Curriculum Vitae of Oliver Gutsche

Oliver Gutsche
Fermi National Accelerator Laboratory / Scientific Computing Division
P.O. Box 500 / MS. 205, Batavia, IL, 60510
+1-630-840-8909

gutsche@fnal.gov
gutsche.github.io
LinkedIn
InSpire ORCID

I am a particle physicist and am conducting **leading edge research** for New Physics Beyond the Standard Model of Particle Physics as well as precision Standard Model measurements.

I have multiple years of experience in analyzing high-energy collisions at different particle colliders using a multitude of different techniques. I have published many papers in leading journals and am currently a member of the CMS collaboration at the Large Hadron Collider (LHC) at CERN. In my recent studies at the LHC, I have lead searches for evidence of physics beyond the Standard Model using top quarks, and contributed to searches for Supersymmetry and Dark Matter. One of my most noticeable publications is the Observation of the Higgs Boson in 2012.

I am a **leader in scientific computing** and have acquired deep knowledge and expertise in scientific software and computing infrastructure. My active involvement in HEP science allows me to guide the science community to benefit from the latest computing developments, bridging the worlds of science and IT.

Particle physics is based on particle detection by sophisticated experimental devices and their comparison to accurate simulations. Scientific software consists of millions of lines of C++ and python code and is needed to extract these physics results. I am an expert in object oriented software development, statistical data analysis methods and Monte Carlo simulation techniques as well as various optimization and machine learning techniques.

High Energy Physics (HEP) requires very large amounts of computing resources to analyze simulations and data recorded by the detectors. I have extensive experience in planning, developing, and operating distributed computing infrastructures that provide access to several hundred-thousand computing cores and hundreds of petabytes of disk space. I am intimately familiar with scientific grid sites, academic and commercial clouds and the largest U.S. supercomputers.

I am part of a worldwide community process to prepare the software and computing infrastructure for the High Luminosity LHC (HL-LHC, 2026), which will require many times the computing resources as today. I was part of the community planning exercise to provide the basis for an R&D program to reduce this increase significantly through disruptive changes to software and computing. Examples are new analysis paradigms using industry technologies, and vectorized and SIMD programming technologies and machine learning approaches to exploit new hardware architectures like accelerators and GPUs, as well as new concept in the area of data organiztion, management and analysis (DOMA). I was co-author for the overview white paper of the community and was co-editor of the topical white paper about the future of data analysis in High Energy Physics.

My recent interest is to explore industry technologies for analysis of petabyte scale datasets. For this purpose, I started the CMS Big Data Project. One component was a collaboration with Intel on scaling of analysis facilities to Petabytes using Apache Spark, that was concluded successfully in January 2019. Fermilab joined CERN openlab to make this collaboration a reality, which opens up the possibility for more collaborations and projects in the future.

I held many management positions at the Fermi National Accelerator Laboratory and within the international CMS collaboration, supervising up to 100 individuals across many time zones. In March 2019, I was appointed U.S.CMS Software and Computing Operations Program manager overseeing the U.S. CMS Tier-1 and Tier-2 facilities as well as software maintenance and development efforts for core software, computing infrastructure software and analysis systems. In this context, I am working together with DOE and NSF and partners at the universities and other labs to enable analysis of LHC particle collisions in the U.S. for the 2500 physicist strong CMS collaboration, and prepare for HL-LHC.

I regularly speak at interational conferences and workshops and am member of the editorial board of the journal for Computing and Software for Big Science.

Research Positions

09/2014 - Present	Scientist	Fermi National Accelerator Laboratory (Fermilab)
06/2009 – 09/2014	Application Physicist I	Fermi National Accelerator Laboratory (Fermilab)
06/2005 – 05/2009	Research Associate	Fermi National Accelerator Laboratory (Fermilab)
09/2001 – 02/2005	Doctoral Candidate	Deutsches Elektronen Synchrotron (DESY)

Education

2001-2005	**University of Hamburg**, *Doctor of Natural Sciences*, Hamburg, Germany
Thesis title	*Measurement of beauty quark cross sections in photoproduction with the ZEUS experiment at the electron proton collider HERA* (thesis)
Advisors	Prof. Dr. Robert Klanner, Dr. Achim Geiser
1996 - 2001	**University of Hamburg**, *Diploma in Physics*, Hamburg, Germany
Thesis title	*Development of the trigger algorithm for the MONOLITH experiment* (thesis)
Advisors	Prof. Dr. Robert Klanner, Dr. Achim Geiser

Assignments: U.S. CMS Software and Computing Operations Program

03/2019 - present	U.S.CMS Software and Computing Operations Program manager
10/2016 - 02/2019	U.S.CMS Software and Computing Operations Deputy Program manager
10/2016 - 02/2019	L2 manager for Software and Support in the U.S. CMS Software and Computing Operations Program Execution Team
03/2014 – 09/2016	L2 manager for Operations in the Software and Computing Operations Program Execution Team

Assignments: Fermi National Accelerator Laboratory - Scientific Computing Division

10/2016 - present	Deputy Head of the Scientific Services Quadrant
10/2014 – 09/2016	Assistant Head of the Scientific Computing Division for Science Operations and Workflows
10/2014 – 12/2014	Interim Department Head of the Scientific Data Processing (SDP) Department of the Scientific Services Quadrant
10/2013 – 09/2014	Deputy Department Head of the Scientific Data Processing (SDP) Department of the Scientific Services Quadrant
09/2012 – 09/2014	CMS Operations Group Leader in the Scientific Data Processing (SDP) Department of the Scientific Services Quadrant

Assignments: CMS Collaboration - Offline & Computing Project

09/2015 – present	CMS Offline & Computing Project Focus Area Lead for Infrastructure and Resources

10/2014 – 08/2015	Member of the CMS Offline & Computing Project Management Board
01/2012 – 09/2014	CMS Offline & Computing Project Computing Operations L2 Manager
07/2009 – 12/2011	CMS Offline & Computing Project Data Operations L2 Manager
01/2007 – 07/2009	CMS Offline & Computing Project Release Validation Manager

Scholarships and Awards

2013	CMS Young Researcher Prize
2008	Exceptional Performance Recognition Award of Fermilab
2003	German Academic Exchange Service scholarship for an academic stay at University College London (UCL), London (Great Britain)

Research Experience

CMS collaboration: 2005 – Present

I joined the CMS collaboration at the LHC in 2005 and my research focus has been the search for New Physics Beyond the Standard Model of Particle Physics as well as precision Standard Model measurements.

I was a founding member of an analysis group with members from Fermilab/UCSD/UCSB, focusing on final states with leptons. The approach proved to be successful; after early publications such as a measurement of the top quark cross section, the focus shifted to new physics and beyond the Standard Model processes. We were leaders of the WW to dilepton analysis in the CMS Higgs discovery paper, and searches for SUSY in same-sign and opposite-sign dilepton as well as single lepton channels. The group is currently continuing the searches for SUSY in lepton final states as well as Standard Model processes.

I have been supervising several Fermilab postdoctoral researchers helping me to pursue my research interests.

- Together with Jacob Linacre, I concentrated on exploiting the dilepton signature to search for pair production of a heavy top-like quark (t'). I continued studying the properties of top quarks exploiting angular distributions of the dilepton final state. We were the first to use the dilepton final state to measure the top pair charge asymmetry at the LHC to further investigate the deviations seen at the Tevatron. We published LHC Run 1 papers for top pair spin correlations and top quark polarization for the 7 TeV and 8 TeV datasets as well as the top pair charge asymmetry for the 7 TeV and 8 TeV datasets.
- Since 2015, I am supervising Fermilab PostDoc Matteo Cremonesi. He created a new dark matter analysis effort at the Fermilab LHC Physics Center (LPC), searching for dark matter particles in various channels. The first publication presented the search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum in the 13 TeV 2016 dataset of LHC Run 2. The second publication describes the search for dark matter produced in association with a Higgs boson decaying to a pair of bottom quarks in the same dataset. The group is now concentrating on other mono-object channels with an expanding effort at the LPC.
- Since 2018, I am supervising Fermilab PostDoc Nick Smith. He joined the Higgs efforts of the LPC and is contributing to the analysis of the Higgs decay channel into two bottom quarks.

ZEUS collaboration: 2001 – 2005

I started my graduate research at the University of Hamburg in 2001 joining the ZEUS collaboration during the upgrade of the HERA collider and the integration of the new ZEUS micro-vertex silicon strip detector. Because of its precise tracking capabilities, I planned to use secondary vertex and impact parameter techniques to identify beauty quarks decaying into muons.

I first concentrated on the identification of muons. In a team of graduate students, I improved the ZEUS muon reconstruction by exploiting redundancies of the different detector components from tracking, calorimetry, to the muon detectors. I then used the complete pre-upgrade 1996-2000 dataset to extract beauty quark cross sections in photo production. I used the transverse momentum of the muon relative to a jet to identify beauty quarks and measured differential angular correlation cross sections, which were found to be in agreement with NLO predictions. With the first data after the upgrade, the micro-vertex detector became available. I used the dataset from 2003-2004 to identify beauty quarks using impact parameter techniques for the first time in ZEUS. Both measurements were documented in my thesis.

Technical Experience

Particle physics is based on particle detection by sophisticated experimental devices and their comparison to accurate simulations. High Energy Physics (HEP) requires very large amounts of computing resources to analyze simulations and data recorded by the detectors. I have deep knowledge of planning, developing, maintaining and operating distributed computing infrastructures providing access to several hundred-thousand computing cores and many hundred of petabytes of disk space. I am intimately familiar with scientific grid sites, academic and commercial clouds and the largest supercomputers at High Performance Computing centers in the U.S. and across the world. The infrastructure then execute scientific software consisting of millions of lines of C++ and python code that is needed to extract physics results. I am an expert in object oriented software development, statistical data analysis methods and Monte Carlo simulation techniques as well as various optimization and machine learning techniques.

These technical aspects of my work are closely connected to my physics research, as they enable the analysis of particle physics detector data and simulations as a basis to extract physics results. My active involvement in HEP science allows me to guide the science community to benefit from the latest computing developments, bridging the worlds of science and IT.

Computing Infrastructure

In March 2019, I was appointed the U.S. CMS Software and Computing Operations Program manager enabling U.S. CMS collaborators to analyze the CMS experiment's data. From October 2016 to February 2016, I was the deputy manager for the same operations program. I am overseeing the operation of the U.S. CMS Tier-1 site at Fermilab and 7 U.S. Tier-2 sites at Caltech, Florida University, MIT, University of Nebraska-Lincoln, Purdue University, UC San Diego, and Unversity of Wisconsin-Madison. The program funds over 80 FTE of effort to administer the sites, maintain the computing infrastructure and aid in strategic R&D projects.

I am employing my extensive knowledge of scientific software and computing in contributing to the worldwide community efforts to plan for the software and computing infrastucture for the High Luminosity LHC (HL-LHC). Starting in 2026, the HL-LHC will produce many times the amount of data of the current LHC running periods. In addition, the collisions and the corresponding simulations will be many times as complex as today. I was an integral part of the community planning process and my input was documented in the Roadmap for HEP Software and Computing R&D for the 2020s. In addition, I was co-editor of the HEP Software Foundation Community White Paper Working Group - Data Analysis and Interpretation.

My recent research interest in computing infrastructure is asking the question if analysis in HEP can be conducted more efficiently using tools developed and used by industry. Instead of employing the ROOT toolkit that was entirely developed by the HEP community, I am exploring using toolkits like Apache Spark or similar technologies. I created a research group spanning researchers from Fermilab, CERN and the Universities Princeton, Padova and Vanderbilt. The CMS Big Data Project is also very closely working together with industry, latest in a project with Intel concluded in January 2019 in the context of CERN openlab. To realize this project, Fermilab joined CERN openlab and I organized the DOE approval process with the help of the Fermilab Office of Partership

and Technology Transfer and the Fermilab Legal Office. I also managed a Laboratory Directed Research and Development project (LDRD) to develop innovative technology for Big Data delivery to array-based analysis code, the Striped Data Server for Scalable Parallel Data Analysis. The project concluded successfully in January 2019 and produced a prototype that is currently being used by a diverse set of experiments from collider physics to astro-particle physics.

The CMS collaboration appointed me Focus Area Lead for Services and Infrastructure in the CMS Software and Computing project in 2015. I am coordinating the efforts of the worldwide submission infrastructure, innovative new ways of using resources at commercial clouds and supercomputing centers, and the development of computing infrastructure services like data management and workflow management systems.

From September 2014 to September 2016, I was appointed Assistant Scientific Computing Division Head for Science Operations and Workflows in the Scientific Computing Division of Fermilab. I was responsible for the delivery of scientific computing services to all Fermilab experiments, including High Energy Physics experiments (e.g. CMS), Neutrino Physics experiments (e.g. NOvA, Minerva), Intensity Frontier experiments (e.g. mu2e, Muon g-2) and Astroparticle Physics experiments (e.g. DES). As member of the senior management team, I developed strategic plans to evolve the infrastructure and operational procedures. For example, I developed a new storage strategy that simplifies the operation and usage of the more than 30 PB of disk space at Fermilab. I was also responsible for maintaining the computing strategy as part of the Laboratory Strategy Effort and reported to the laboratory directorate.

In October 2016, I was appointed Deputy Head of the Scientific Services Quadrant. This quadrant is the user facing arm of the Scientific Computing Division, and develops computing infrastructure software components for data and workload management for the whole scientific program of Fermilab, supporting neutrino, muon, and astro-particle experiments as well as CMS.

The CMS collaboration appointed me lead of the Data Operations Project in 2009. Using my deep involvement in analysis and my expertise in computing, I was responsible for the timely delivery of all data and MC samples for analysis, a significant contribution to the overall success of the experiment. In 2012, CMS extended my responsibilities and appointed me to lead all of the CMS Computing Operations Project, adding the care of over 70 computing centers distributed all over the world and all central computing services of CMS. I was supervising the contributions of more than 60 scientists and engineers to the Computing Operations Project worldwide. The team was overseeing the readiness of all the computing facilities and monitor both central workflows and analysis and the transfers of data and MC samples between the sites. At the same time, I was a L2 manager in the U.S. CMS Software & Computing Operations Program responsible for Computing Operations. In this capacity and also before, I reported regularly to funding agencies and took part in reviews of DOE and NSF.

Software

In the U.S. CMS Software and Computing Operations Program, I was responsible for the Software and Support area from October 2016 to February 2019. Under my guidance, the CMS software framework CMSSW was maintained and evolved, as well as critical R&D was started in the areas of vectorized tracking software, machine learning and novel analysis facilities.

Since 2005 and before the start of the LHC data taking in 2010, I was deeply involved in getting the CMS software ready for data taking. I was the lead developer of an innovative tracking algorithm that was used during the commissioning of the CMS detector called RoadSearch. I conducted the first software tutorials in CMS, teaching the CMS community the basics of analysis software and how to perform analysis on the GRID, using a user-friendly GRID analysis tool called CRAB, of which I was one of the lead developers.

During my graduate student time at DESY from 2001 to 2005, I was one of the proponents and lead developers of a new object-oriented and ROOT-based event display. The upgrade of the ZEUS detector made it necessary to integrate the new and changed detector components in the event visualization solution of ZEUS. A client-server

structure allowed physicists to display events without direct access to the event store. Online events could also be displayed worldwide with very small latency during data taking.

Committees

- October 2015 - October 2016: Fermilab's Scientific Advisory Council
- October 2018 - October 2020: Fermilab's Users Executive Committee

Other Professional Activities

- In 2017, I was asked to join the editorial board of the journal for "Computing and Software for Big Science" published by Springer.
- In 2018, I was asked to be Co-Editor of the American Physics Society (APS) Division of Particles and Fields (DPF) white paper as input to the European Particle Physics Strategy Update 2018 – 2020, responsible for the computing section.
- Organizer of the Computing & Machine Learning parallel session at The CPAD Instrumentation Frontier Workshop 2018 "New Technologies for Discovery IV".
- Active in public outreach:
  - Regular participation in Fermilab's Ask-A-Scientist program as lecturer and answering questions of the general public.
  - Tour guide for Fermilab's Saturday Morning Physics program, especially of the computing facilities.
  - Regular question-and-answer sessions for high school classes visiting Fermilab.

Physics Publications with Major Personal Contributions

A.M. Sirunyan et al., **Search for dark matter produced in association with a Higgs boson decaying to a pair of bottom quarks in proton–proton collisions at** $\sqrt{s} = 13TeV$, *Eur. Phys. J.* C79 (2019) 280, doi:10.1140/epjc/s10052-019-6730-7, arXiv:1811.06562 [hep-ex]

A.M. Sirunyan et al., **Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at** $\sqrt{s} = 13$ **TeV**, *JHEP.* 06 (2018) 027, doi:10.1007/JHEP06(2018)027, arXiv:1801.08427 [hep-ex]

V. Khachatryan et al., **Measurements of t t-bar spin correlations and top quark polarization using dilepton final states in pp collisions at sqrt(s) = 8 TeV**, *Phys. Rev.* D93 (2016) 052007, doi:10.1103/PhysRevD.93.052007, arXiv:1601.01107 [hep-ex]

V. Khachatryan et al., **Measurements of** $t\bar{t}$ **charge asymmetry using dilepton final states in pp collisions at** $\sqrt{s} = 8$ **TeV**, *Phys. Lett.* B760 (2016) 365–386, doi:10.1016/j.physletb.2016.07.006, arXiv:1603.06221 [hep-ex]

S. Chatrchyan et al., **Measurements of** $t\bar{t}$ **spin correlations and top-quark polarization using dilepton final states in** pp **collisions at** $\sqrt{s} = 7$ **TeV**, *Phys. Rev. Lett.* 112 (2014) 182001, doi:10.1103/PhysRevLett.112.182001, arXiv:1311.3924 [hep-ex]

S. Chatrchyan et al., **Measurements of the** $t\bar{t}$ **charge asymmetry using the dilepton decay channel in pp collisions at** $\sqrt{s} = 7$ **TeV**, *JHEP.* 04 (2014) 191, doi:10.1007/JHEP04(2014)191, arXiv:1402.3803 [hep-ex]

S. Chatrchyan et al., **Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC**, *Phys. Lett.* B716 (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235 [hep-ex]

Computing Publications with Major Personal Contributions

J. Albrecht et al., **A Roadmap for HEP Software and Computing R&D for the 2020s**, *Comput. Softw. Big Sci.* 3 (2019) 7, doi:10.1007/s41781-018-0018-8, arXiv:1712.06982 [physics.comp-ph]

M. Cremonesi et al., **Using Big Data Technologies for HEP Analysis**, (2019), arXiv:1901.07143 [cs.DC]

L. Bauerdick et al., **HEP Software Foundation Community White Paper Working Group - Data Analysis and Interpretation**, (2018), arXiv:1804.03983 [physics.comp-ph]

D. Berzano et al., **HEP Software Foundation Community White Paper Working Group – Data Organization, Management and Access (DOMA)**, (2018), arXiv:1812.00761 [physics.comp-ph]

J. Chang et al., **Striped Data Server for Scalable Parallel Data Analysis**, *J. Phys. Conf. Ser.* 1085 (2018) 042035, doi:10.1088/1742-6596/1085/4/042035

O. Gutsche et al., **CMS Analysis and Data Reduction with Apache Spark**, *J. Phys. Conf. Ser.* 1085 (2018) 042030, doi:10.1088/1742-6596/1085/4/042030, arXiv:1711.00375 [cs.DC]

O. Gutsche et al., **Big Data in HEP: A comprehensive use case study**, *J. Phys. Conf. Ser.* 898 (2017) 072012, doi:10.1088/1742-6596/898/7/072012, arXiv:1703.04171 [cs.DC]

List of Presentation and Talks

O. Gutsche, **CMS Software, Computing and Analysis Tools**, (2019), Plenary Talk at the CMS Data Analysis School 2019 (CMSDAS 2019), (Material)

O. Gutsche, **Exploring the Energy Frontier using Amazon Web Services**, (2018), Amazon Booth at Supercomputing 2018 in Dallas, TX, (Material)

O. Gutsche, **The Future of Large Scale Scientific Computing - Trends in computing hardware**, (2018), US ATLAS Workshop 2018 Open Day at the University of Pittsburgh, (Material)

O. Gutsche, **Particle Physics - A world wide journey from recording particles to analysis using big computing**, (2017), Plenary Talk at the Chicago Council on Science & Technology Panel: Fermilab and the New Frontiers of Physics, (Material)

O. Gutsche, **50 years Fermilab - Computing Innovations**, (2017), Talk at the Fermilab 50th Anniversary Symposium, (Material)

List of Media Articles and Mentionings

A. Purcell, **Oliver Gutsche: Fermilab joins CERN openlab, works on data reduction project with CMS experiment**, (2017), Article in CERN openlab News, (Article)

M. May, **Oliver Gutsche: A Spark in the dark**, (2017), Article in ASCR Discovery, (Article)

M. May, **Oliver Gutsche: Open-source software for data from high-energy physics**, (2017), Article in Phys.Org, (Article)

- Full List of Physics Publications with Major Personal Contributions can be found here.
- Full List of Computing Publications with Major Personal Contributions can be found here.
- Full List of Publications from all Collaborations and Experiments can be found here.
- Full List of Presentations and Talks can be found here.
- Full List of Articles and Media Mentionings can be found here.