Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Scientific computing and data preservation

Andrii Verbytskyi and Oliver Schulz

MPP Annual project review, MPP, München,
December 14, 2021

Computing Resources

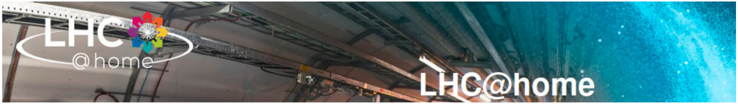
Available computing resources

MPP users have access to a hierarchy of computing resources:

- In-house desktop workstations with Condor batch-system
- MPP Linux-cluster at MPCDF
- MPG supercomputers Cobra and Raven at MPCDF
- HPSS data archive at MPCDF
- Experiment-specific resources (Grid, ...)

MPP move to Garching in 2022:
Opportunity for tighter integration of computing systems with MPCDF!

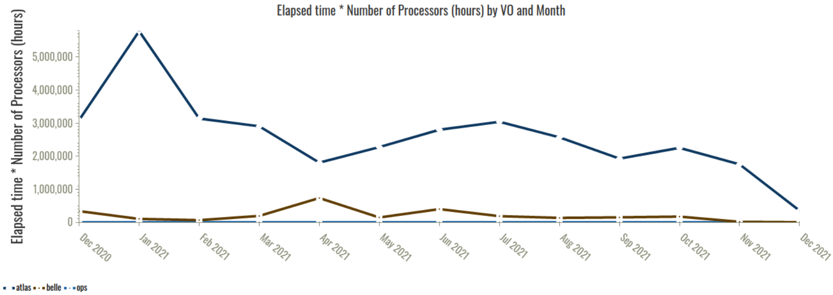
MPP Desktop Batch Systems



Rank	Name	Recent average credit	Total credit	Country
1	Agile Boincers	6,322,059	7,304,534,294	Switzerland
2	AGLT2	3,286,336	3,285,773,431	United States
3	praguelcg2	2,806,034	275,695,323	Czech Republic
4	TRIUMF-LCG2	1,056,172	1,054,498,293	Canada
5	MPI für Physik	514,013	317,099,069	Germany

- Currently 220 SuSE-Linux desktops workstations, backed by 2.5 PB CephFS storage cluster, running
  - Condor with 1821 cores
  - BOINC with about 200 multi-core slots
- Condor prioritised over BOINC, BOINC only uses free resources
- We're the largest desktop-system contributor to ATLAS@home.
- Support for Singularity containers!
- There's free capacity on MPP Condor for your jobs!

MPP Linux-cluster at MPCDF



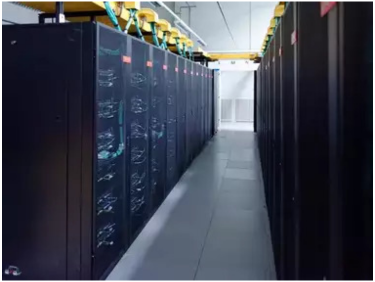
ATLAS grid jobs at MPP@MPCDF, EGI accounting

MPP operates a dedicated compute cluster at MPCDF:

- Compute capacity: 186 nodes, 9572 cores, 4 TB RAM
- Storage capacity: 600 TB Storage (GPFS), 6 PB dCache (3PB used, mainly Atlas)
  - GPFS: 630TB, 90% used (plus 600TB more in Q1 2022)
  - dCache: 6 PB (3PB used, mainly my ATLAS)
- Operating system: CentOS-7
- SLURM share 65% grid jobs (80% ATLAS, 20% Belle), 35% local jobs

MPG supercomputers at MPCDF

Cobra (since 2018):
3188 nodes, 136960 CPU cores with 529 TB RAM,
128 Tesla V100 plus 240 Quadro RTX 5000 GPUs with 7.9 TB RAM

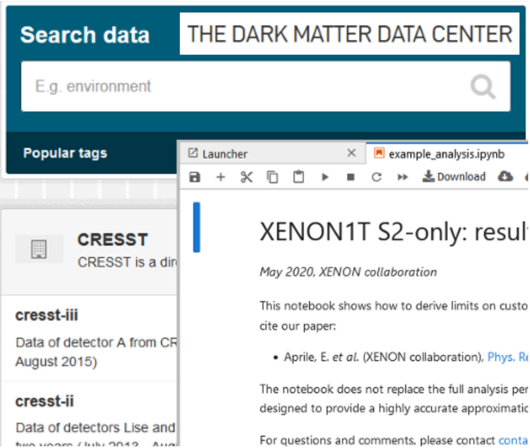


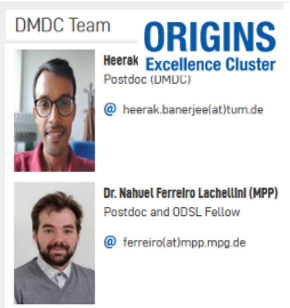
Raven (since 2020/21):
1592 nodes, 114624 CPU cores with 375 TB RAM,
768 Nvidia A100 GPUs with 30 TB RAM.

- MPP people get access to these machines (Nov '21: 16 users, 5.8%)

Software

The Dark Matter Data Center





Data and model repository for dark matter searches

- CKAN instance for data, Gitlab for code/models, Binders for example analyses and workflows

Cuba

- Cuba is a multidimensional numerical integration library
- Motivation: Very common problem,
  but efficient and stable solutions highly non-trivial
- Developer: Thomas Hahn
- Four different integration algorithms,
  all with C/C++, Fortran, Mathematica and Julia interface
- Multi-purpose, used in many physics software projects
- Also non-physics / industry users
- New release 4.2.1 in June 2021

http://www.feynarts.de/cuba/

[T. Hahn, "CUBA: A Library for multidimensional numerical integration," CPC168 (2005)[2]

Developments in computer algebra

- LoopTools, evaluation of scalar and tensor one-loop integrals: minor release in November 2021
- FeynHiggs, code for calculations of the masses, mixings and much more of the Higgs bosons in the MSSM: minor release in July 2021
- FormCalc, tree-level and one-loop Feynman diagrams: minor release in December 2021
- and even more supported by Thomas Hahn https://wwwth.mpp.mpg.de/members/hahn/

BAT.jl: Bayesian Analysis Toolkit in Julia

- Motivation: Bayes' theorem simple on paper, but numerics are hard
- O. Schulz, A. Caldwell et al., involvement from MPP, TUM/ODSL, TU-Dortmund
- BAT.jl v2.0.5 release in September 2021, written in Julia
- Multiple algorithms: MH, HMC, AHMI (Caldwell et al.), bridge sampling, nested sampling, Cuba (T. Hahn)
- Automatic parameter space transformation and differentiation
- Several successful Bachelor theses in 2021: Harmonic mean integration, bridge sampling and inference on ODE-based Covid-models.
- BAT.jl 3.0 planned for Spring 2021 with new features like normalising flows

https://github.com/bat/BAT.jl

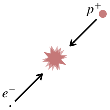
[Schulz et al. "BAT.jl: A Julia-Based Tool for Bayesian Inference", SNCS (2021)

[Caldwell et al. "Integration with an Adaptive Harmonic Mean Algorithm", IJMPA (2020)]

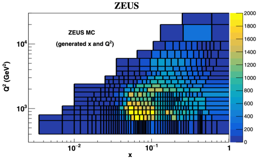
[Hafych et al. "Parallelizing MCMC Sampling via Space Partitioning", arXiv:2008.03098 (2020)

[Caldwell et al. "Infections and Identified Cases of COVID-19 ...", arXiv:2005.11277 (2020)

Parton density inference: QCDNUM + Julia + BAT

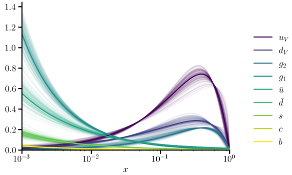


Collider experiment



Count data in x and Q² bins

MCMC methods to fit data



Visualisation of PDF results

- Can probe the structure of hadrons via electron-proton collisions to infer parton distribution functions (PDFs)
- Motivated by the ZEUS high-x results with low event numbers, aim to go beyond using Bayesian methods
- Wrapped Fortran QCDNUM (M. Botje) for Julia: new registered Julia package QCDNUM.jl
- currently preparing prior parametrisation forward model to run first inference soon.

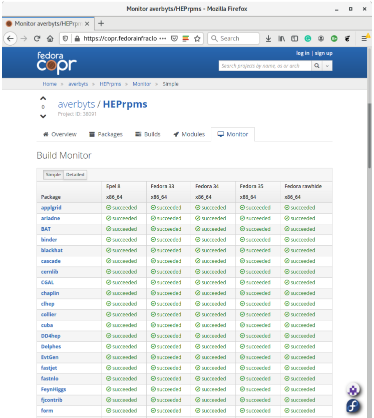
https://github.com/cescalara/QCDNUM.jl, https://github.com/cescalara/PartonDensity.jl

[R. Aggarwal, M. Botje, A. Caldwell, F. Capel, A. Verbytskyi, O. Schulz], supported by ORIGINS ODSL

HEPrpms: software deployment done correctly

HEP software packaged as RPMs, see
https://copr.fedorainfracloud.org/coprs/averbyts/HEPrpms
&contribution to the ACAT2021 conference.

- Support of CentOS/Fedora/SuSe on x86_64.
- ≈ 60 packages, many with MPP roots, mostly related to MCEGs.
- COPR by IBM/RedHat as cloud storage/building backend.



Package	Epel 8 x86_64	Fedora 33 x86_64	Fedora 34 x86_64	Fedora 35 x86_64	Fedora rawhide x86_64
applgrid	succeeded	succeeded	succeeded	succeeded	succeeded
ariadne	succeeded	succeeded	succeeded	succeeded	succeeded
BAT	succeeded	succeeded	succeeded	succeeded	succeeded
binder	succeeded	succeeded	succeeded	succeeded	succeeded
blackhat	succeeded	succeeded	succeeded	succeeded	succeeded
cascade	succeeded	succeeded	succeeded	succeeded	succeeded
cernlib	succeeded	succeeded	succeeded	succeeded	succeeded
CGAL	succeeded	succeeded	succeeded	succeeded	succeeded
chaplin	succeeded	succeeded	succeeded	succeeded	succeeded
clhep	succeeded	succeeded	succeeded	succeeded	succeeded
collier	succeeded	succeeded	succeeded	succeeded	succeeded
cuba	succeeded	succeeded	succeeded	succeeded	succeeded
DD4hep	succeeded	succeeded	succeeded	succeeded	succeeded
Delphes	succeeded	succeeded	succeeded	succeeded	succeeded
EvtGen	succeeded	succeeded	succeeded	succeeded	succeeded
fastjet	succeeded	succeeded	succeeded	succeeded	succeeded
fastnlo	succeeded	succeeded	succeeded	succeeded	succeeded
FeynHiggs	succeeded	succeeded	succeeded	succeeded	succeeded
fjcontrib	succeeded	succeeded	succeeded	succeeded	succeeded
form	succeeded	succeeded	succeeded	succeeded	succeeded

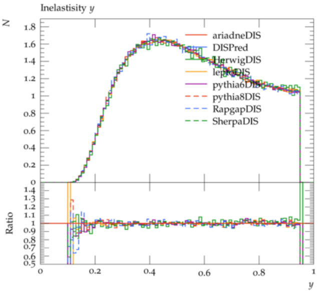
- **All software from HEPrpms is easily available on centrally managed MPP machines thanks to Thomas!**
- Quick, reproducible, and reliable zero-maintenance installation for PC, virtual machines and containers.
- CI in multiple HEP projects, debug of MC event generators for ATLAS, usage in pheno analyses.

Developments in MC generator tools, part I

- THEP8I, the Pythia8 hadronization interface for Herwig/ThePEG.
  - Minor release 2.0.1, adoption at LHC.
- HEPMC3, the Monte Carlo Event record library, release 3.2.4.
  - Optimisations and technical improvements, e.g. compression support.
  - Will be **the event record library** in ATLAS@LHC run II $\rightarrow$ huge preparation work.
  - Significant efforts are made to use it as a common tool for HEP and Heavy Ion communities.

Developments in MC generator tools, part II

- Studies on Monte Carlo tuning using Bayesian Analysis and ThePI8: collaboration with (Salvatore La Cagnina *et al.*@TU Dortmund). Using BAT.jl, containers with HEPrpms and Julia tools!



- Promote the usage of HERA measurements in RIVET (A. Buckley, J. Butterword, M. Wing *et al.*@Rivet), (H. Jung *et al.*@DESY), (M. Diefenthaler *et al.*@JLAB) and assure these are used correctly.

Figure: Inelasticity distribution obtained with Rivet from simulated $e^{\pm}p$ data.

Data Preservation @MPP

ZEUS & H1 @MPP

ZEUS:

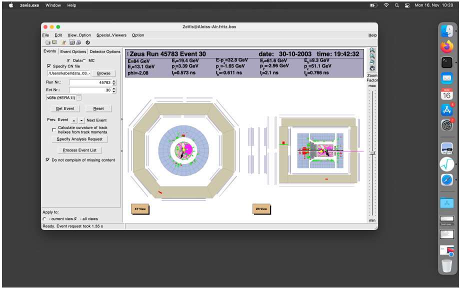
People: I. Abt, A. Caldwell, A. Verbytskyi, (H. Abramowicz, A. Levy).
Software: ROOT is the analysis framework for ZEUS. ZEUS approach does not require re-reconstruction of data.
The **mutable active code** is in (private) Github repositories, with enabled GitHub Actions CI, scheduled builds, documentation. **Compiles on CentOS7/CentOS8/(CentOS9)/(Win)/(OSX) @x86_64**.
+password-protected yum repository → good for containers.
Data: all data

H1:

People: D. Britzger, V. Chekelian, G. Grindhammer, Ch. Kiesling.
H1 software: "the good old way". Code is maintained in DESY, external builds are possible. The software requires regular rebuilds from scratch and is kept in shape by DESY. Depends on DESY software stack and/or **recent LCG releases on CentOS7 @x86_64**. This approach has its advantages.
Data: all data

ZEUS@MPP: offer MC and keep software in shape & portable.

- ZeVis, ROOT-based ZEUS event display: multiplatform support, CentOS9 support in progress, bugfix release in 2021.



- CNINFO, ZEUS database utility: multiplatform support.
- FORMOZA, ZEUS MC event generators + HepMC3 interface to **any modern MCEGs**: CentOS8 support, CentOS9 in progress.

OPAL and JADE@MPP

JADE:

People: S. Bethke, S. Kluth, H. von der Schmitt + A. Verbytskyi.

- Fortran software in **public** github repository with CI, and docs.
- Works with generic CentOS7+/OSX10.15+ with GNU/Intel Fortran.
- Generation of MC events is possible with **any modern MCEG**.
- A ROOT-based framework for analyses of JADE/OPAL ntuples.
- A perfect candidate to join CERN OpenData.

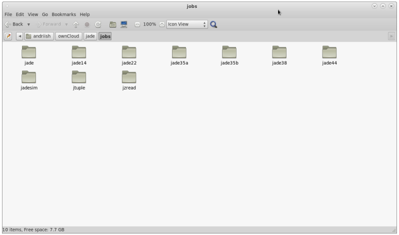
OPAL:

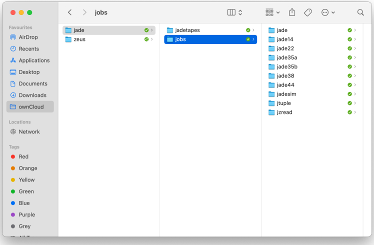
People: S. Bethke, S. Kluth, H. von der Schmitt + A. Verbytskyi.

- A ROOT-based framework for analyses of JADE/OPAL ntuples.
- A more modern software stack is possible with newer CERNLIB.

The data

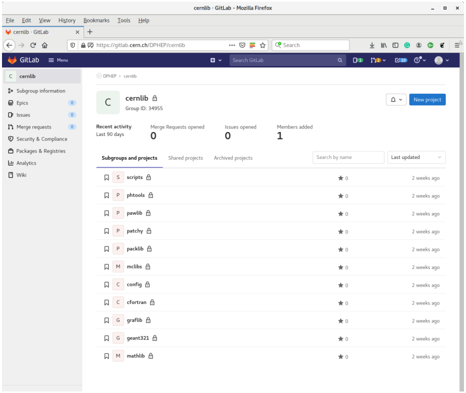
- The OPAL, JADE, ZEUS, & H1 data → MPP/MPCDF dCache/archive system.
- ZEUS, OPAL and JADE data → plain ROOT ntuples and can be analysed immediately for all physics processes (ZEUS) or for typical QCD (JADE, OPAL).
- 2021: MPCDF dCache instance upgrade & ongoing data migration.
- JADE data is ≈600Gb → is also in MPCDF ownCloud → Works on Windows, MacOSX and Linux.





Broader DPHEP efforts and software

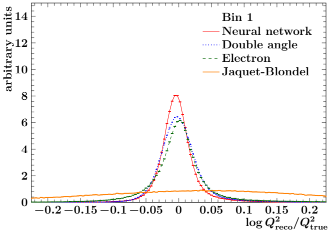
- Regular meetings and discussions.
- Revitalisation of CERNLIB. Suggested long time ago by DP@MPP, but get CERN support just this year.

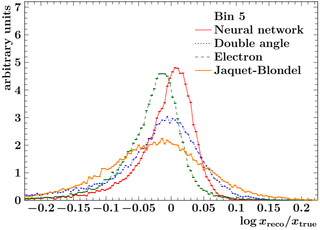


- Improve compatibility of popular MCEGs@HERA with modern tools. Collaboration & thanks to RAPGAP and CASCADE, (H. Jung *et al.*@RAPGAP,@CASCADE)

Reuse the preserved data@MPP: ZEUS Common Ntuples

M. Diefenthaler, A. Farhat, A. Verbytskyi and Y. Xu, "Deeply Learning Deep Inelastic Scattering Kinematics,"[arXiv:2108.11638], &here.





Figures: Resolution of Q^2 and x observables obtained with different methods in simulated and reconstructed ZEUS data.

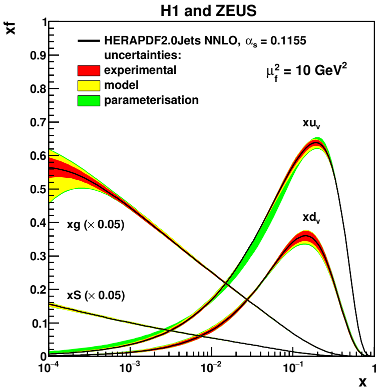
- Use deep NN to reconstruct the kinematic observables in DIS using the combination of classical methods and information on HFS/electron.
- Collaboration with computer scientists, **create** the NN architecture.
- Emphasise physics and understanding, **NOT** blackboxing the NN.
- Analysis of ZEUS data as a first step to future analyses on EIC.

H1 & ZEUS papers: good and diverse modern physics

H1 & ZEUS papers

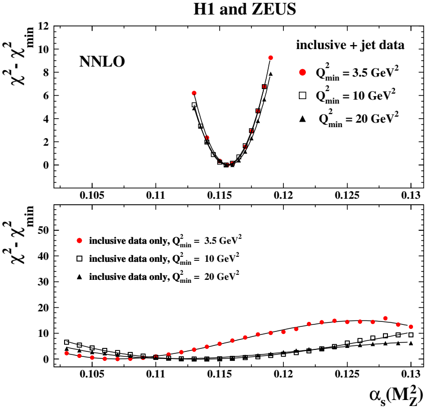
- I. Abt *et al.* [ZEUS and H1], "Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions," [arXiv:2112.01120 [hep-ex]].
- I. Abt *et al.* [ZEUS], "Azimuthal correlations in photoproduction and deep inelastic **ep** scattering at HERA," [arXiv:2106.12377 [hep-ex]].
- V. Andreev *et al.* [H1], "Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding," [arXiv:2108.12376 [hep-ex]].
- V. Andreev *et al.* [H1], "Measurement of charged particle multiplicity distributions in DIS at HERA and its implication to entanglement entropy of partons," Eur. Phys. J. C **81** (2021) no.3, 212 [arXiv:2011.01812 [hep-ex]]

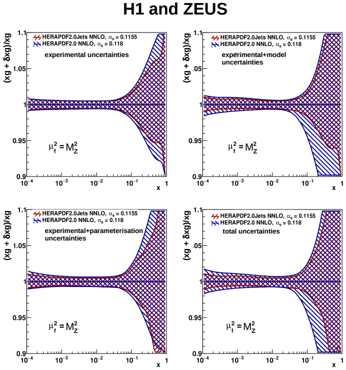
Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions, part I



- **A global QCD analysis of HERA data now with jets.**
- **One of the most important legacies of HERA.**
- Thanks to **8** jet data-sets.
- and close collaboration with theoreticians [3].
- Two new precise and consistent PDF sets: HERAPDF2.0Jets NNLO with $\alpha_s = 0.1155$ and $\alpha_s = 0.118$.

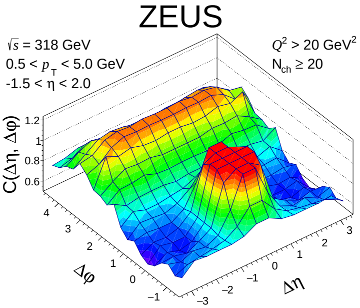
Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions, part II

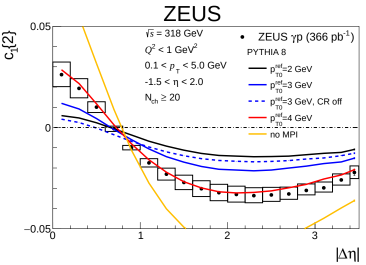




- Important input for the α_S average, the result from the QCD fit $0.1156 \pm 0.0011\ (\mathrm{exp})\ ^{+0.0001}_{-0.0002}\ (\mathrm{model}\ +\mathrm{par})\ \pm 0.0029\ (\mathrm{scale})$
- Reduced uncertainties on the gluon DF vs. HERAPDF2.0.

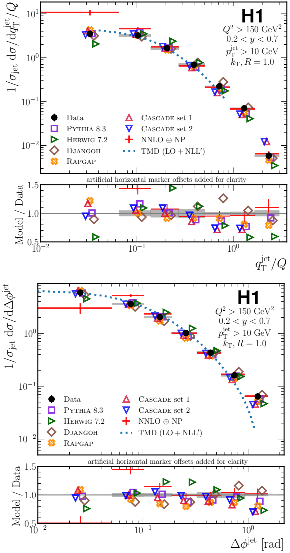
Azimuthal correlations in photoproduction and deep inelastic **ep** scattering at HERA





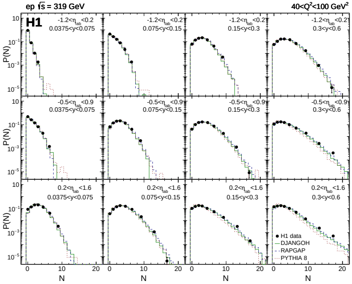
- A study of the famous "ridge" effect – a study of collective effects, correlations between particles in DIS and for the first time in γp.
- Seen in heavy-ion collisions, later in *pp*, but not in *ep* DIS (see JHEP **04** (2020), 070) [1]) and not in e^+e^-.
- No clear indication of the ridge effect in γp, however, the comparison with different models suggest presence of MPI.
- Very interesting in the context of large interest of MCEG community to MPI and future physics at EIC.

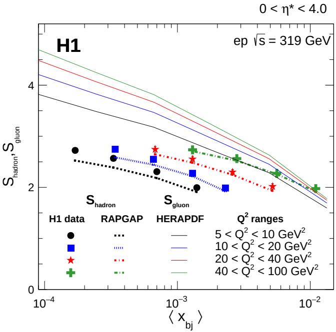
Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding



- One of the most inclusive jet measurements: the final state is a jet + scattered lepton.
- The LO configuration should give back-to back configuration in $\Delta\phi^{jet} \approx 0$.
- The deviations provide an insight into the NLO QCD effects and could serve as a probe for transverse-moment dependent PDFs.
- Unfolding with neural network code.

Measurement of charged particle multiplicity distributions in DIS at HERA and its implication to entanglement entropy of partons



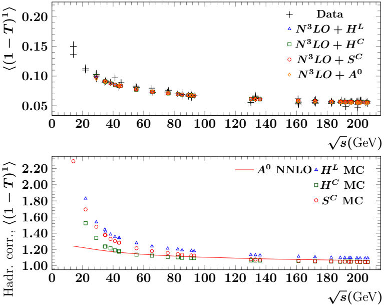


- The charged particle multiplicity distributions in bins of Q^2, y, η.
- In the analysis the hadron entropy S_{hadron} was measured and compared to the entropy calculated from the PDF functions S_{gluon}.
- The prediction $S_{hadron} = S_{gluon}$ is not confirmed by the measurement.

More diverse physics contributions

QCD analyses beyond NNLO/thoughts on FCC-e^+e^-

A. Kardos, G. Somogyi and A. Verbytskyi, “Determination of α_S beyond NNLO using event shape averages,” Eur. Phys. J. C **81** (2021) no.4, 292 [arXiv:2009.00281 [hep-ph]]



$\alpha_S(M_Z)_{NNLO} = 0.12911 \pm$ 0.00177(*data and hadronisation*)$\pm$ 0.0123(*scale*)

- Estimation of N^3LO contributions to event shape moments and α_S extraction at N^3LO.
- The study aims to clarify if we will reach better precision in α_S with more e^+e^- data and more precise predictions e.g. N^3LO?
- Most likely – no. The main obstacle will be the understanding of hadronisation.

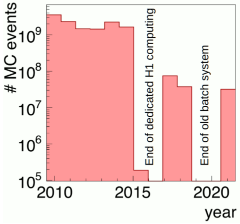
Ultimate solution: dedicate efforts @FCC-e^+e^- & study the hadronisation.

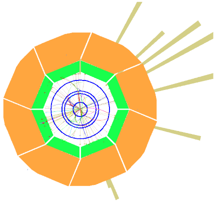
Preservation through Modernisation [arXiv:2106.11058]

- The entire software stack of HERA's H1 experiment was migrated to a modern computing infrastructure (`CentOS7`, `gcc9`, `C++20`, DESY's HPC facilities, `gitlab`, `LCG`-dependencies, `ROOT6`, `python`-analysis, relocatable, backward-containerization w/ SL5, etc. . .)

  D. Britzger, S. Levonian, S. Schmitt, D. South, EPJ Web Conf. 251 (2021) 03004 [arXiv:2106.11058].

- ∼1.2M LOCs: `FORTRAN` sim./reco. software, and the `C++` analysis framework
- A full MC production campaign was performed with modernized framework
- New MC generators become available (here: `Pythia8.3`)
- New collaborators from EIC Community are joining (several PhDs, Postdocs, Staff)



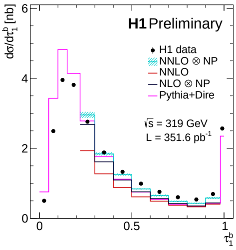


Rapid physics results with new framework

First measurement of 1-jettiness event shape in NC DIS

[H1prelim-21-032], J. Hessler for H1 [arXiv:2111.11364], MA J. Hessler

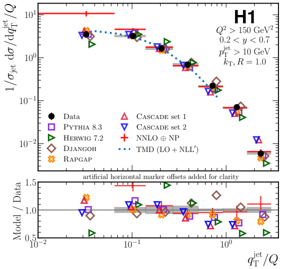
- $\tau_1^b = \frac{2}{Q^2} \sum_{i \in X} \min\{xP \cdot p_i, (q + xP) \cdot p_i\}$
- Equivalent to thrust, when normalised to $Q/2$
- Single- and triple-differential data compared to NNLO & various event generators
- MC generators provide unsatisfactory description



First measurement lepton-jet correlation in NC DIS

H1 Collaboration, subm. to PRL (2021) [arXiv:2108.12376]

- First application of unbinned unfolding using machine learning in HEP (Omnifold)
- Variables potentially sensitive to TMD's
- TMD predictions describe low q_T/Q
- Fixed-order predictions describe high q_T/Q



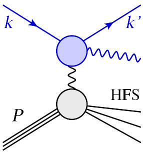
Three more editorial boards were formed in 2021

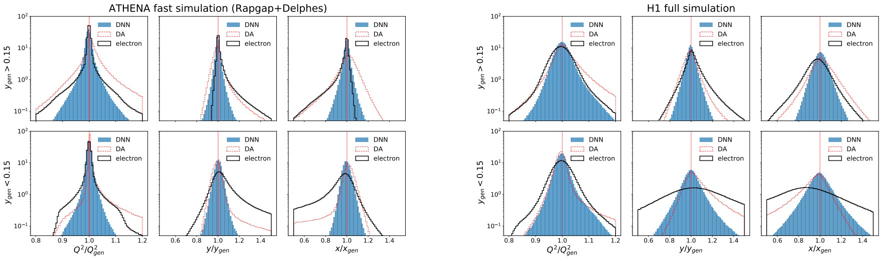
Bridge made between HERA and EIC

Reconstructing the Kinematics of Deep Inelastic Scattering with Deep Learning

M. Arratia, D. Britzger, O. Long, B. Nachman, in print, NIM A 1023 (2022) 166164 [arXiv:2110.05505]

- Feed-forward deep-neural-net (DNN) with 7 hidden layers is used to combine 15 observables for classification and regression of QED radiation in NC DIS events
- DNN can also be used to improve DIS variables Q^2, x and y
- First analysis to use HERA and EIC simulated data
- EIC fast simulation is benchmarked against H1 full sim.



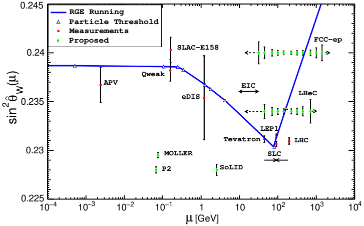


High-luminosity energy frontier DIS: LHeC and FCC-eh

Electroweak physics at the FCC-eh

D. Britzger, M. Klein, H. Spiesberger [MPP-2021-208]

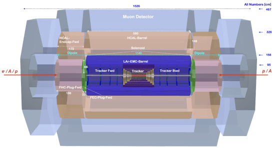
- Study sensitivity of inclusive DIS data at FCC-eh to parameters of Electroweak theory
- Unique measurements feasible with high precision
- Running of weak mixing angle can be measured
- Plot: J. Erler, R. Ferro-Hernandez and X. Zheng.



An Experiment for Electron-Hadron Scattering at the LHC

K.D.J. Andre, *et al.*, subm. to EPJ C [MPP-2021-95]

- New interaction region design for LHeC
- Same IP for *ep*, *eh*, *pp* and *hh* collisions possible
- Updated and symmetrized detector layout for extended physics scope, incl. now *hh*/*pp*
- Joint proposal with ALICE3 appears to be feasible



Conclusions

- The computing projects in MPP are diverse and have a huge impact on the global landscape of scientific computing.
- MPP/MPCDF are important suppliers of computing resources for the HEP, astroparticle physics, theory and other communities.
- The preserved data allows for a large number of new physics results.

Backup slides

DRAFT:Bibliography I

[1] I. Abt *et al.* [ZEUS], "Two-particle azimuthal correlations as a probe of collective behaviour in deep inelastic *ep* scattering at HERA," JHEP **04** (2020), 070 doi:10.1007/JHEP04(2020)070 [arXiv:1912.07431 [hep-ex]].

[2] T. Hahn, "CUBA: A Library for multidimensional numerical integration," Comput. Phys. Commun. **168** (2005), 78-95 doi:10.1016/j.cpc.2005.01.010 [arXiv:hep-ph/0404043 [hep-ph]].

[3] J. Currie, T. Gehrmann, A. Huss and J. Niehues, "NNLO QCD corrections to jet production in deep inelastic scattering," JHEP **07** (2017), 018 [erratum: JHEP **12** (2020), 042] doi:10.1007/JHEP07(2017)018 [arXiv:1703.05977 [hep-ph]].