sequential streams are served by a single disk, the disk has to perform head movements to switch between these streams and the total throughput drops. This results in a near-random-access pattern for a large number of concurrent streams.

SSDS

The behavior of SSDs or RAM disks differs from spinning disks. SSDs and ramdisks do not have mechanical components and are not slowed down by concurrent access. As applications usually perform I/O requests synchronously – that is, the application is blocked until a request returns – multiple I/O streams are useful for fully utilizing such a fast storage device.

MIXED-ACCESS PATTERNS

Our evaluation has shown that mixing sequential and random-access patterns on a single device results in a nonoptimal resource utilization [57]. Adding a single random-access stream to an HDD that serves sequential-access streams results in a significant drop of total throughput. To simplify the device models and keep the reservation scheduler device agnostic, we also do not mix workloads on other device types.

MODELING PERFORMANCE PROFILES

Due to the complex behavior of many uncertainties – for example, caused by different cache layers or unknown software components like the local file system on a storage device – we built an empirical performance profile for each OSD that enters an XtreemFS cluster by running a benchmark that measures sequential and random throughput at different levels of concurrent access.
The capacity and performance in terms of sequential or random throughput, can be controlled by either selecting an OSD that can serve a reservation or striping a reservation over multiple OSDs that provide the necessary resources in sum. Each OSD provides a set of resources like capacity, sequential throughput or random throughput. The available resources of an OSD depend on the already allocated resources and the number of reservations that are currently mapped to it.

Our reservation scheduler receives a stream of incoming requests that have to be scheduled to OSDs immediately (or rejected) while the lifetime of a reservation is unknown. Our scheduling algorithm aims to minimize the number of necessary OSDs while guaranteeing the requested capacity and performance.

If a reservation does not fit on a single OSD, we preferably stripe the reservation over multiple OSDs that are already in use. If that is not possible, unused OSDs are taken into consideration. To separate workload patterns, when scheduling decisions we only take OSDs into account that either are unused or have reservations with the same access pattern.

The selection of individual OSDs at each of these steps is a multidimensional bin-packing problem where each OSD and each reservation is represented by a vector. Each dimension of the vectors represents one of the resources: capacity, sequential throughput random throughput. The sum of all reservation vectors that are scheduled to an OSD must not exceed its resource vector. We make use of the heuristic Toyoda’s Algorithm [59] to find a sufficient OSD for each reservation vector. Toyoda’s Algorithm minimizes the angle between the OSD vector and the sum of all reservation vectors that are scheduled to the OSD. Thereby reservations are scheduled to OSDs that have a similar ratio between different resources – for instance, a similar ratio between sequential throughput and capacity. This strategy should eventually result in a schedule in which as few resources as possible remain free.

The request scheduler has to ensure that the fraction of the total performance, which each reservation gets, is identical to its fraction of all throughput reservations. We implement this property by adding a weighted fair-queuing scheduler that replaces the existing request queue of the OSD.

The most critical part in a shared storage environment is the performance isolation that consists of appropriate reservation scheduling to find proper XtreemFS OSDs and reservation enforcement to ensure the promised performance for each reservation [58].
HETEROGENEOUS CLOUDS

APPLICATION 1

APPLICATION 2

APPLICATION 3

PLATFORM

COMPUTATION

COMMUNICATION

STORAGE

CPUs, GPGPUs, FPGAs, ...

OPEN-FLOW SWITCHES, SOFTWARE ROUTERS, ...

MEMORY, SSDs, HARD DISKS, ...

THE EU PROJECT “HARNESS”

To provide performance guarantees on the application level, guaranteed performance is also required for other resources like computing resources and communication. We developed our solutions in the context of the EU project HARNESS, which has the aim of building a prototype of a platform-as-a-service (PaaS) cloud that provides performance guarantees for heterogeneous infrastructures.

PROGRAMMING MODEL

We have seen how to use a heterogeneous storage cluster, consisting of HDDs, SSDs, or other devices, efficiently. Heterogeneous computing resources may be CPUs, GPUs, and FPGAs, for instance. Writing applications that can run on all of these architectures requires a programming model with abstractions that can be compiled to all of these target architectures. All resources that are used by an application are connected by virtual network links that can be provided by an Open Virtual Switch (Open vSwitch).

A user submits an application containing a manifest that describes possible configurations and cross-resource constraints to the HARNESS platform. The platform itself is resource agnostic and delegates the discovery and allocation of particular resources to different infrastructure resource managers. A cross-resource scheduler ensures that resource constraints like distance limits or throughput requirements between allocated resources are fulfilled. The platform layer builds application models by running a submitted application with different configurations.

EFFICIENT RESOURCE USAGE

Our prototype of a PaaS cloud stack provides quality of service guarantees for arbitrary applications on heterogeneous infrastructure with automatic application profiling. Running applications on the best-fitting devices saves resources and helps to reduce the energy consumption of a data center.
Getting the best out of tomorrow’s HPC processors

Right now, High-Performance Computing (HPC) is going through the transition from the multi-core to the many-core era. This development comes with new challenges for application and system software developers - from modernizing legacy codes to making efficient use of the changing hardware toward developing new programming models for future applications. At ZIB, we are searching for solutions to these problems in close collaboration with leading industry partners.
THE TREND

FROM MULTI- ...

The multi-core era started when physical and technical limitations prevented further increases of the clock frequencies of single core CPUs. Instead, multiple cores were put into a single chip, which led to exponential growth in computing power. Current HPC multi-core processors have around one dozen cores. Leveraging their performance in an application requires software developers to assign tasks, which can be processed in parallel, to each core. Multi-cores are designed to perform well for a wide variety of tasks, such as running desktop operating systems, office programs, multimedia applications, web services, databases, games, and scientific simulations.

... TO MANY-CORES

High-performance computing for science and engineering deals with very specific applications exposing common characteristics. Many-core processors are designed specifically for such workloads. This allows for the further improvement of computing power within the current limitations like chip area and power consumption. Many-cores have multiple dozens of rather simple compute cores that are optimized for numerical computations. The increased chip area is spent on arithmetical units for floating-point computations instead of processor logic and general-purpose optimizations.

The development of many-cores started when graphics processors (GPUs) became freely programmable. Similarities between computer games and scientific simulations allowed scientists to exploit the processing power of the graphics hardware to compute scientific problems. GPU manufacturers quickly recognized the new market and created general-purpose GPUs (GPGPUs) specifically for accelerating HPC systems. Today, GPGPUs are an integral part of many of the fastest systems on the planet. However, GPGPUs are mere accelerators that reside in a host system, just like a graphics card in a PC. Only compute-intensive program parts called kernels are offloaded from the host CPU to the GPGPU for faster processing. To utilize a modern GPGPU, a problem must be partitioned into thousands of small tasks that are computed in parallel.

Alongside the GPGPU trend, Intel defined a Many Integrated Cores (MIC) architecture, which is more similar to x86 multi-cores. The first MIC implementation, the Xeon Phi coprocessor (Knights Corner), has around 60 cores that draw their computing power from vector processing units. Enabling vectorization in a program, especially for legacy codes, is a major problem we face and is closely connected to memory access optimization.

1 Intel(R) Xeon Phi(TM) (Knights Corner) coprocessor.
From the early 1970s to the 1990s, vector processors were rated the quasi architectural standard for the design of supercomputers. Representatives like the legendary Cray 1M, the Cray X-MP, and later the Y-MP have been installed at Zuse Institute Berlin. The latter was operated until 1995. However, the increasing performance and the decreasing prices of conventional processors triggered (with a few exceptions) the end of the vector processor era in the mid-1990s. Around the year 2000, vector processing re-emerged in the form of short vector SIMD units, starting with SSE and Altivec. Current multi-core CPUs can perform four double-precision operations at once using AVX or AVX2. The Intel Xeon Phi many-core coprocessor can do eight per vector unit of which it has one per core.
AUTO-VECTORIZING COMPILERS

Modern optimizing compilers are not only capable of inserting SIMD vector instructions when explicitly told to by the programmer but manage to introduce SIMD execution even into larger code segments on their own.

Auto-vectorization by the compiler turns out to be satisfactory for (highly) regular data processing – for example, in dense linear algebra operations with its regular loop structures. Many other cases – that is, with more complex loops - require explicit vectorization, meaning the introduction of compiler directives into the code in order to give additional information to the compiler or manual vectorization. While the latter is notably intrusive to the code, the directive based approach is not so at first sight. However, for codes that are not prepared for SIMD at all, heavy code re-writing together with a careful selection of the most appropriate compiler directives is required. Additionally, deficiencies to present specific optimization strategies to the compiler may render the directive based approach impractical in certain situations. In these cases, manual vector coding can be considered as the a resort.

As the compiler vendors are aware of these issues, auto-vectorization capabilities are pushed strongly. Besides loop transformations and code reorganization, the current research has its focus on managing irregularities within loops to approach SIMD execution. Emerging SIMD instruction sets therefore will cover masked SIMD operations – for example, to enable the separation of divergent execution paths. A further feature programmers can draw on when optimizing their codes for SIMD is SIMD-enabled functions. By means of specific directives, the compiler can be encouraged to expand scalar function definitions to work on vector arguments and, hence, be callable from within vectorizable loops. Investigations regarding the usability of these new features are carried out within ZIB’s activities at the Intel Parallel Computing Center (IPCC).

Application performance gain due to SIMD vectorization. The application has been executed on a Xeon Phi cluster using up to eight compute nodes. Only some hotspots were optimized for SIMD; that is, there is further potential for additional speedups over the reference execution.
Besides automatic vectorization by the compiler, there are also means to manually vectorize application code. For instance, the OpenCL programming language provides vector types. These types allow for operation with vectors of multiple—like instance-eight—floating point numbers in parallel. Arithmetic operations with such vectors can then be directly mapped to the corresponding instructions of the hardware. For conventional languages like C, most compilers offer intrinsics, which are low-level operations that allow directly programming of the hardware’s vector units. For the C++ language, there are libraries that encapsulate those intrinsics for different hardware platforms behind a common interface. This approach leads to more portable code. The advantage of manual vectorization is control over the applied vectorization strategy by the programmer. This allows for the optimization of the memory layout for contiguous load and store operations to and from the processor’s vector registers.

We used the GPU-HEOM code to conduct a case study with different vectorization approaches. GPU-HEOM uses the Hierarchical Equations of Motion (HEOM) method to simulate the energy transfer in natural and artificial photoactive systems. We found the memory layout to be the most important factor for efficient vectorization on both multi- and many-cores. Together with other optimizations, we could improve the performance of the Hexciton computation of HEOM by up to 6.3 times on the Xeon Phi coprocessor. The continuation of this work was successfully proposed as an interdisciplinary DFG (German Research Foundation) project that starts in early 2015.
Since 2013, ZIB has been operating the Research Center for Many-Core High-Performance Computing, which was established together with Intel as one of the first Intel Parallel Computing Centers (IPPC) in Europe. The goal is to modernize existing HPC codes, especially those relevant for the HLRN user community, and to develop novel algorithms and programming models for the many-core era. The IPCC program provides ZIB researchers with early access to upcoming hardware products and allows for close collaboration with Intel engineers.

VECTORIZATION

Besides modern codes that can be designed with fine-grained parallelism and vectorization in mind directly from the beginning, legacy codes often do not fit the new multi- and many-core processors without adaptations. Particularly, many of the complex community codes from the different fields of science have their origins some decades ago, and their design originates from the very different architectural features found in processors up to now. For instance, codes that once were written for vector processors have been extended with new functionalities, but not necessarily with vector processors as the target platform in mind. As SIMD vectorization has already experienced a revival on the current multi-cores. For many-cores it is an absolute necessity, among other adaptations, to redesign certain codes so as to exploit SIMD execution.

Optimizing an application for vector processing may cover the whole range from redesigning algorithmic elements to meet SIMD requirements, to choosing appropriate data layouts, to finally writing down the code in a SIMD-friendly way. For the latter, we focus on generic vector programming so that codes are portable across different platforms. The latter is mainly achieved by using compiler directives and vector libraries that encapsulate assembly or SIMD intrinsic functions in the C/C++ programming language.

CONCURRENT KERNEL OFFLOADING

We not only have large-scale computations on the screen, but we also aim to have small-scale workloads benefit from multi- and many-core performance. If a single problem instance cannot utilize all computing resources, why not running multiple of them simultaneously? Concurrent kernel offloading approaches share multi- and many-cores to increase the overall application throughput.

EFFICIENT INTRA- AND INTER-NODE OFFLOADING

We developed an offloading framework that focuses on efficiency and on unifying local (intra-node) offloading and remote (inter-node) offloading within a single programming interface. This allows for fine-grained offloading and scaling applications beyond the many-core coprocessors installed in a single node.
The concurrent kernel offload (CKO) mode approaches the inherent underutilization of modern many-core processors when workloads are too small to exploit the available computing resources on their own. Instead of processing only a single problem instance, multiples of them are handled concurrently, thereby sharing the many-core processor for an increased overall load. Within the IPCC activities at ZIB, the CKO feature has been integrated into the GLAT program as so as to leverage the immense computing power of the Intel Xeon Phi coprocessor for the simulation of small drug-like molecules. One of the key insights in that respect is the necessity to partition the many-core processor appropriately so that concurrent computations do not interfere with each other. Results of extensive studies regarding CKO on the Xeon Phi coprocessor have found their way into the book High Performance Parallelism Pearls by J. Jeffers and J. Reinders.

Current many-core solutions – that is, GPGPUs and the Intel Xeon Phi – are accelerators that run inside a host system that uses one or more multi-core CPUs. The standard programming model for this heterogeneous system architecture is offloading, where only compute-intensive portions of the program are computed on the accelerator. These compute kernels can be computed faster on the accelerator, but the process of offloading takes time and resources by itself, which adds additional costs to the program execution. For instance, input data must be transferred to the accelerator’s memory, the code for the computation must be transferred and executed in coordination with the host, and finally the results must be transferred back to the host. Only program parts where the gain overcompenses the costs are suitable for offloading. Less overhead means more potential for offloading even fine-grained tasks. We developed the HAM-Offload framework with the goal to minimize the offloading costs and thus make many-cores accessible for a wider class of applications. By utilizing modern C++ metaprogramming techniques, HAM-Offload can reduce the cost per offload by a factor of 28 compared with Intel’s Language Extension for Offload (LEO). For a small problem instance of GLAT that causes fine-grained offloads, this translates into an application speedup factor of two.

Existing solutions are also limited to intra-node, or local, offloading within a single node. Offloading tasks to remote accelerators over a network requires the usage of another programming model, like the Message Passing Interface (MPI). HAM-Offload unifies local and remote offloading and even allows reverse offloading from a many-core coprocessor towards a multi-core host system. This is completely transparent for the programmer, who uses a single API (Application Programming Interface) to put work on local and remote targets. Behind the scenes, HAM-Offload uses MPI for network communication, which ensures compatibility with most systems installations. For existing offload applications, this allows for scaling beyond the limited number of local coprocessors without introducing explicit inter-node communication into the application code. We were able to demonstrate this for the GLAT application using the coprocessors of 16 nodes instead of just one. Inter-node, or remote, offloading with HAM-Offload results in only half as much offloading costs as Intel LEO does for local offloads.

Remote offloading using a hybrid approach combining MPI and Intel LEO (Language Extension for Offload) vs. HAM-Offload. The box labeled MIC is the remote Xeon Phi™ accelerator.
ZIB PUBLICATIONS

BOOKS


ZIB REPORTS


ZIB PUBLICATIONS

DISSERTATIONS


Ivan Davics (2014), TUB, Robust Tail Assignment.

Jesco Humpola (2014), TUB, Gas Network Optimization by MINLP.


Timo Berthold (2014), TUB, Heuristic algorithms in global MINLP solvers.


Sharon Hüffner (2014), FUB, Modularity in Biological Networks.

Alexey Shadrin (2014), FUB, Positional Information Storage in Sequence Patterns.

REFERENCES

DESIGN OF NANOPHOTONIC DEVICES


IMAGE ANALYSIS FOR BIOLOGY


REFERENCES

FAIRNESS AND LOAD IN DISTRIBUTED FILE SYSTEMS


