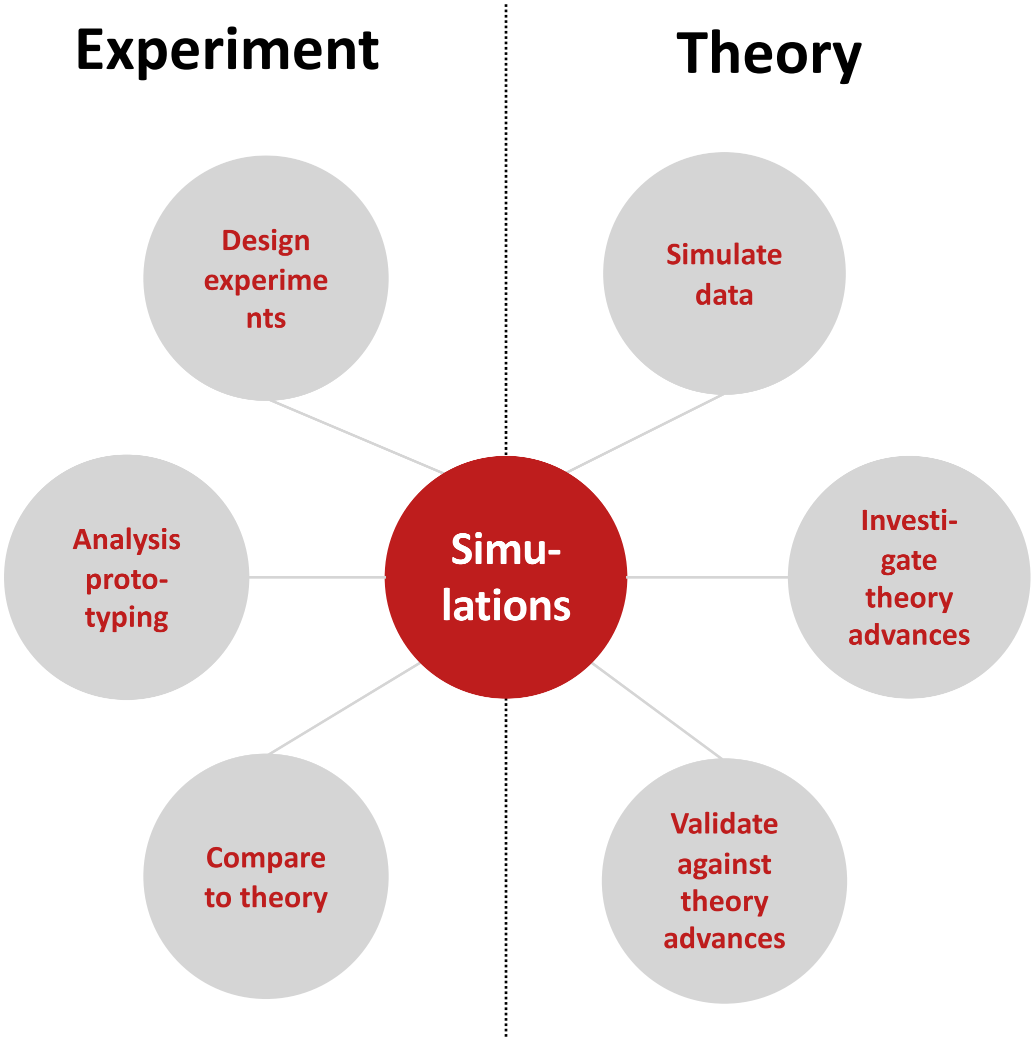
Introduction to EIC Software with a focus on simulations

Markus Diefenthaler

Jefferson Lab

Simulations in Experiment and Theory



Software tutorials

- **Introduction: Emphasizing the interplay of science and software & computing**
  - Markus Diefenthaler (JLab)
- **Software for Detector Design**
  - Sylvester Joosten (ANL)
  - Joe Osborn (ORNL)
- **Software for Physics Studies**
  - Wouter Deconinck (University of Manitoba)
  - Kolja Kauder (BNL)

Prepare for Kolja's tutorial

- Install singularity on your system: https://github.com/ECCE-EIC/Singularity#troubleshooting
- Start the singularity container following the instructions on:
  https://github.com/ECCE-EIC/Singularity#singularity-container-for-ecceeic-fun4all

No preparations needed for the other tutorials.

Electron-Ion Collider

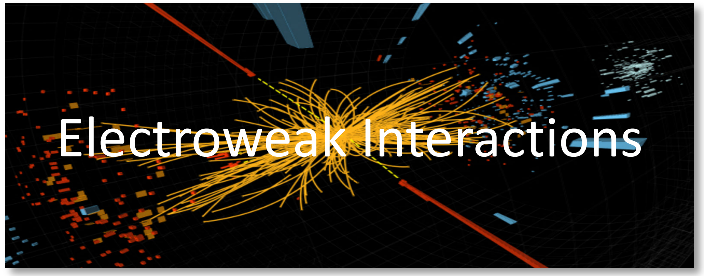
A new frontier in science

Nuclear Physics

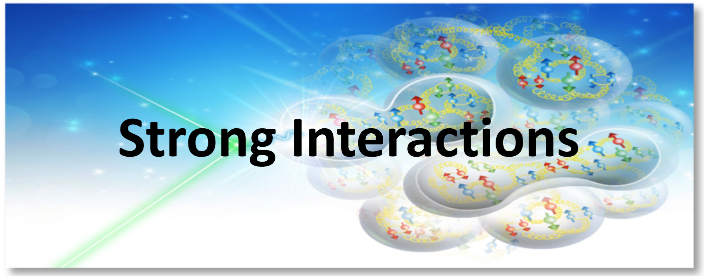
Further exploration of the Standard Model

Dark matter searches

Electroweak symmetry breaking



Deeper understanding of QCD



Mission of Nuclear Physics

- discover, explore, and understand all forms of nuclear matter

Frontiers in Nuclear Physics

- One of the enduring mysteries of the universe is the nature of matter—what are its basic constituents and how do they interact to form the properties we observe? The largest contribution by far to the mass of the matter we are familiar with comes from protons and heavier nuclei.
- Although the fundamental particles that compose nuclear matter—quarks and gluons—are themselves relatively well understood, exactly how they interact and combine to form the different types of matter observed in the universe today and during its evolution remains largely unknown.

Nobel Prizes in Physics and Nuclear Physics / EIC Science

Hideki Yukawa (1949) "for his prediction of the existence of mesons on the basis of theoretical work on nuclear forces"
But the quark-gluon origin of the nuclear binding force remains an unknown.

Robert Hofstadter (1961) "for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons"
But the 3D quark-gluon structure of nucleons remains an unknown.

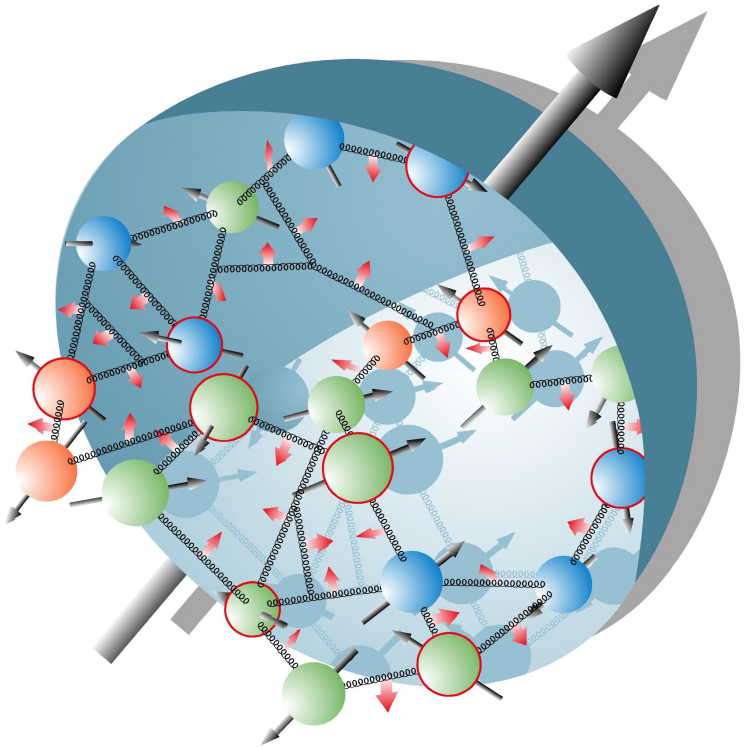
Jerome Friedman, Henry Kendall, Richard Taylor (1990) "for their pioneering investigations concerning deep inelastic scattering of electrons on protons and bound neutrons, which have been of essential importance for the development of the quark model in particle physics"
But the role of gluons in protons and bound neutrons remains unknown.

David Gross, David Politzer, Frank Wilczek (2004) "for the discovery of asymptotic freedom in the theory of the strong interaction"
But the confinement aspect of the theory remains unknown.

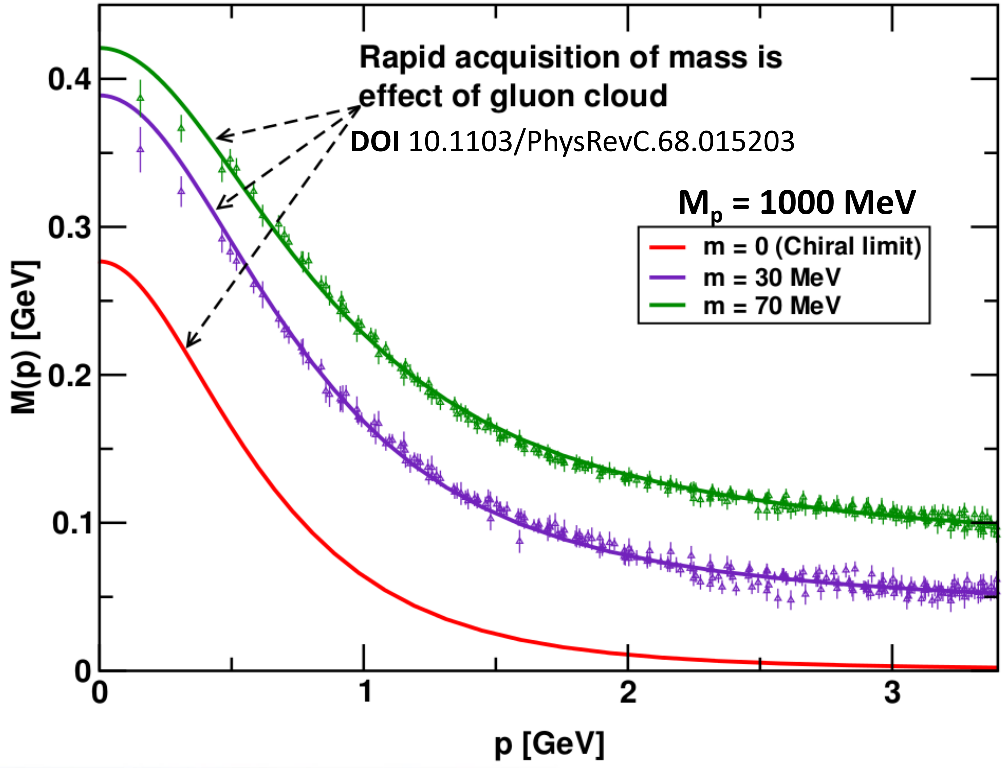
Yoichiro Nambu (2008) "for the discovery of the mechanism of spontaneous broken symmetry in subatomic physics "
But how dynamical chiral symmetry breaking shapes the mass and structure of quark-gluon systems remains unknown.

The dynamical nature of nuclear matter

Nuclear Matter Interactions and structures are inextricably mixed up



Observed properties such as mass and spin emerge out of the complex system



Ultimate goal Understand how matter at its most fundamental level is made

To reach goal precisely image quarks and gluons and their interactions

Advances in Nuclear Physics

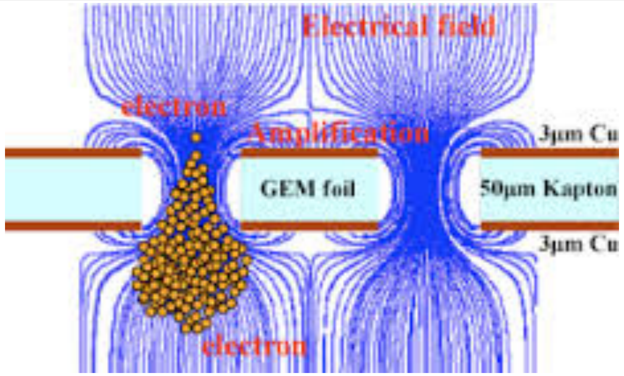
Theory of the strong interaction

$$
\begin{aligned}
\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\,\mathrm{d}y\,\mathrm{d}q_\mathrm{T}^2} = \frac{4\pi^2\alpha^2}{9Q^2 s} \sum_{j,j_A,j_B} e_j^2 \int \frac{\mathrm{d}^2\boldsymbol{b}_\mathrm{T}}{(2\pi)^2} e^{i\boldsymbol{q}_\mathrm{T}\cdot\boldsymbol{b}_\mathrm{T}} \\
\times \int_{x_A}^{1} \frac{\mathrm{d}\xi_A}{\xi_A} f_{j_A/A}(\xi_A;\mu_{b_*})\, \tilde{C}^{\mathrm{CSS1,\ DY}}_{j/j_A}\left(\frac{x_A}{\xi_A}, b_*; \mu_{b_*}^2, \mu_{b_*}, C_2, a_s(\mu_{b_*})\right) \\
\times \int_{x_B}^{1} \frac{\mathrm{d}\xi_B}{\xi_B} f_{j_B/B}(\xi_B;\mu_{b_*})\, \tilde{C}^{\mathrm{CSS1,\ DY}}_{\bar{j}/j_B}\left(\frac{x_B}{\xi_B}, b_*; \mu_{b_*}^2, \mu_{b_*}, C_2, a_s(\mu_{b_*})\right) \\
\times \exp\left\{ -\int_{\mu_{b_*}^2}^{\mu_Q^2} \frac{\mathrm{d}\mu'^2}{\mu'^2} \left[A_{\mathrm{CSS1}}(a_s(\mu');C_1) \ln\left(\frac{\mu_Q^2}{\mu'^2}\right) + B_{\mathrm{CSS1,\ DY}}(a_s(\mu');C_1,C_2) \right] \right\} \\
\times \exp\left[-g^{\mathrm{CSS1}}_{j/A}(x_A, b_\mathrm{T}; b_{\max}) - g^{\mathrm{CSS1}}_{\bar{j}/B}(x_B, b_\mathrm{T}; b_{\max}) - g^{\mathrm{CSS1}}_{K}(b_\mathrm{T}; b_{\max}) \ln(Q^2/Q_0^2) \right] \\
+ \text{suppressed corrections.}
\end{aligned}
$$

Quantumchromo-dynamics (QCD)

Accelerator technologies

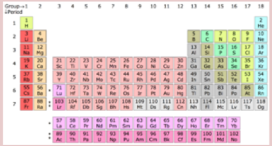
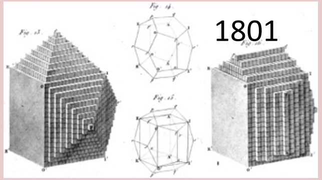
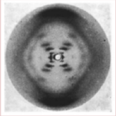
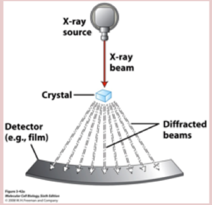
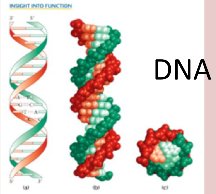
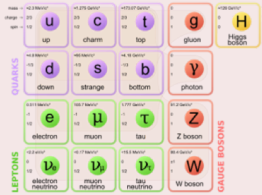
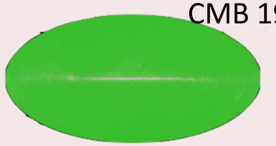
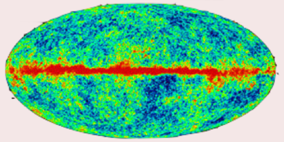
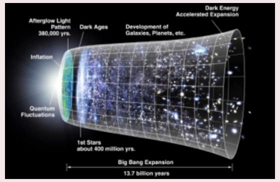
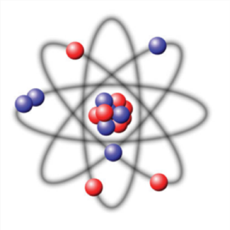
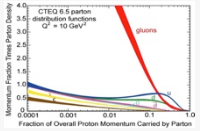
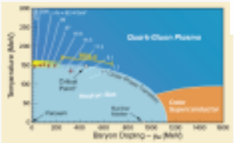
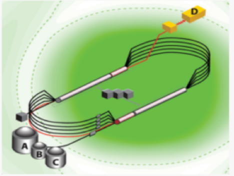
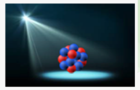
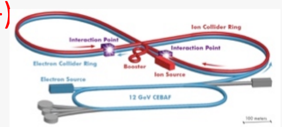
Detector technologies



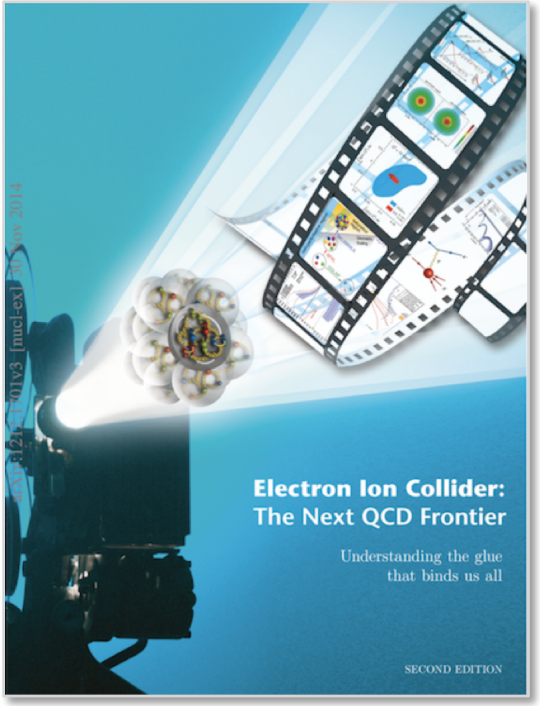
Computer technologies

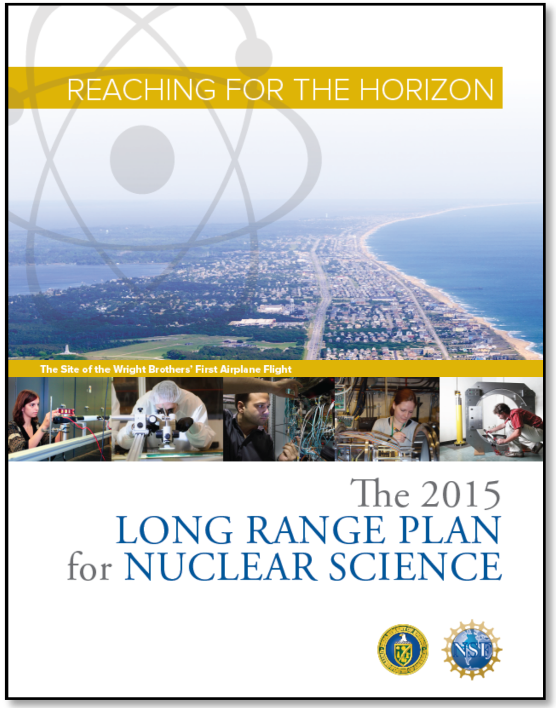
Steady advances in all of these areas mean that →

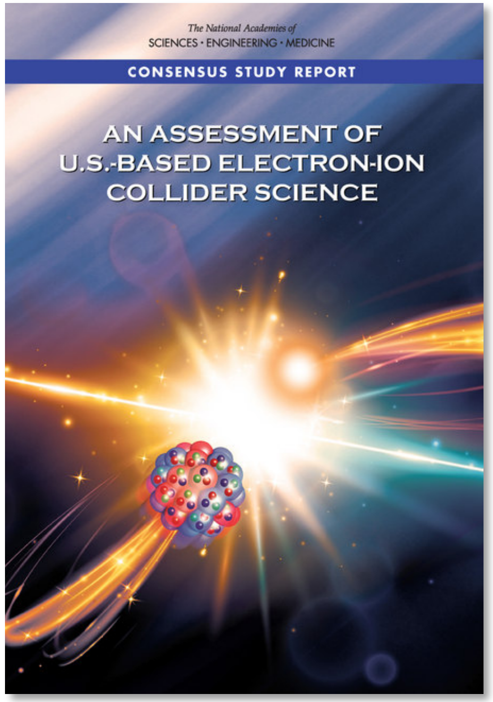
EIC: A new frontier in science

Dynamical System	Fundamental Knowns	Unknowns	Breakthrough Structure Probes (Date)	New Sciences, New Frontiers
Solids 	Electromagnetism Atoms 	Structure 1801 	X-ray Diffraction (~1920)  	Solid state physics Molecular biology DNA 
Universe 	General Relativity Standard Model 	Quantum Gravity, Dark matter, Dark energy. Structure CMB 1965 	Large Scale Surveys CMB Probes (~2000)  	Precision Observational Cosmology 
Nuclei and Nucleons 	Perturbative QCD Quarks and Gluons $\mathcal{L}_{\rm QCD} = \overline{\psi}(i\partial\!\!\!/ - gA\!\!\!/)\psi - \frac{1}{2}\mathrm{tr}\, F_{\mu\nu}F^{\mu\nu}$	Non-perturbative QCD Structure 2017   	CEBAF12 (2018)   Electron-Ion Collider (2025+) 	Structure & Dynamics in QCD 

Why an Electron-Ion Collider?







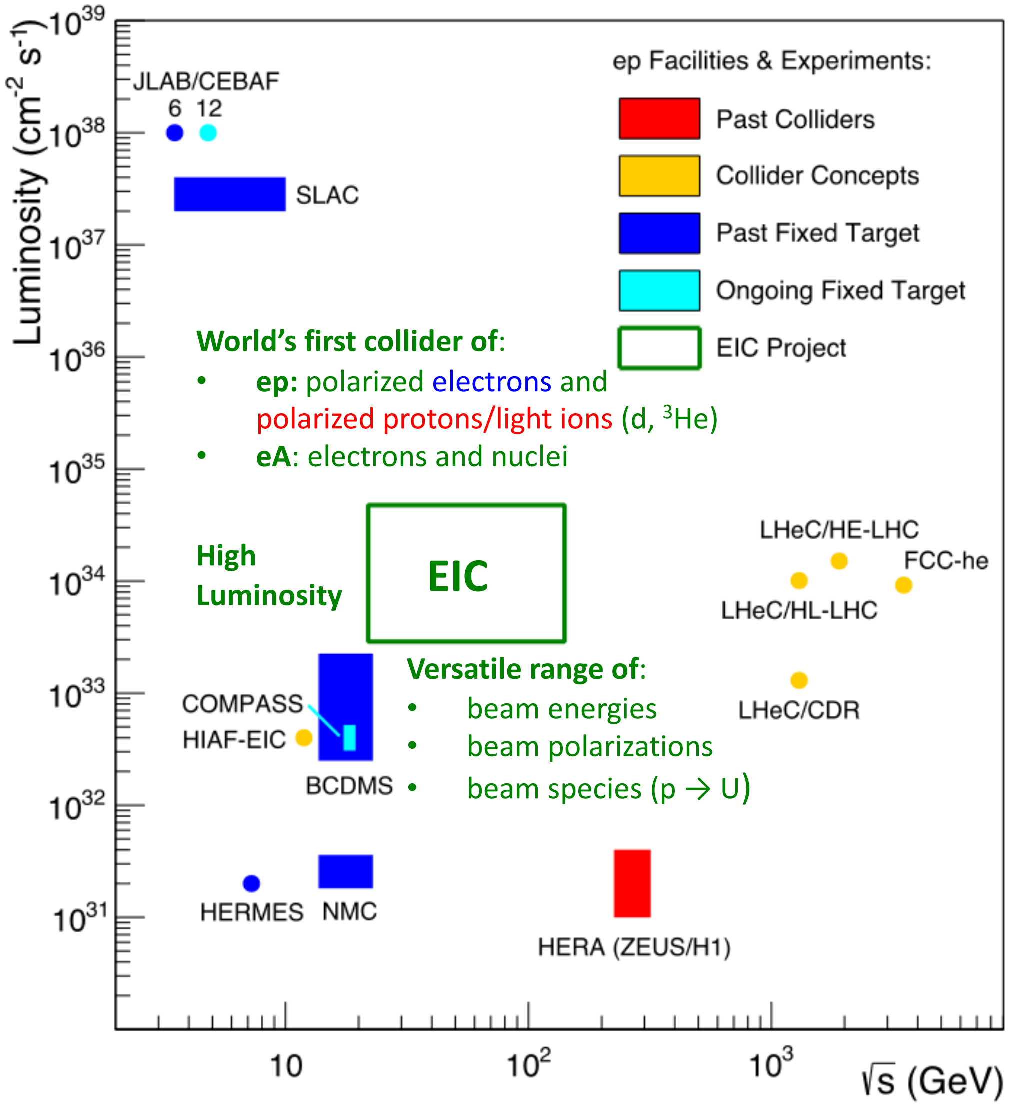
Right tool

- to precisely **image quarks and gluons** and their interactions
- to explore the new **QCD frontier of strong color fields in nuclei**
- to understand **how matter at its most fundamental level is made**.

Understanding of nuclear matter is **transformational**,
perhaps in an even more dramatic way than how the understanding of the atomic and molecular structure of matter led to new frontiers, new sciences and new technologies.

Jefferson Lab

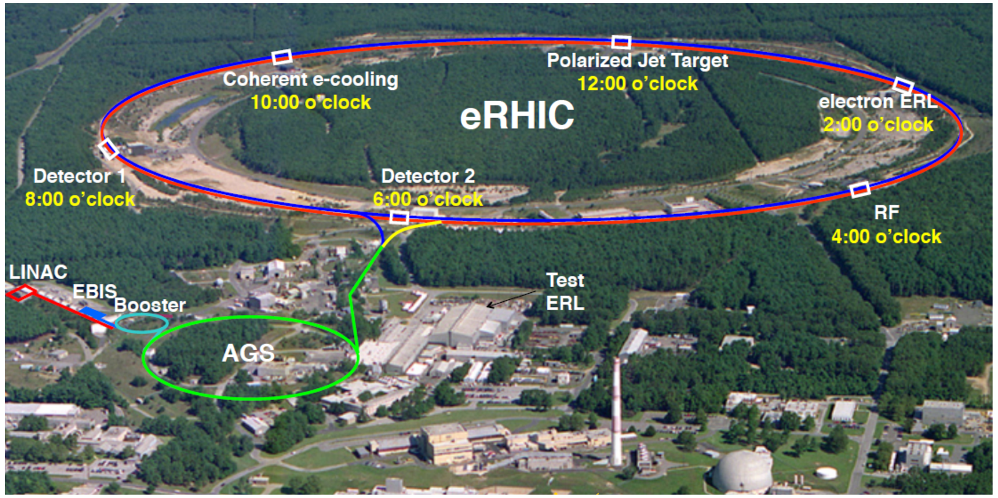
The Electron-Ion Collider: Frontier accelerator facility in the U.S.



Department of Energy

U.S. Department of Energy Selects Brookhaven National Laboratory to Host Major New Nuclear Physics Facility

JANUARY 9, 2020

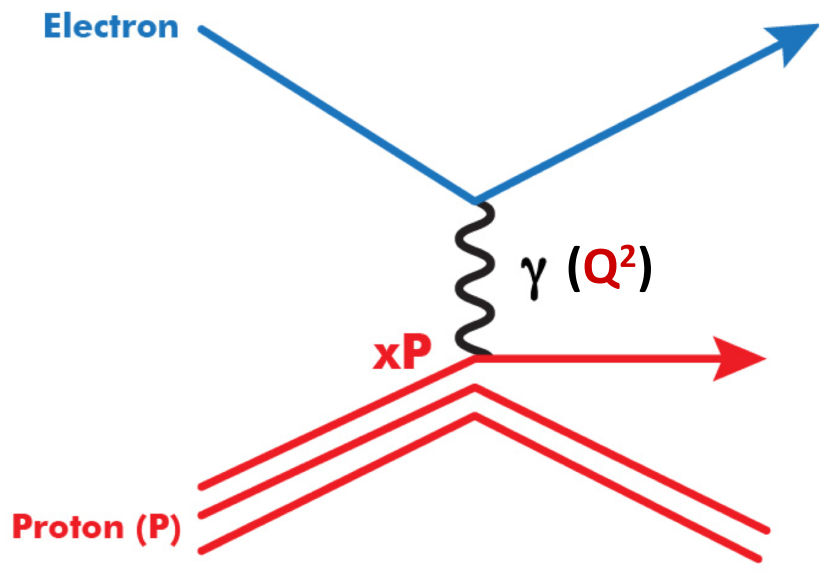


Thomas Jefferson National Accelerator Facility in Newport News, VA will be a major partner in realizing the EIC, and several other DOE laboratories are expected to contribute to EIC construction and to the groundbreaking nuclear physics research program that will be accomplished there.

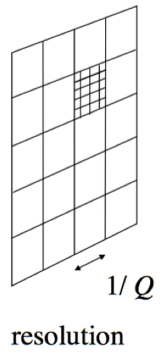
Detector design

General design considerations

Electron-Proton Scattering

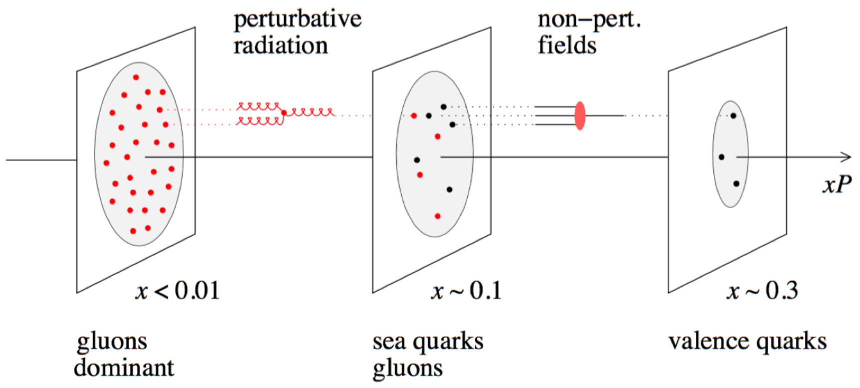


Ability to change Q^2 changes the resolution scale



$Q^2 = 400\ GeV^2$
=> $1/Q = 0.01$ fm

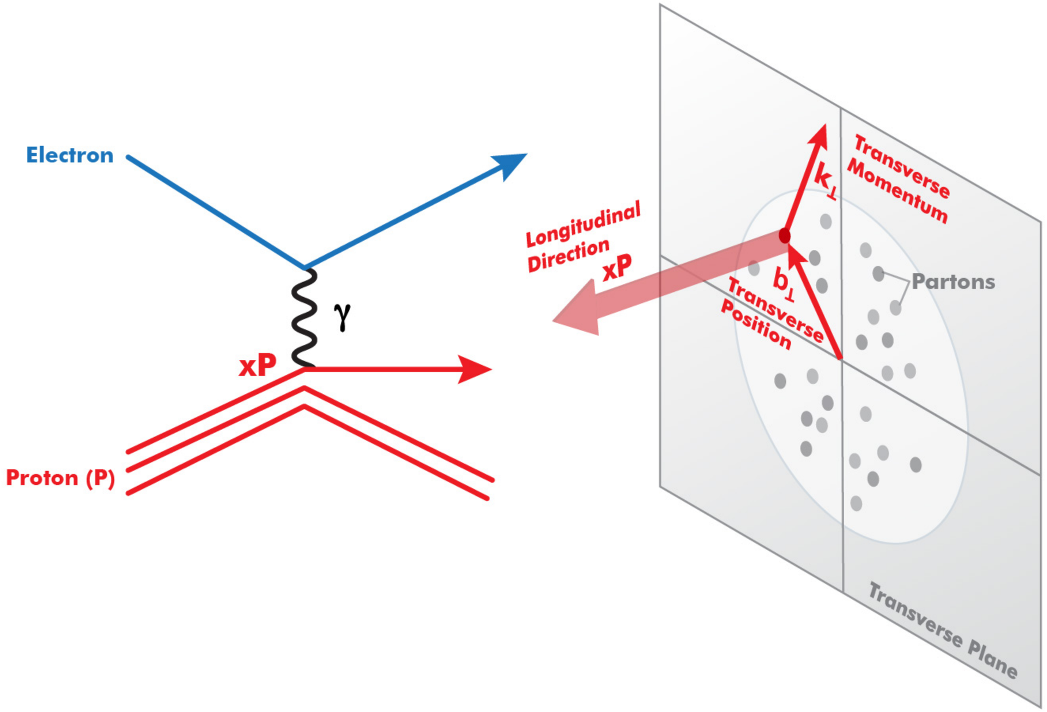
Ability to change **x** projects out different con-figurations where different dynamics dominate



Jefferson Lab

Mapping position and motion of quarks and gluons

Study nuclear matter **beyond longitudinal description** makes the **requirements for IR and detector design different** from all previous colliders including HERA.



3D imaging in space and momentum

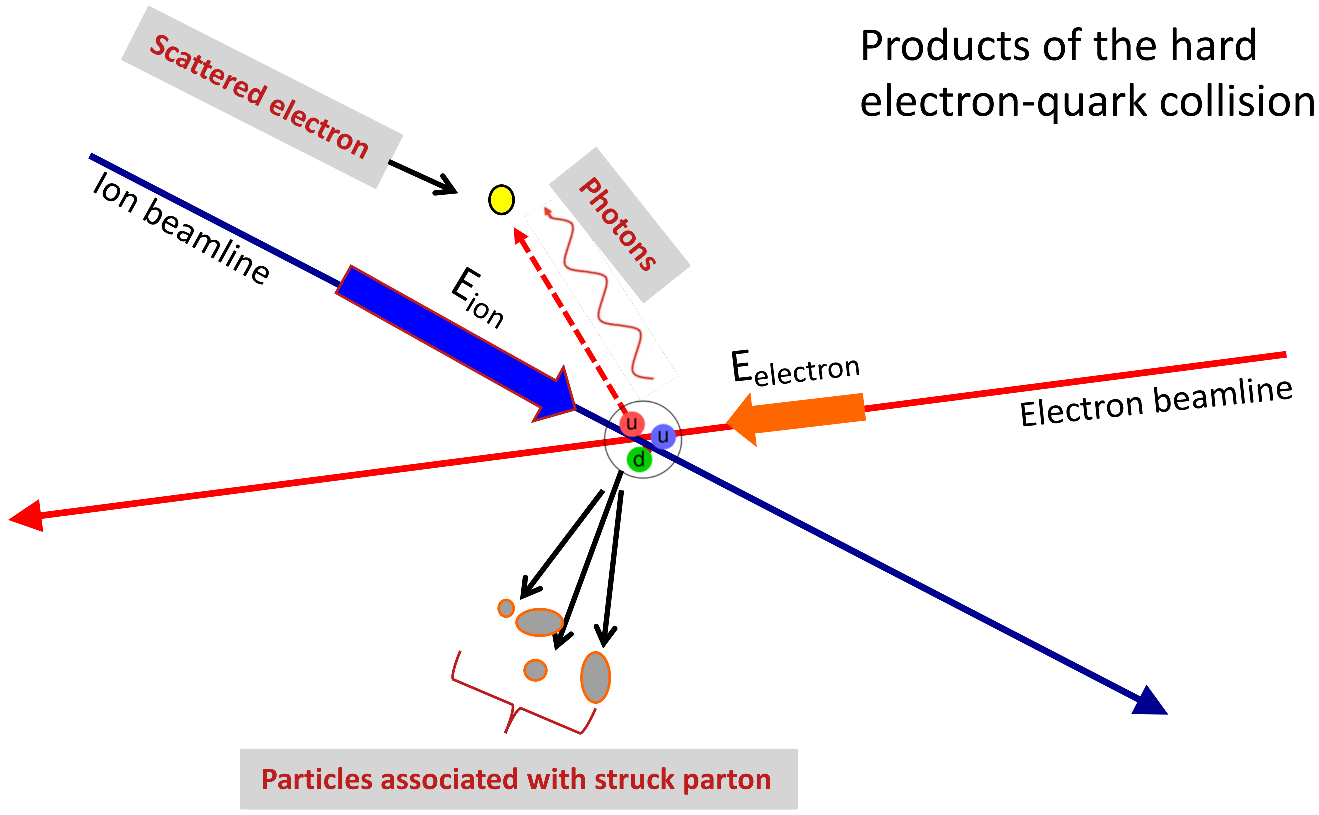
longitudinal structure (PDF)
+ transverse position Information (GPDs)
+ transverse momentum information (TMDs)

order of a few hundred MeV **measurement**

Jefferson Lab

Particle Identification

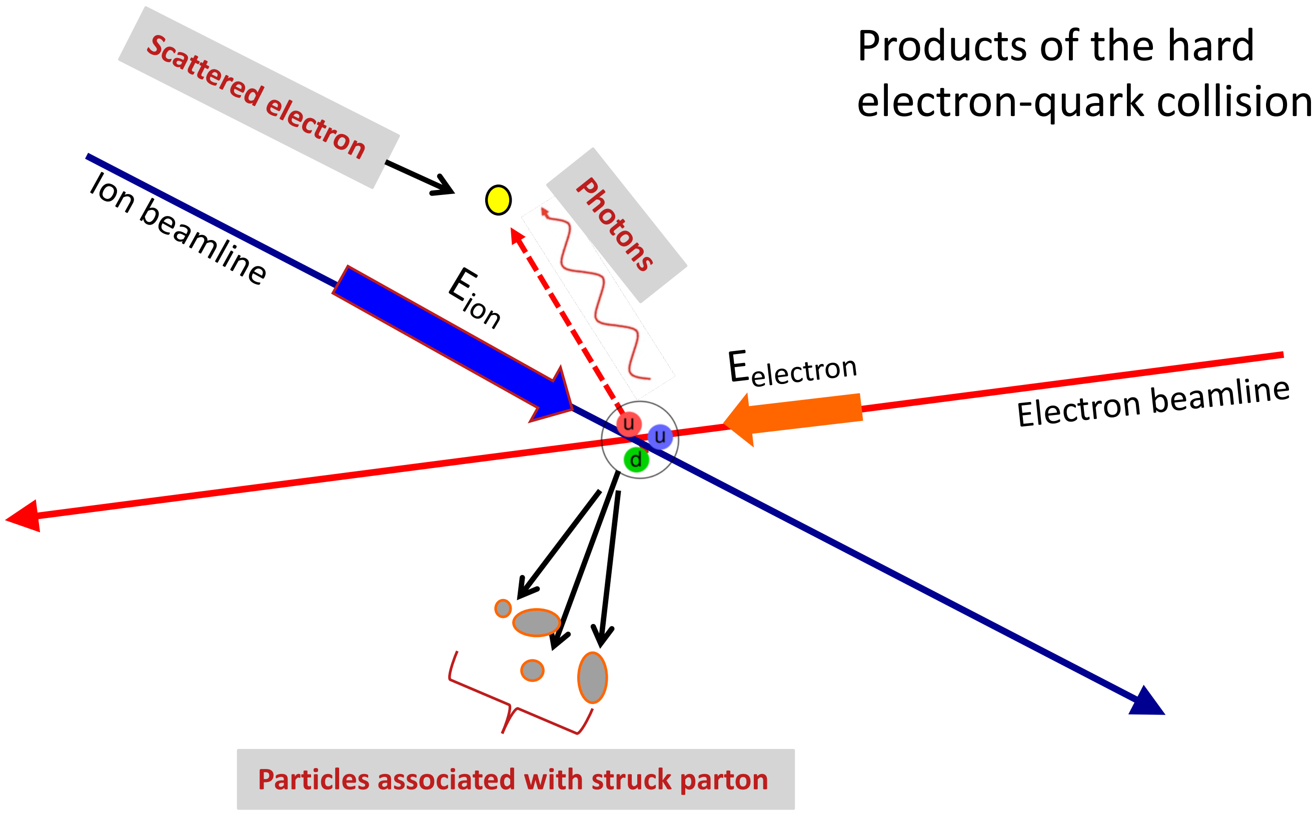
Products of the hard electron-quark collision



Transverse and flavor structure measurement of the nucleon and nuclei: The particles associated with struck parton must have its species identified and measured. **Particle ID much more important than at HERA** colliders.

Jefferson Lab

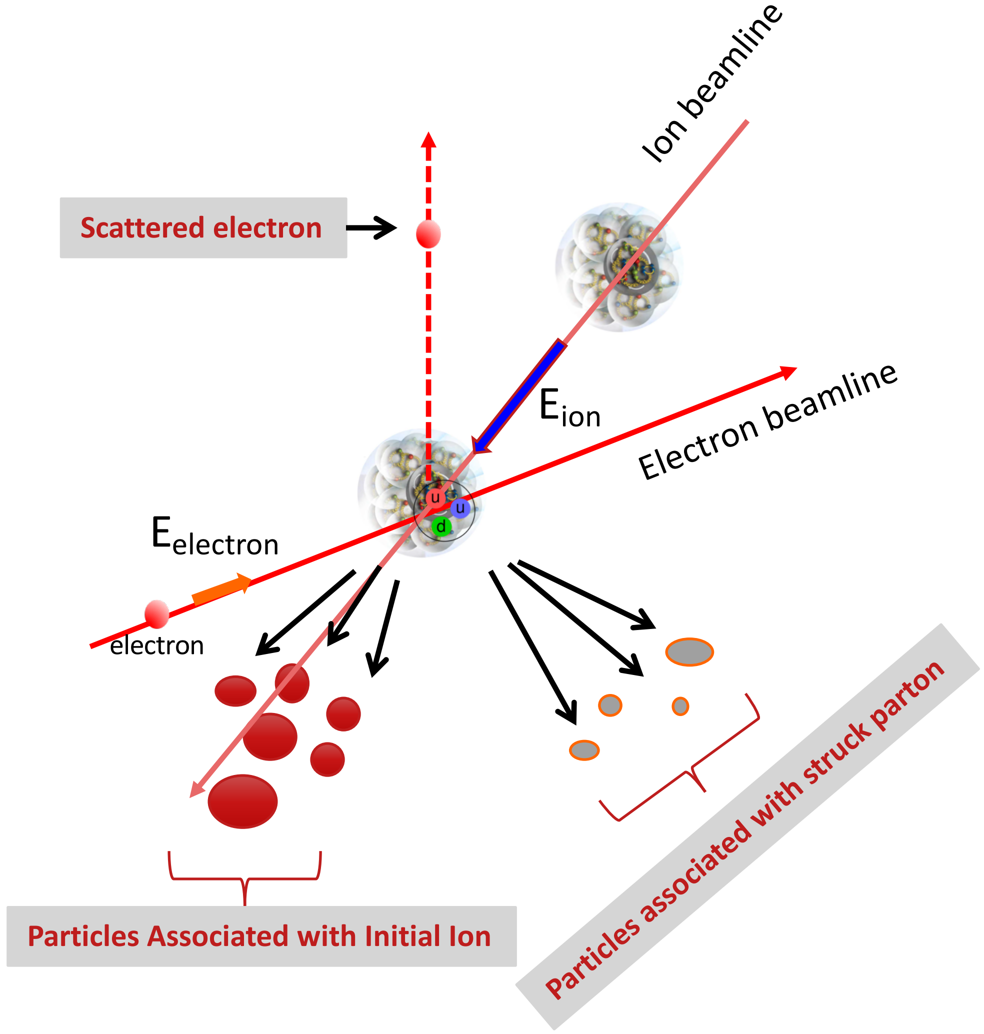
Final-state particles in the central rapidity



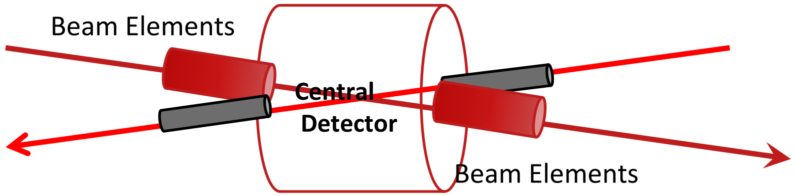
Products of the hard electron-quark collision

Asymmetric collision energies will boost the final state particles in the ion beam direction: **Detector requirements change as a function of rapidity.**

Final-state particles



The aim is to get **~100% acceptance** for all final state particles, and measure them with good resolution.

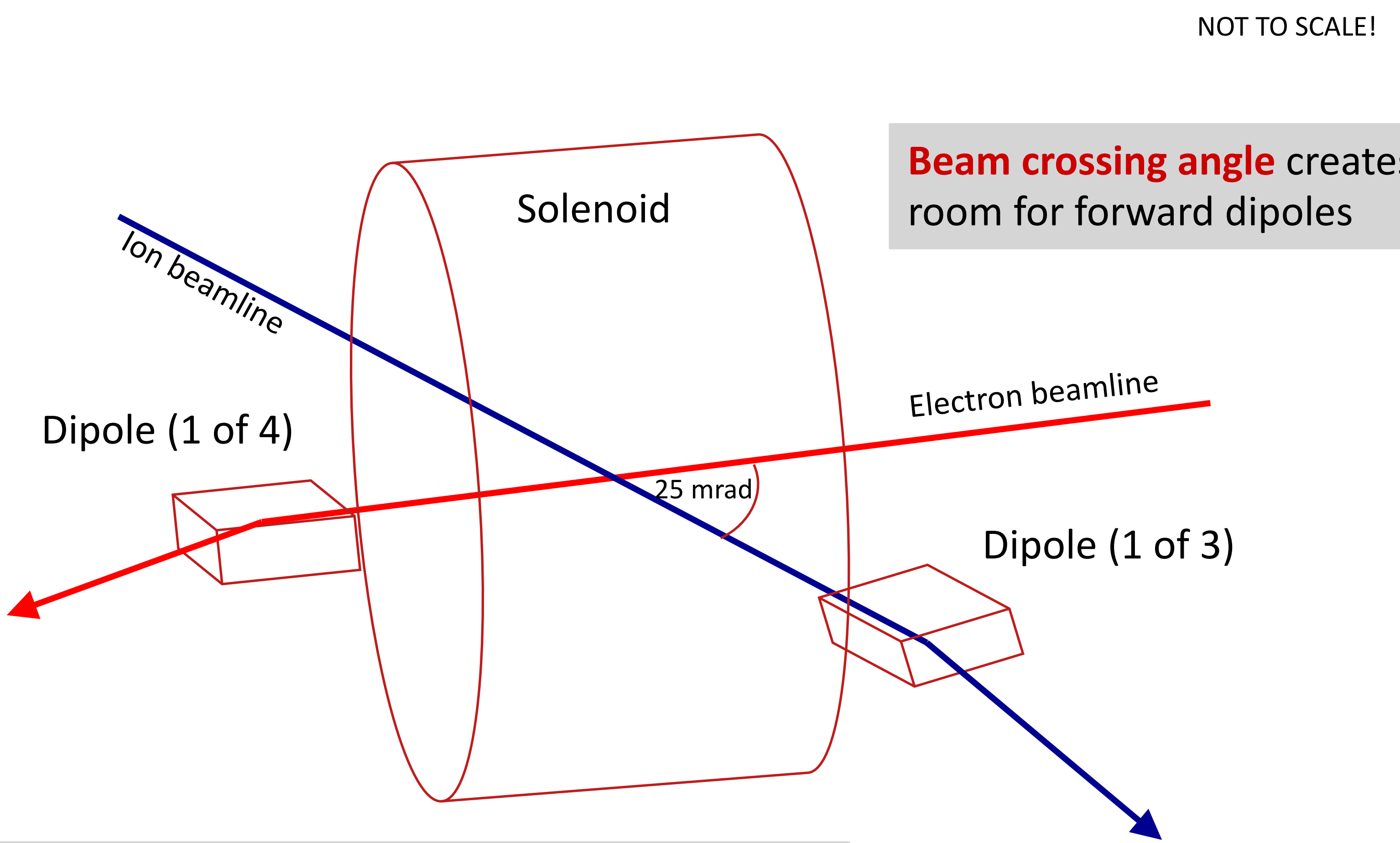


Experimental challenges:

- beam elements limit forward acceptance
- central Solenoid not effective for forward

Jefferson Lab

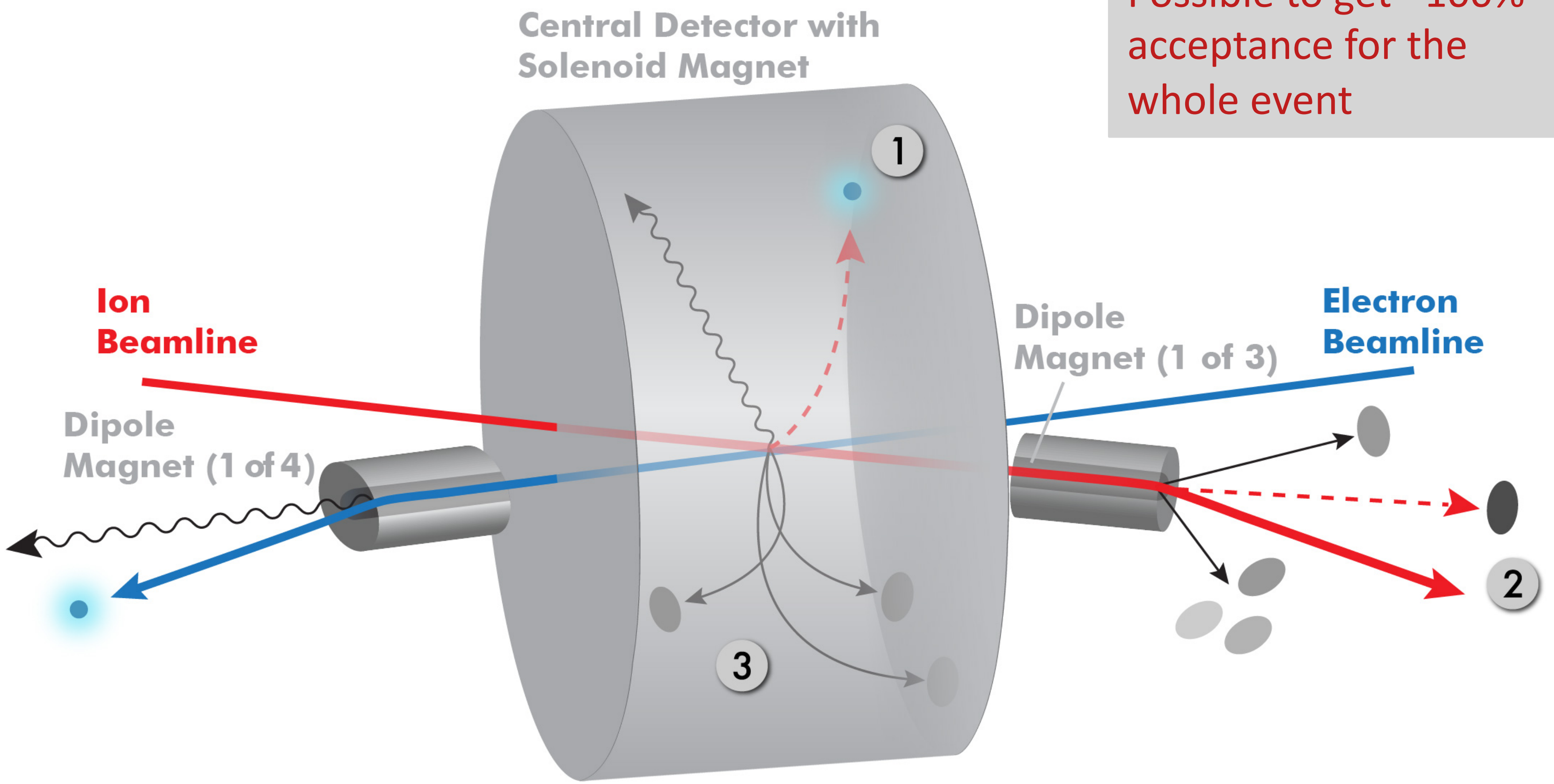
Interaction region concept



Beam crossing angle creates room for forward dipoles

Dipoles analyze the forward particles
and create space for detectors in the forward direction

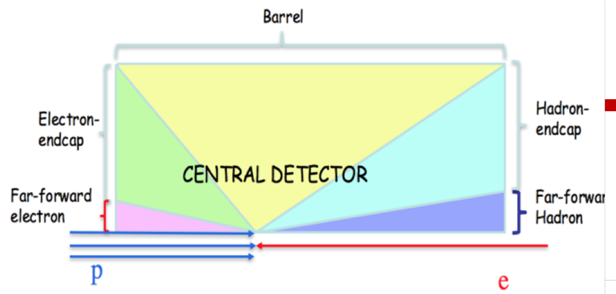
Interaction region concept

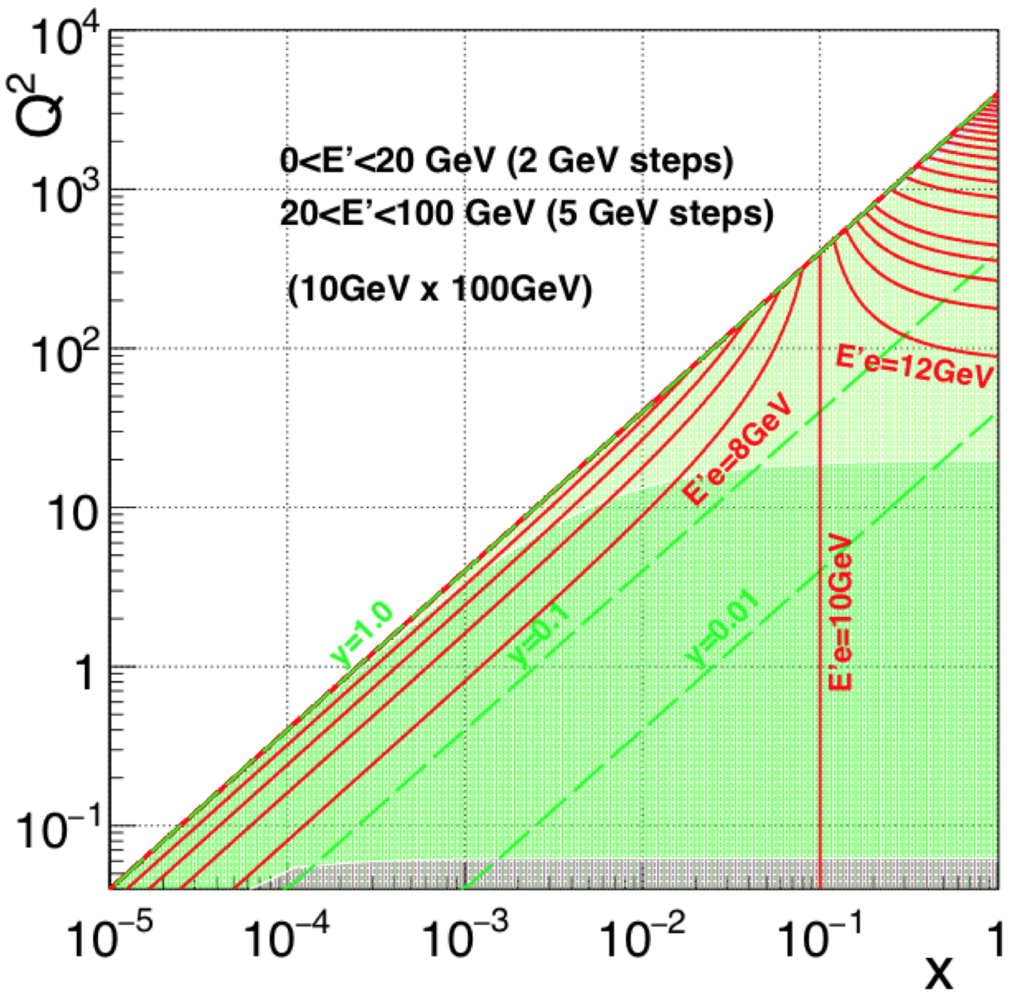


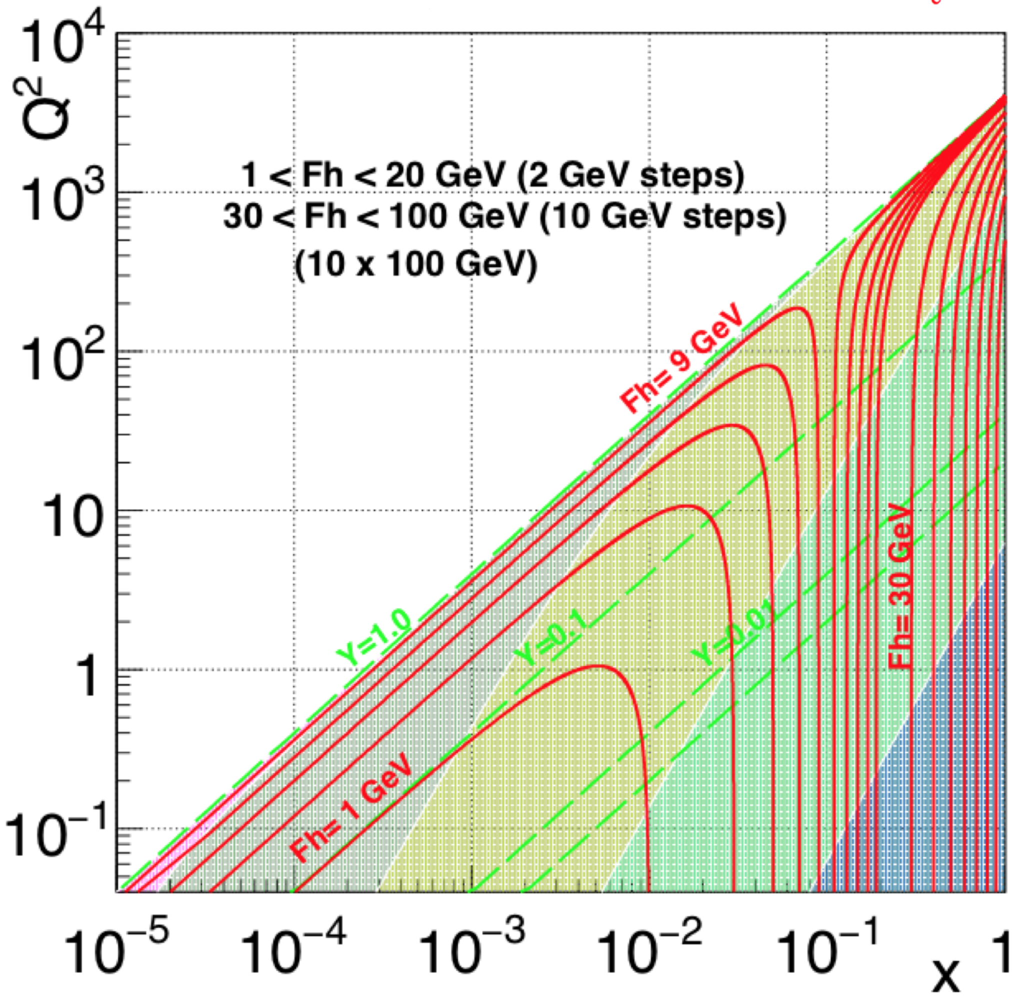
Possible to get ~100% acceptance for the whole event

Total acceptance detector (and IR)

Scattered lepton and jet in the EIC detector







Jefferson Lab

Detector concepts for the EIC

After a presentation on "Breakthroughs in Detector Technology", Ian Shipsey (Oxford) was asked about the role of software.

Anecdote

"***Software is the soul of the detector,***" Ian Shipsey replied in a poetic way and emphasized the **importance of great software for great science**. He added that we need to **work together**, on a global scale and with other fields, to achieve this goal.

Immediately after my overview

- Sylvester Joosten (ANL) will introduce the **ATHENA** detector concept.
- Joe Osborn (ORNL) will introduce the **ECCE** detector concept.

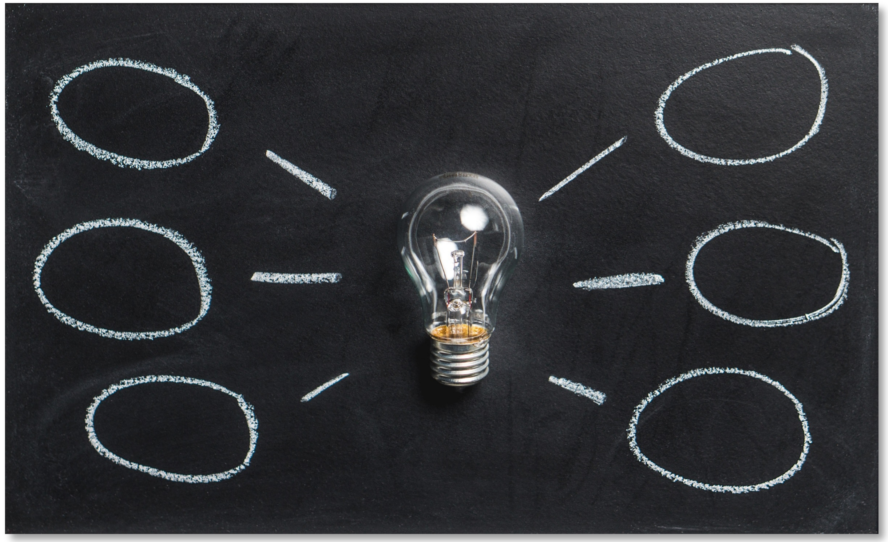
Streaming Readout

Seamless data processing from detector readout to analysis

Towards the next-generation research model in Nuclear Physics

Science & Industry remarkable advances in electronics, computing, and software over last decade

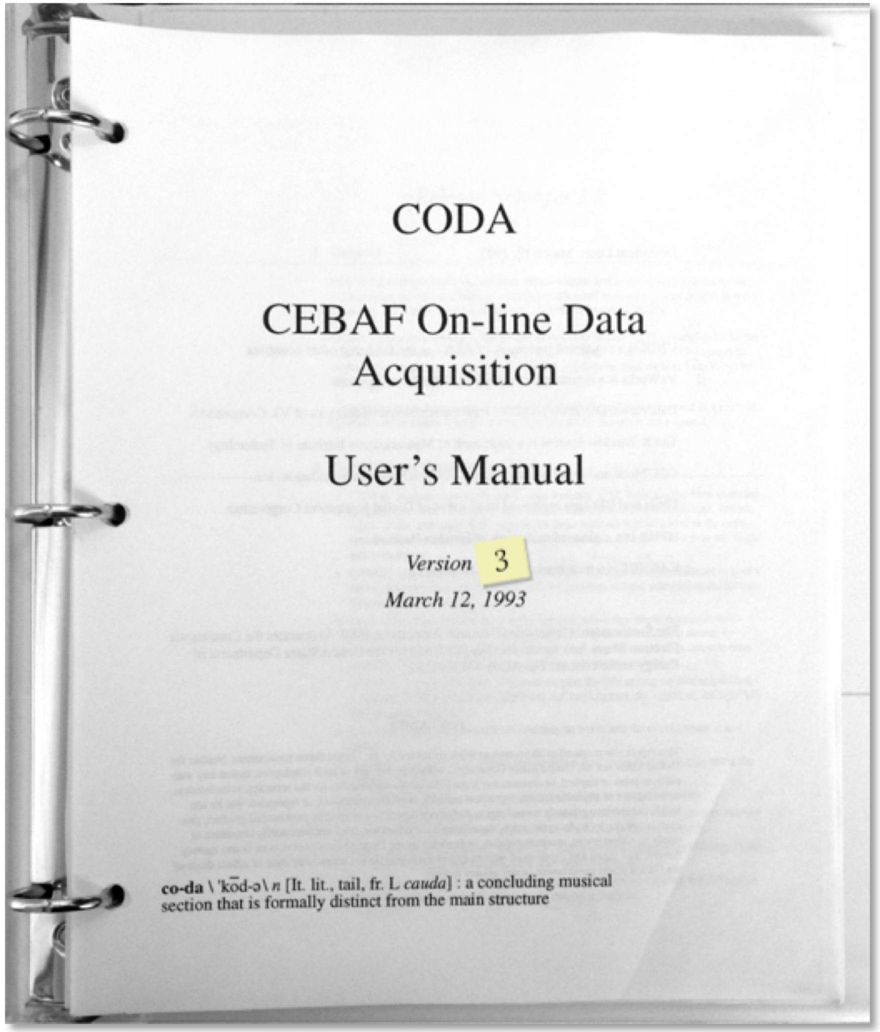
evolve & develop **Nuclear Physics research model** based on these advances



Roles of computing Data processing from data acquisition (DAQ) to analysis largely shaped by kinds of computing that has been available
Example Trigger-based readout systems

Advances in electronics, computing, and software Unique opportunity to think about new possibilities and paradigms
Example Streaming readout systems

CODA: Trigger-based readout system



Based upon assumptions in traditional DAQ design

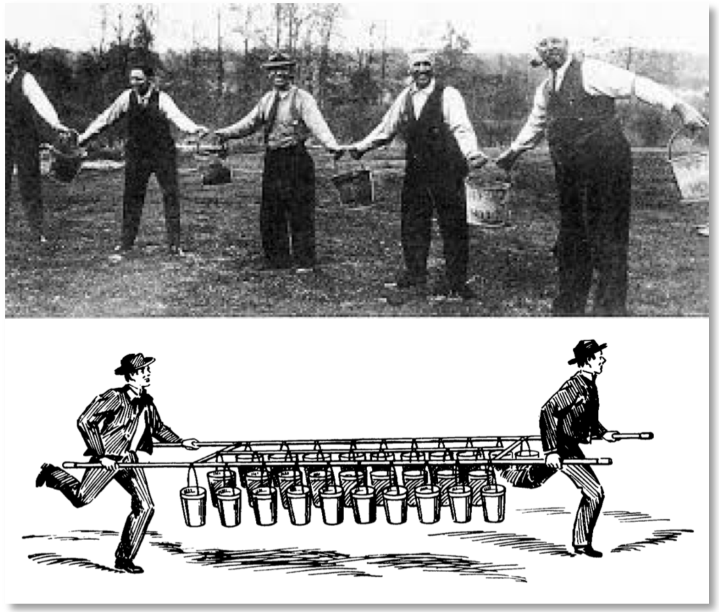
- The data rate from a detector is impossible to capture with an affordable data acquisition system without a trigger to reduce event rates.
- Even if the untriggered data rate could be captured, it would be impossible to store.
- Even if it could be stored the full dataset would represent a data volume that would require impractically large computing resources to process.

With computing advances Assumptions no longer valid

Limitation in trigger-based readout systems

- bias to low-energy particles
- do not deal well with event-pileup
- not an ideal for complex, general-purpose detectors

Alternative readout mode: Streaming



Traditional trigger-based readout

- data is digitized into buffers
- trigger starts readout
- parts of events are transported to an event builder where they are assembled into events
- at each stage the flow of data is controlled by *back pressure*
- data is organized sequentially by events

Streaming readout

- data is read continuously from all channels
- validation checks at source reject noise and suppress empty channels
- data then flows unimpeded in parallel channels to storage or a local compute resource
- data flow is controlled at source
- data is organized in multiple dimensions by channel and time

Streaming Readout: Trigger-less data acquisition

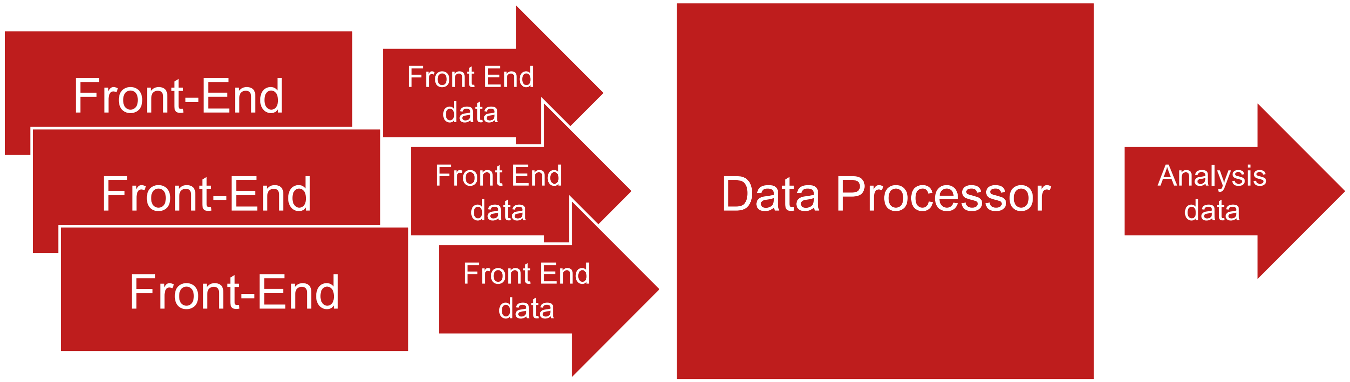
Definition of Streaming Readout

- Data is digitized at a fixed rate with thresholds and zero suppression applied locally.
- Data is read out in continuous parallel streams that are encoded with information about when and where the data was taken.
- Event building, filtering, monitoring, and other processing is deferred until the data is at rest in tiered storage.

Advantages of Streaming Readout

- simplification of readout (no custom trigger hardware and firmware)
- trigger-less readout:
  - beneficial for experiments that are limited by event-pileup or overlapping signals from different events
  - beam time is expensive so data mining or taking generic datasets shared between experiments is becoming popular: loosen triggers to store as much as possible
- opportunity to streamline workflows
- take advantage of other emerging technologies

Streaming Readout and (near) real-time processing



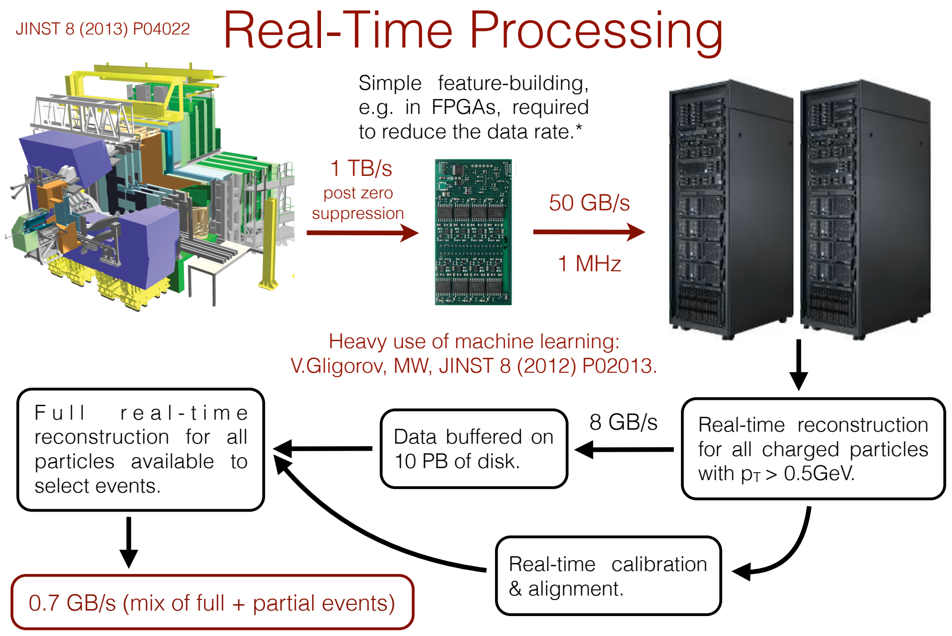
Data Processor

- assembles the data into events
- outputs data suitable for final analysis (**Analysis data**)

Features

- ideal for AI
- automated calibration in (near) real time
- automated alignment in (near) real time
- reconstruction in (near) real time
- event filtering into analysis streams based on full event information
- automated anomaly detection
- responsive detectors (conscious experiment)

LHCb Example



*LHCb will move to a triggerless-readout system for LHC Run 3 (2021-2023), and process 5 TB/s in real time on the CPU farm.

Integration of DAQ, analysis and theory to optimize physics reach

Research model with seamless data processing from DAQ to data analysis

- blurring of online and offline analysis evolves into near real-time analysis
- not about building the best detector but the best detector that fully supports:
  - streaming readout
  - fast algorithms for alignment, calibration, and reconstruction

Artificial Intelligence

Accelerating simulations

The role of A.I. in simulations

Lesson learned High-precision QCD measurements require high-precision simulations

Statistical accuracy for precise hypothesis testing

- up to trillion of simulated events required (HL-LHC)
- often computationally intensive, in particular calorimeter simulations

Common alternatives

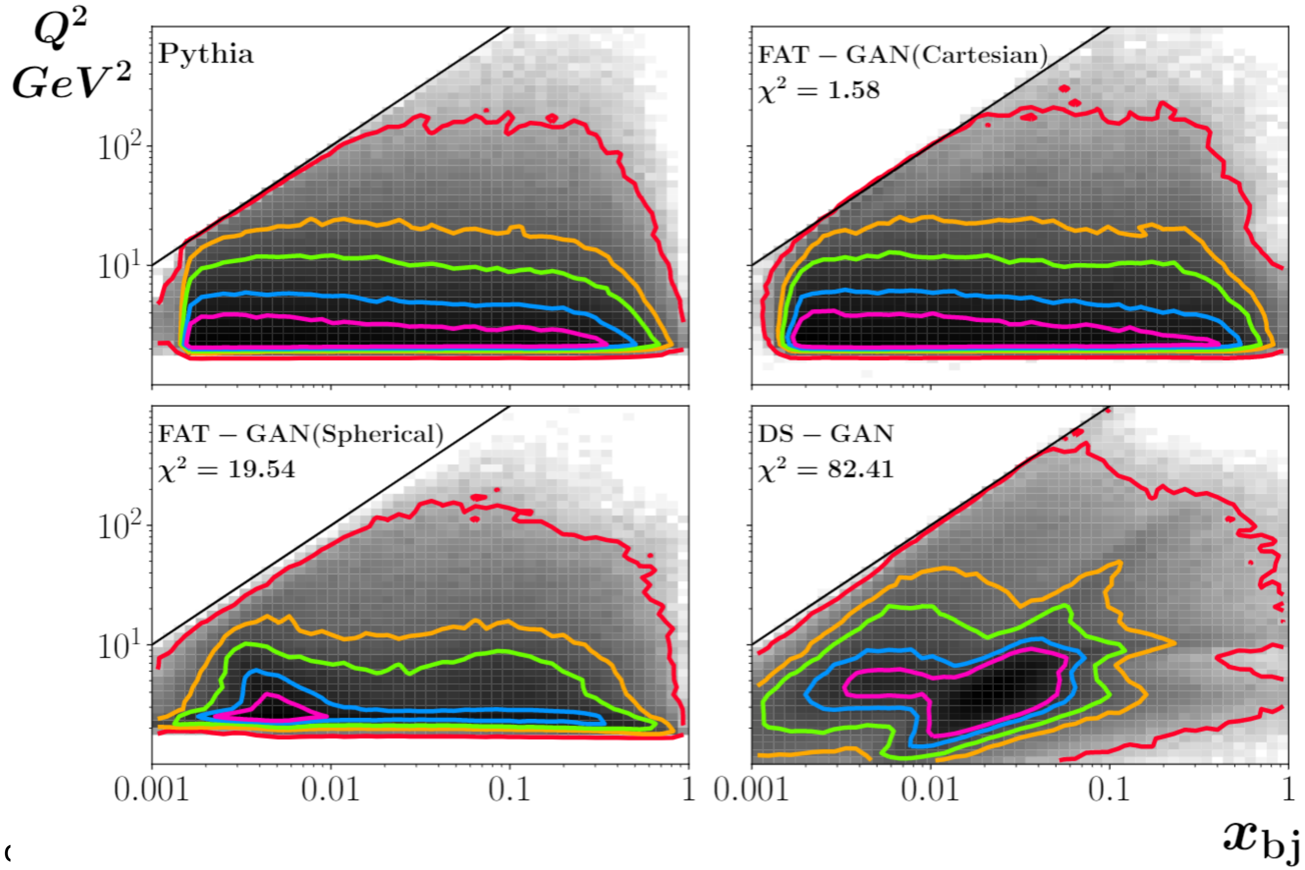
- fast simulations with computationally efficient approximations, e.g., parameterizations or look-up tables
- **still** insufficient accuracy for high-precision measurements

Promising alternatives

- fast generative models, e.g., GANs or VAEs
- A.I. driven design, e.g., Bayesian optimization

Empirically Trained Hadronic Event Regenerator (ETHER) Yaohang Li (ODU)

- LDRD project at Jefferson Lab: theorists interpolate across many different experiments, in a way that they could never do by stitching all the experiments together
- currently: study GAN as a repository of the behavior of the theory as expressed in Pythia (later real data)
- working well for single beam energy and inclusive single electrons / single electron and pion
- varying beam energy facing difficulty (variational GAN based event generators)

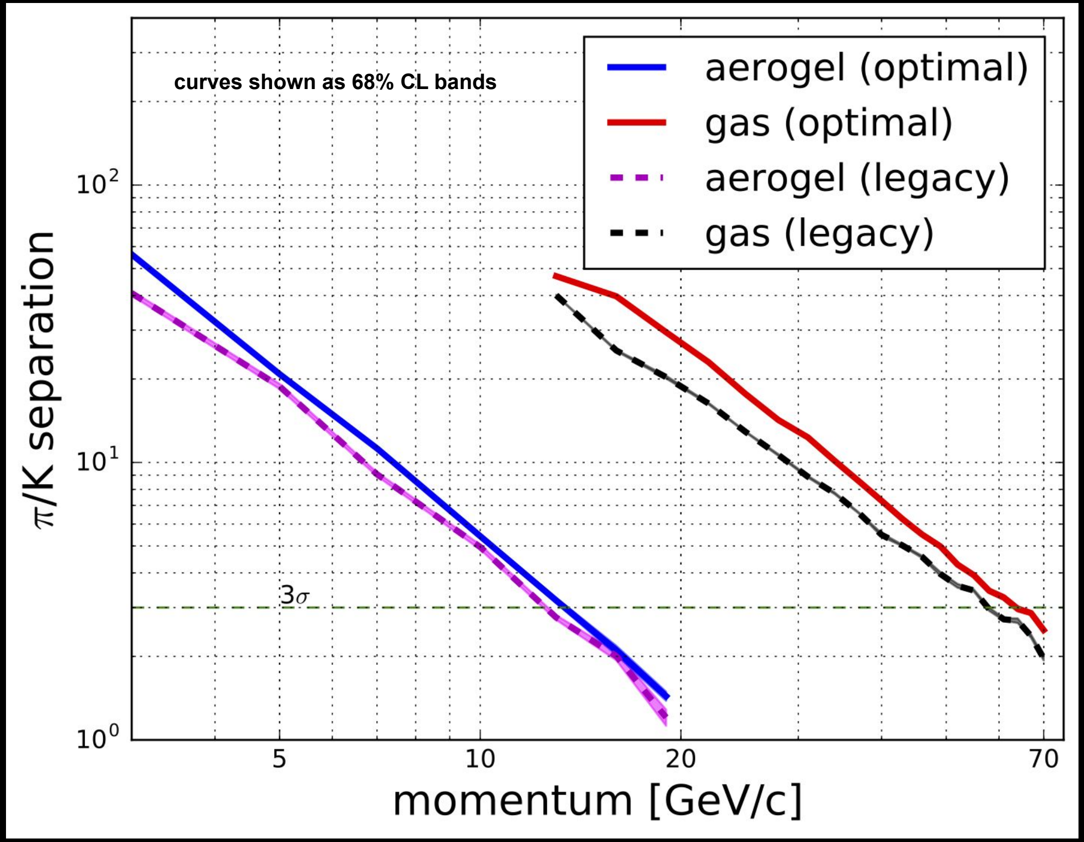


Jefferson Lab

AI-optimized Detector Design

Cristiano Fanelli (MIT, EIC^2)

- automated, highly parallelized, self consistent framework for detector design
- specific application for the dual-RICH of the future EIC has been shown
- statistically significant improvement w.r.t. baseline design found
- tested with O(20) parameters, ways to deal with O(100) parameters, possible to add cost



- Statistically significant Improvement in both parts.
- In particular in the gas region where the 5σ threshold shifted from 43 to 50 GeV/c and the 3σ one extended up to
- Notice that before this study we did not know "how well" the legacy design was performing.

CF et al. AI-optimized detector design for the future EIC *arXiv:1911.05797* (2019)

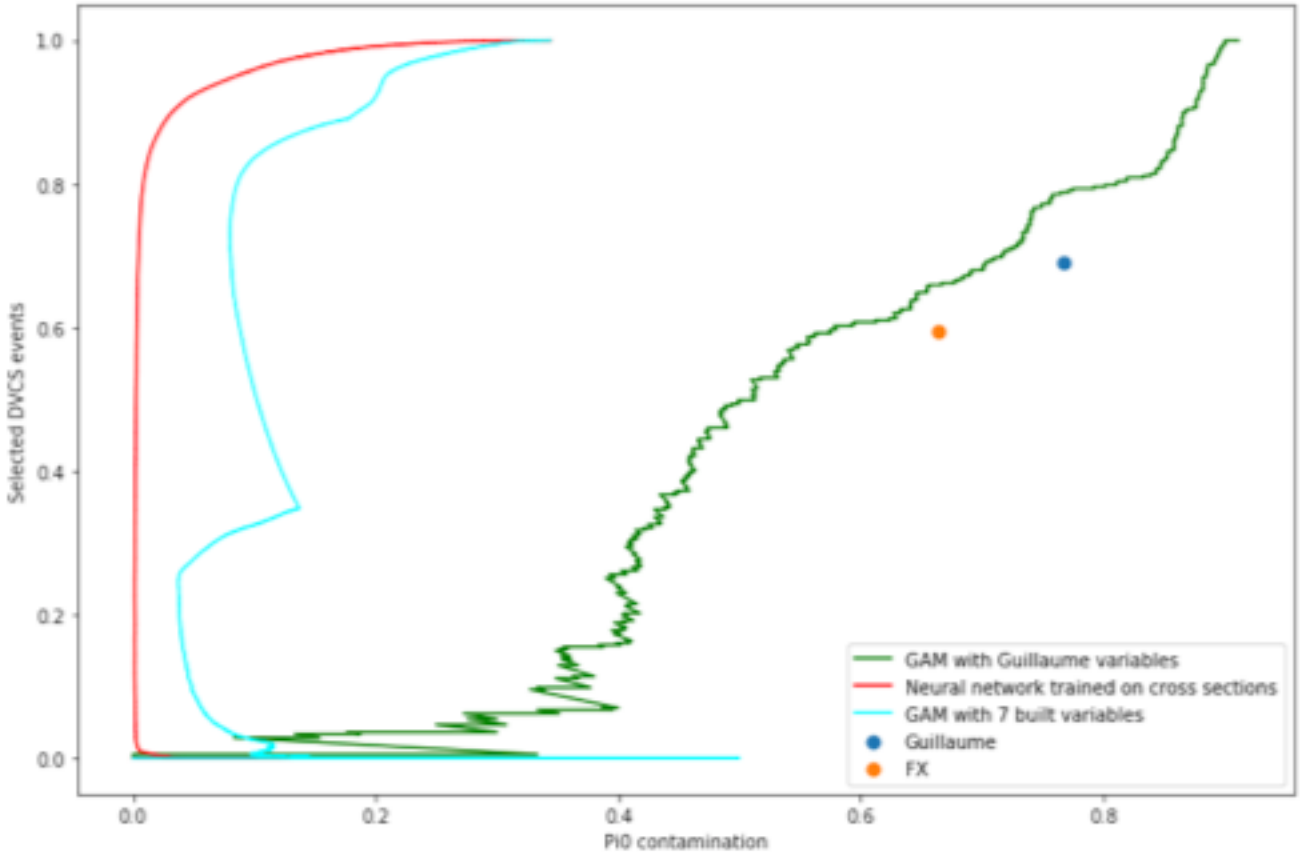
Semantic Interaction

- Exploratory platform to find High-D relationships
- Offload the hard stuff to ML, but let the expert drive
- Manipulate CINEMA thumbnails (eg 3D projections)

Visualization

- makes debugging models and code easier
- key component of discovery and communication
- better visualization tools can help build better models and analytic capabilities for A.I. / ML

- build a selector for DVCS events
- uses feature construction to get new discriminative variables
- implementation in generalized additive models (GAMs)
- GAM makes better use of the correlations between the variables than other approaches, out-performs conventional approaches on efficiency and purity

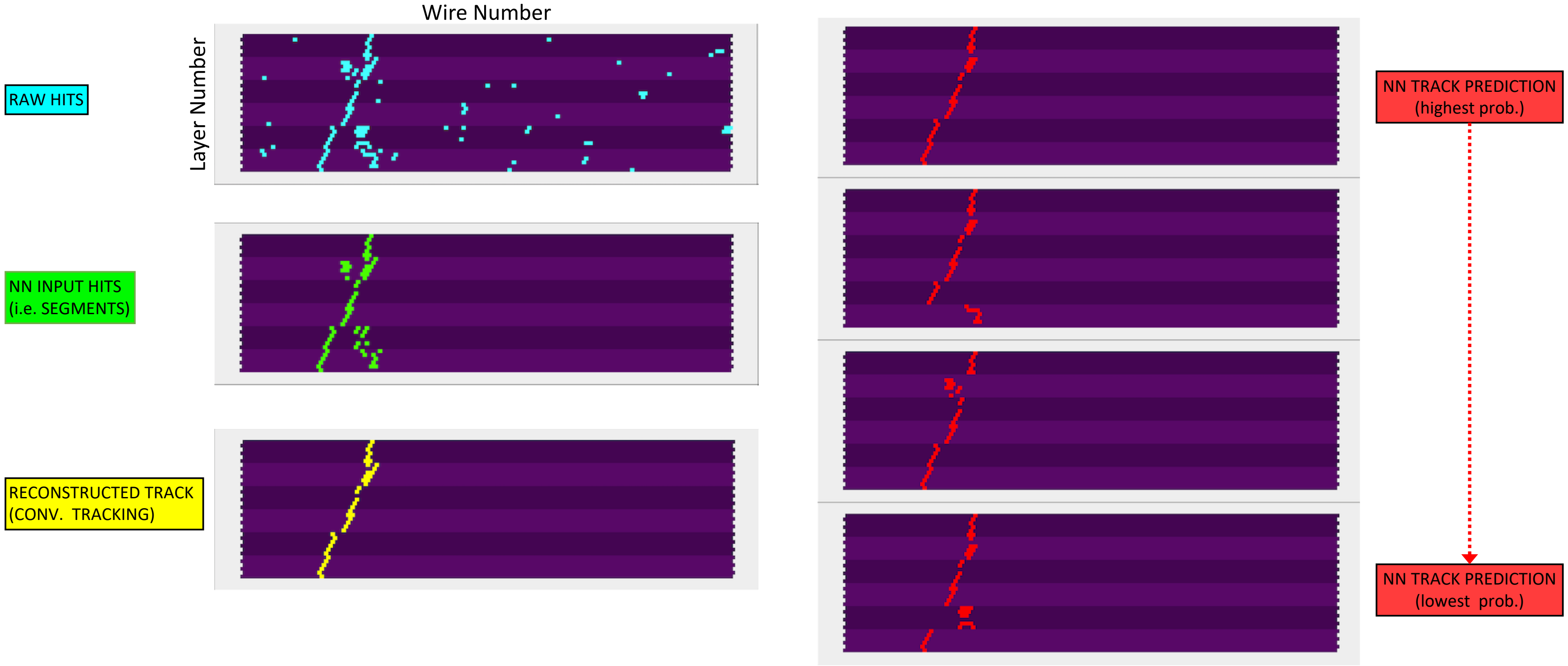


Open questions

- fair comparisons of the different methods
- objectively assess interpretability
- how to apply to real data

CLAS12 Tracking with ML

- combinatorics in resolving ghost tracks, noise rejection takes considerable time:
  - AI-assisted tracking to speed it up
  - evaluate different NN approaches
- track reconstruction is ~5x faster using NN for segment finding
- NN tracking finds tracks missed with conventional tracking, in presence of high background. But also the reverse happens. Studying tracking efficiency is underway (currently ~99.5% accuracy).



identify which track segments are consistent with being on-track and use these NN-found segments for seeding recons

ML to reconstruct DIS kinematics

Abdullah Farhat (ODU, EIC^2)

Method	Required Measurements	Strengths	Limitations
Electron	$E_{e'}, \theta_{e'}$	Precise	Sensitive to QED radiation
Jacques-Blondel	$\delta_{\mathcal{H}}, P_{T,\mathcal{H}}$	Resistant to QED radiation	Needs precise energy measurements
Double-Angle	$\theta_{e'}, \gamma_{\mathcal{H}}$	Does not need precise energy measurements	Poor resolution at low x, low Q^2

Bin	Events	Q^2 (GeV^2)	x	x RMSE		Q^2 RMSE	
1	114606	80 – 160	0.0024 – 0.010	NN: 0.0040 JB: 0.0042	EL: 0.0029 **DA: 0.0012**	NN: 22.705 JB: 204.39	**EL: 14.810** DA: 20.753
2	65501	160 – 320	0.0024 – 0.010	NN: 0.0049 JB: 0.0053	EL: 0.0014 **DA: 0.0013**	NN: 35.068 JB: 405.88	**EL: 29.609** DA: 36.397
3	74382	320 – 640	0.01 – 0.05	**NN: 0.0053** JB: 0.0086	EL: 0.0226 DA: 0.0063	**NN: 60.198** JB: 311.52	EL: 64.426 DA: 82.069
4	47055	640 – 1280	0.01 – 0.05	**NN: 0.0046** JB: 0.0103	EL: 0.0061 DA: 0.0047	**NN: 96.406** JB: 792.58	EL: 105.55 DA: 151.91
5	60684	1280 – 2560	0.025 – 0.150	**NN: 0.0102** JB: 0.0194	EL: 0.0262 DA: 0.0154	**NN: 195.70** JB: 1012.1	EL: 216.84 DA: 283.20
6	46242	2560 – 5120	0.05 – 0.25	**NN: 0.0154** JB: 0.0303	EL: 0.0333 DA: 0.0249	**NN: 410.11** JB: 1694.9	EL: 435.00 DA: 509.29
7	47380	5120 – 10240	0.06 – 0.40	**NN: 0.0197** JB: 0.0452	EL: 0.0358 DA: 0.0327	**NN: 712.45** JB: 3368.6	EL: 745.37 DA: 831.62
8	28507	10240 – 20480	0.10 – 0.6	**NN: 0.0288** JB: 0.0791	EL: 0.0454 DA: 0.0433	**NN: 1553.4** JB: 7096.9	EL: 1660.8 DA: 1796.4

- reconstruct kinematic variables x and Q^2 at collider via ML
- using ZEUS MC at HERA

- still working on the low kinematic range, outperforms conventional methods elsewhere
- did consider dividing the network into several for the different regions of the detector, for now decided to work with a single network covering the full detector

Jefferson Lab

AI: Multidisciplinary approach

- interplay between **Mathematics**, **Computer Science**, and **NP**
  - computer scientists need problems to solve
  - NP problems give insights into research in computer science and mathematics
  - great opportunity for education

- **related to** in NP (and HEP) need closer connection between experiment and theory

- **A.I./ML research**
  - scientific, systematic approach to applying A.I. / ML approaches to NP problems
  - activation functions, network design particular to NP applications
  - building efficient networks no more complex than necessary
  - NP analysis:
    - want to extract information from **all** the data and find correlations / common features
    - key difference with respect to HEP
  - **need to trust A.I. / ML and AI**
    - drive for *explainable AI* and *uncertainty quantification*
    - human interaction could be applied with great benefit to better understand the requirements and dynamics of such criteria in the NP domain
    - debatable whether *explainable* is a useful criterion for a ML model. We don't have the words for theories we haven't discovered yet

AI: Data science is about data

- **reference datasets for A.I. / ML development in NP**
  - always an issue to get access to big real datasets
    - amount of training data required often unknown
    - often two orders more data for simulations / training required than data
  - important to cultivate ML development
  - but, always difficult to understand from outside the experiment what the data means
    - common struggle for analysis preservation
    - project by the library community to address open data: Open Science Framework
- question to NP community: **Can we as a group figure out what datasets to ask for?**
  - the data was paid for by the DOE in the first place, after all
  - we have to ask, it's not going to magically just appear on the web
  - we need data to make progress
- **pose open challenges and run contests**
  - this has really worked to draw in new young people and new ideas
  - give prizes!
  - Can we think about benchmark problems?

Simulation challenges

Identify challenges coming over the next few years? pick our top problems?

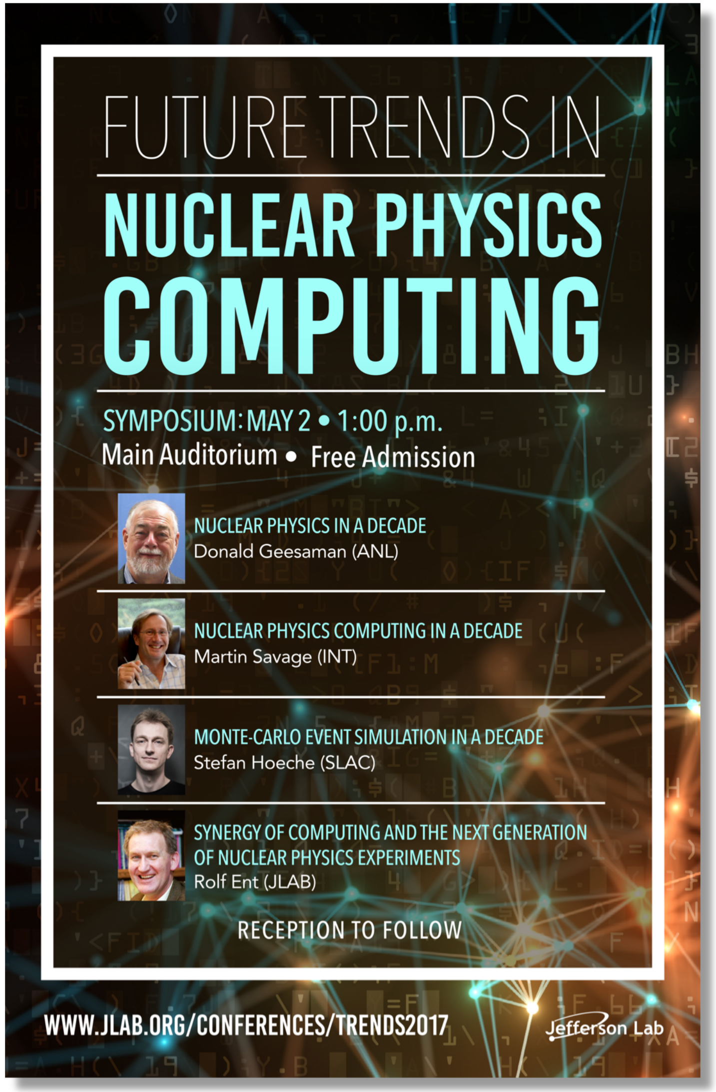
- **Accelerating simulations**
  - calorimeter in particular
  - but also PID, e.g., Cerenkov detectors
- **A.I driven detector design** Bayesian optimization for EIC detector R&D
- **HPC utilization**
  - Experimental NP, HEP have few or no payloads appropriate to the LCF/Exascale which are accelerator based.
  - ML is the best near term prospect for using them effectively.
  - Can we find the ML payloads? Do they use substantial processing resources?
- **ML for event generators**
  - replace models with ML as we do in detector simulations (e.g., LUND string model)

Vision for EIC Software

“The purpose of computing is insight, not numbers.”

Richard Hamming (1962)

Future Trends in Nuclear Physics Computing



Donald Geesaman (ANL, former NSAC Chair) *"It will be **joint progress of theory and experiment** that moves us forward, not in one side alone"*

Martin Savage (INT) *"The next decade will be looked back upon as a **truly astonishing period in Nuclear Physics** and in our understanding of fundamental aspects of nature. This will be **made possible by advances in scientific computing** and in how the Nuclear Physics community organizes and collaborates, and how DOE and NSF supports this, to take full advantage of these advances."*

Computing trends and EIC Computing

EIC rates

- expected data rates similar to next phase LHCb
- not enormous rates creates opportunity for other initiatives

Future compatibility hardware and software

- **Exascale Computing** Most powerful future computers will likely be very different from the kind of computers currently used in Nuclear Physics.
- This requires a modular design with structures robust against likely changes in computing environment so that changes in underlying code can be handled without an entire overhaul of the structure.

Think out of the box

- The way analysis is done has been largely shaped by kinds of computing that has been available.
- Computing begins to grow in very different ways in the future, driven by very different forces than in the past (e.g., Exascale Computing Initiative).
- This is an unique opportunity for Nuclear Physics to think about new possibilities and paradigms that can and should arise.

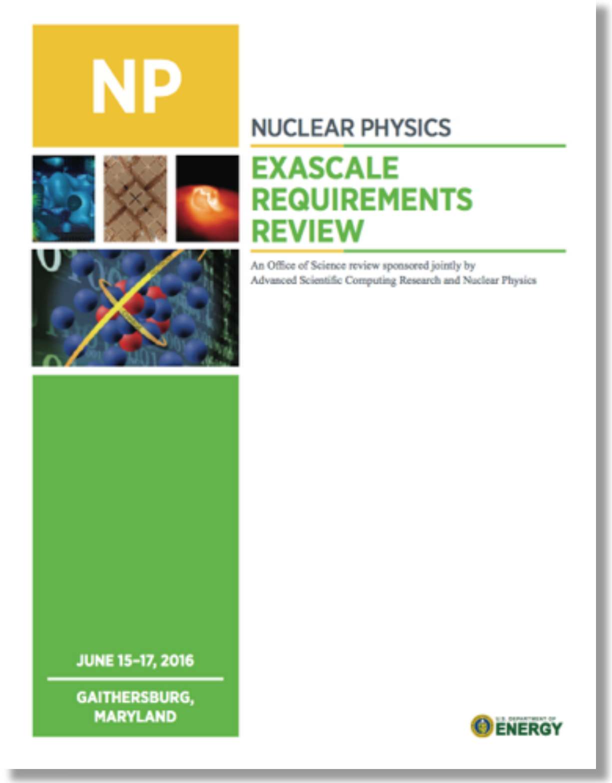
User centered design to enhance scientific productivity

- Engage wider community of physicists, whose primary interest is not computing, in software design to:
  - understand the user requirements first and foremost
  - make design decisions largely based on user requirements.

Implications of Exascale Computing

Past efforts in lattice QCD in collaboration with industry have driven development of new computing paradigms that benefit large scale computation. These capabilities underpin many important scientific challenges, e.g. studying climate and heat transport over the Earth.

The EIC will be the facility in the era of high precision QCD and the first Nuclear Physics facility in the **era of Exascale Computing**. This will affect the interplay of experiment, simulations, and theory profoundly and result in a new computing paradigm that can be applied to other fields of science and industry.



Petascale-capable systems at the beamline

- **unprecedented compute-detector integration**, extending work at LHCb
- requires fundamentally new and different algorithms
- computing model with AI at the DAQ and analysis level and a compute-detector integration to deliver **analysis-ready data from the DAQ system**:
  - responsive calibrations in real time
  - real-time event reconstruction and filtering
  - physics analysis in real time

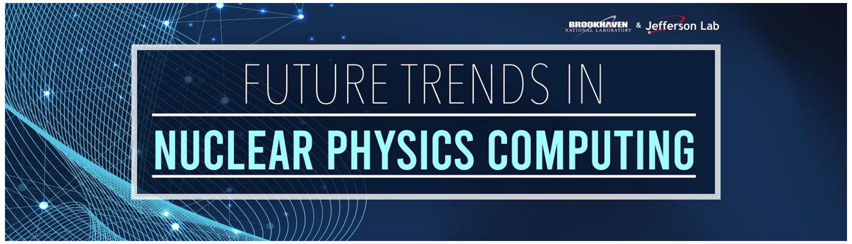
A similar approach would allow **accelerator operations** to use real-time simulations and artificial intelligence over operational parameters to tune the machine for performance.

Future Trends in Nuclear Physics Computing

Focus on the **Nuclear Physics Software & Computing community**

- Identify what is unique about our community
- Discuss how we could strengthen common efforts

Workshop discussion



Future Trends in Nuclear Physics Computing

Meeting Notes

Timetable

This is the live meeting notes document for the Future Trends in Nuclear Physics Computing Workshop held on September 29 - October 1, 2020. This workshop, the third of the series (previous editions were in 2017 and 2016), focuses on the Nuclear Physics Software & Computing community itself. Goals for the workshop are to identify what is unique about our community, find ways to strengthen common efforts, and chart a path for Software & Computing in Nuclear Physics for the next ten years.

We meet for four hours each day in a time window chosen to be as inclusive as possible for participants around the world. Substantial discussion time is included in the agenda, and session conveners will keep speakers to time in order to preserve the discussion time. This google doc will be used in advance to give the discussions structure and focus, as well as during the workshop itself to moderate and record the discussion and gather input from all participants, and after the workshop as the basis for summarizing and report writing. Editing is on, and all participants are encouraged to contribute in all phases.

Each day has a theme. In advance of the workshop, questions and discussion points for each day will be gathered here to guide a moderated common discussion following the talks. A short discussion period will follow each talk to address questions specific to the talk. The content prepared in advance will be augmented during the presentations and discussions.

A brief synopsis of the previous day will be part of an intro talk on days two and three.

The workshop will conclude with a short summary, but summarizing and report writing proper will proceed after the workshop. All participants are welcome and encouraged to join the meeting

Live Notes (26 (!) pages)

Draft (30 pages)

Unique software challenges for Nuclear Physics

Scientific Problem Space

- Focus on non-perturbative QCD phenomena
- MC event generators for spin-dependent measurement, including novel QCD phenomena (e.g., GPDs, TMDs, Wigner functions)
- **Analyses considering large number of signal events simultaneously** (or multiple times)
    - **Contrary** to separating a few events from a large number of background events
        - **Example** Search of rare events with novel topologies
    - **Example** complexity of multi-dimensional, strongly correlated relationships among data (e.g., GPDs, TMDs, Wigner functions)
    - **Example** high-precision results which require complex analyses to control systematic uncertainties
    - Require unique software and computing strategies
- Relatively smaller size of experiments goes along with shorter experimental life cycles and faster changes in scientific goals

Small Group Size

- Collaboration size in average smaller in NP than in HEP
- Tendency for everyone *“doing their own thing”*
    - Larger experiments, individual analyses can be numerous and quite different from another, with a small team on each top.
- Non-unified approach has inhibited progress in the field in the past.
- Transition to experiments with larger data size and more complex analyses
- Old culture cannot effectively address problems of scale of future experiments
- Relatively smaller group size asks for careful planning and design of the software effort: mix of in-house development, adoption of outside packages, and the choice of appropriate scale throughout.
- Challenge in finding the right balance.

Common scientific software

Common Scientific Software – The keys to success

- **The team is the most important** Do not separate development and operations, both ACTS and Rucio benefited from experience with developing and operating a worse software package, crucial experience. Developers keen to use modern software paradigms, open-source and open-minded, proactively searching out best practice and adopting it.
- **The project** Clear, well-focused short-term goals are important, grounded in real-world deliverables. Aligned with the long-term plan of building something sustainable and designed to be used by outside collaborators.
- **The management** Accept that the long-view takes longer to deliver the short-term product, manage expectations of the collaboration and funders to ensure the team have sufficient time and space to succeed.

Scientific software careers need support

- Recognition, encouragement and reward: need to make software citations a priority
- Career paths of Research Software Engineers (RSE) need to be supported and not only at the labs

NP software - should NP participate in HSF or build its own organization?

- Pros and cons, the balance of opinion favored NP participation in HSF. HSF is a do-ocracy, active participation will yield the biggest rewards.
- NP often has small groups developing solutions in-house, work with this reality.

Community-wide effort

Software Working Group

EIC User Group (EICUG)

EICUG Software Working Group

127 members

Electron-Ion Collider User Group

The world's most powerful microscope for studying the "glue" that binds the building blocks of visible matter.

HOME | JOIN EICUG | SCIENCE | ORGANIZATION | CALENDAR | SOFTWARE | DOCUMENTS | YELLOW REPORT | MEDIA | ADMIN

Home » EIC Software

EIC Software

Software Working Group

The Software Working Group (SWG) is open to all members of the EICUG to work on EICUG related software tasks. It communicates via its mailing list and organizes regular online and in-person meetings that enable broad and active participation from within the EICUG as a whole.

The SWG has participated in the call for Expressions of Interest (EoI) and intends to carry its EoI for Software forward as a living document that will evolve towards a work plan for the SWG, setting priorities for the next years and goals for the next decade.

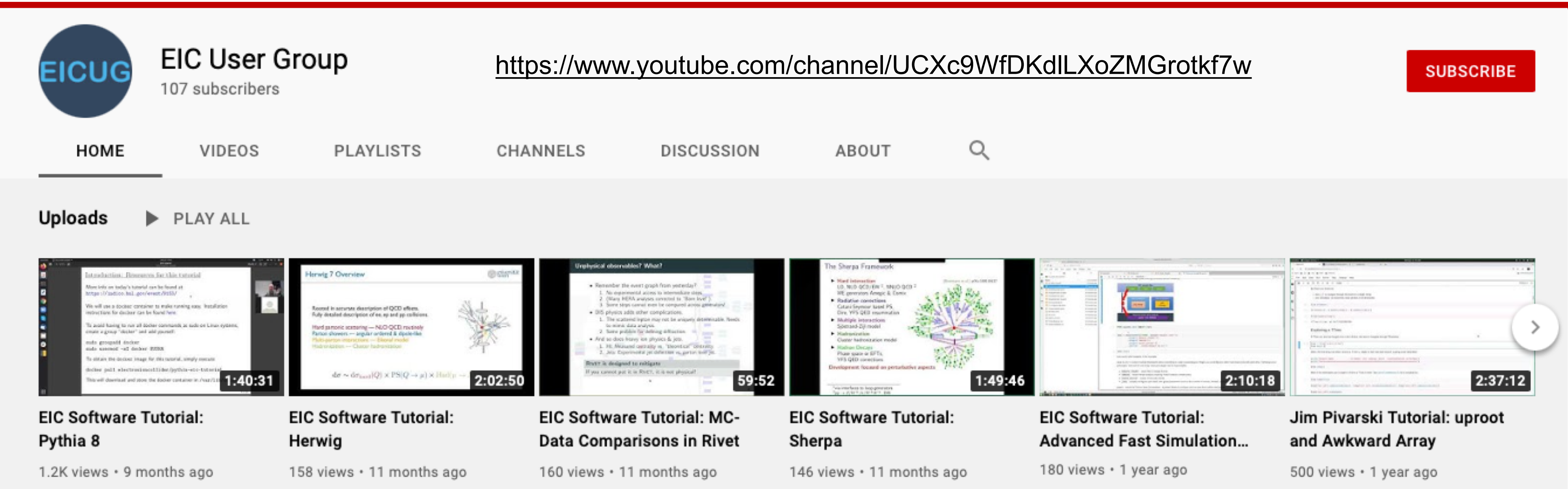
In addition to that, it will support the work on the simulation efforts for the collaboration proposals for detectors at the EIC.

For questions about the Software Working Group, please contact the conveners (Andrea Bressan (Trieste), Markus Diefenthaler (JLab), and Torre Wenaus (BNL)).

Important links

Mailing list	eicug-software@eicug.org (subscribe via Google Group)
EIC organization on GitHub	https://github.com/eic
Website	https://eic.github.io

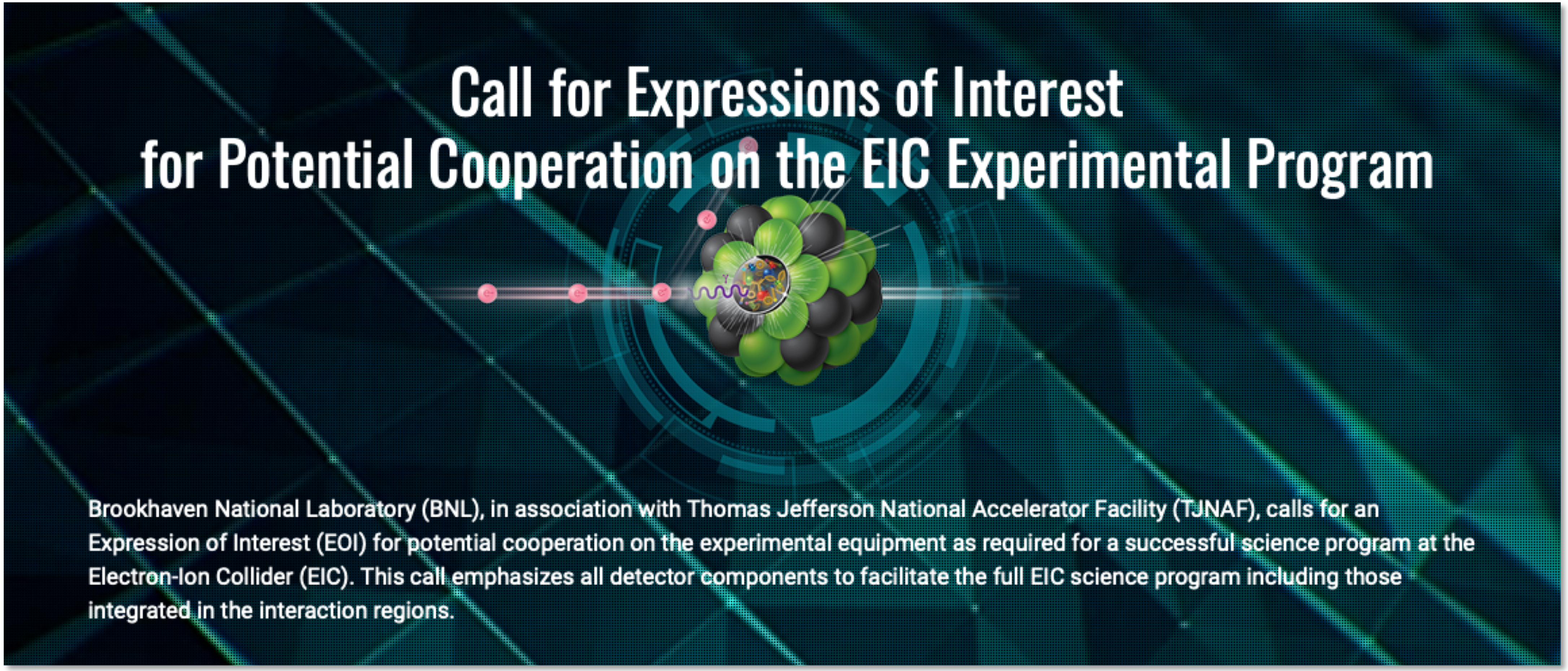
Online tutorials



Recordings from tutorials

- **Advanced Fast Simulation Tutorial** Fast simulations on the command line and in JupyterLab, singularity
- **Detector Full Simulation Tutorials** Geant4 for EIC, how to modify existing detector concepts, and how to integrate a new detector into one of the existing detector concepts.
- **Jim Pivarski Tutorial: uproot and Awkward Array** process and analyze Root files with pure Python libraries
- **MCEG Tutorial** Herwig, Pythia, Sherpa, Rivet, more to follow

SWG Priority: Realize our Expression of Interest



Community Input for Expression of Interest

Software Needs	**Requirements** What software needs for EIC Software would you like to highlight **now**, in a few years, and for the completion of the EIC project?
	Technologies & Techniques What software technologies and techniques should be considered for the EIC?
Meeting Software Needs	What resources can your group contribute?

Common Projects: Expression of Interest for Software

https://eic.github.io/activities/eoi.html

Expression of Interest (EOI) for Software

Please indicate the name of the contact person for this submission:

Conveners of the Software Working Group:

- A. Bressan, M. Diefenthaler, and T. Wenaus
- eicug-software-conveners@eicug.org

Please indicate all institutions collectively involved in this submission of interest:

29 institutions

ANL	Argonne National Laboratory
BNL	Brookhaven National Laboratory
CEA/Irfu	IRFU at CEA /Saclay institute
EIC-India	Akal University, Central University of Karnataka, DAV College Chandigarh, Goa University, Indian Institute of Technology Bombay, Indian Institute of Technology Delhi, Indian Institute of Technology Indore, Indian Institute of Technology Patna, Indian Institute of Technology Madras, Malaviya National Institute of Technology Jaipur, Panjab University, Ramkrishna Mission Residential College Kolkata
IMP-CAS	Institute of Modern Physics - Chinese Academy of Sciences
INFN	Istituto Nazionale di Fisica Nucleare
JLab	Thomas Jefferson National Accelerator Facility
LANL	Los Alamos National Laboratory
LBNL and UC Berkeley	Lawrence Berkeley National Laboratory and University of California, Berkeley
NCBJ	National Centre for Nuclear Research
OhioU	Ohio University
ORNL	Oak Ridge National Laboratory
SBU	Stony Brook University
SLAC	SLAC National Accelerator Laboratory
SU	Shandong University

https://indico.bnl.gov/event/8552/contributions/43221/

- **Software Tools for Simulations and Reconstruction**
  - **Monte Carlo Event Generators**
  - **Detector Simulations**
  - **Reconstruction**
  - **Validation**
- **Middleware and Preservation**
  - **Workflows**
  - **Data and Analysis Preservation**
- **Interaction with the Software Tools**
  - **Explore User-Centered Design**
  - **Discoverable Software**
  - **Data Model**
- **Future Technologies**
  - Artificial Intelligence
  - Heterogeneous computing
  - New languages and tools
  - Collaborative software

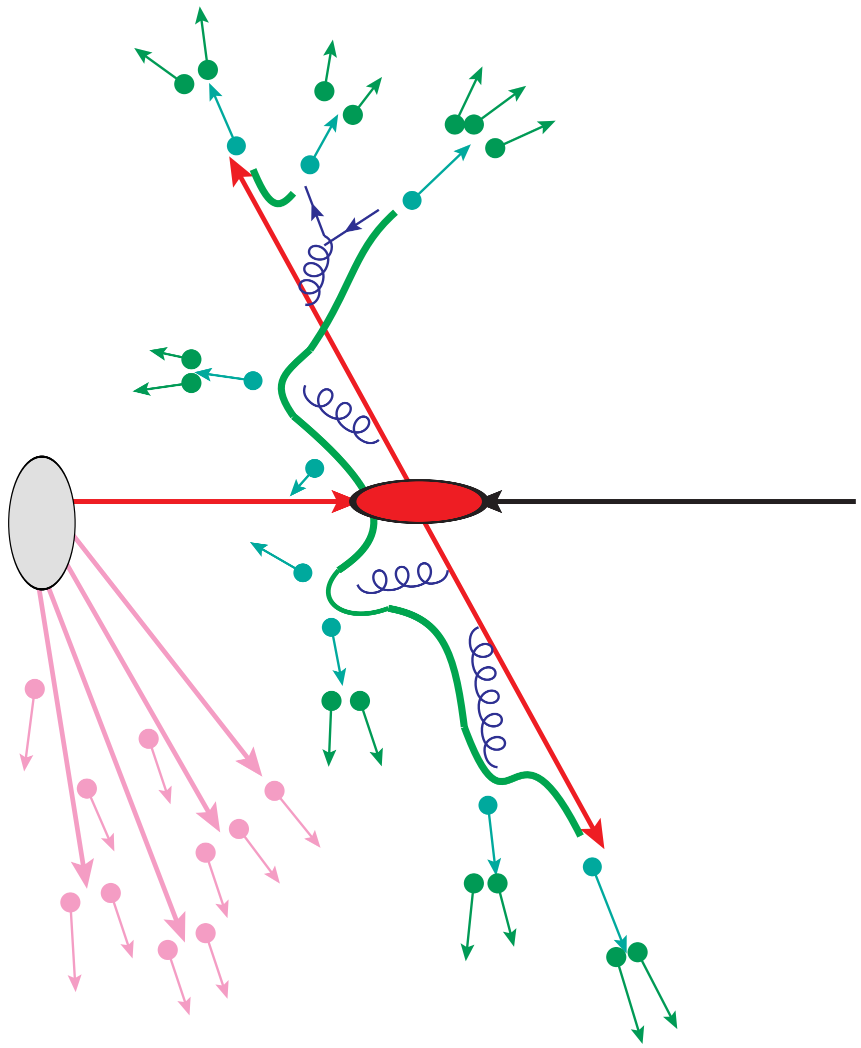
Monte Carlo Event Generator

MCEG

- faithful representation of QCD dynamics
- based on QCD factorization and evolution equations

MCEG algorithm

1. Generate kinematics according to fixed-order matrix elements and a PDF.
2. QCD Evolution via parton shower model (resummation of soft gluons and parton-parton scatterings).
3. Hadronize all outgoing partons including the remnants according to a model.
4. Decay unstable hadrons.



Building a MCEG community for the EIC

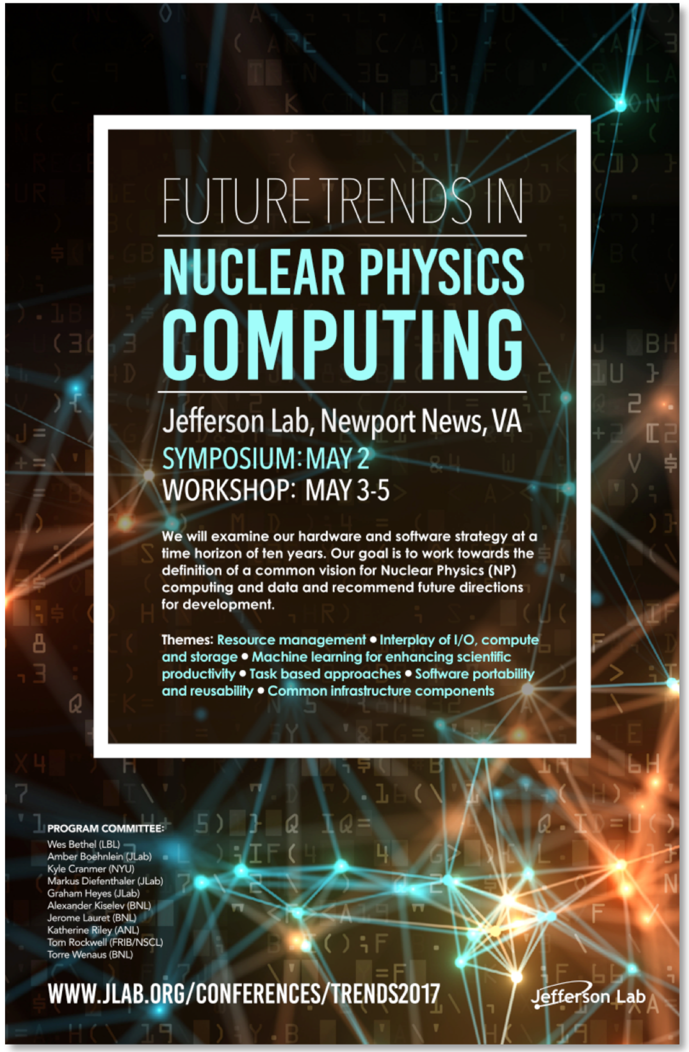
Unique MCEG requirements for EIC Science

- MCEG for polarized ep, ed, and eHe^3
  - including novel QCD phenomena: GPDs, TMDs
- MCEG for eA
- **Merging of QED+QCD effects**

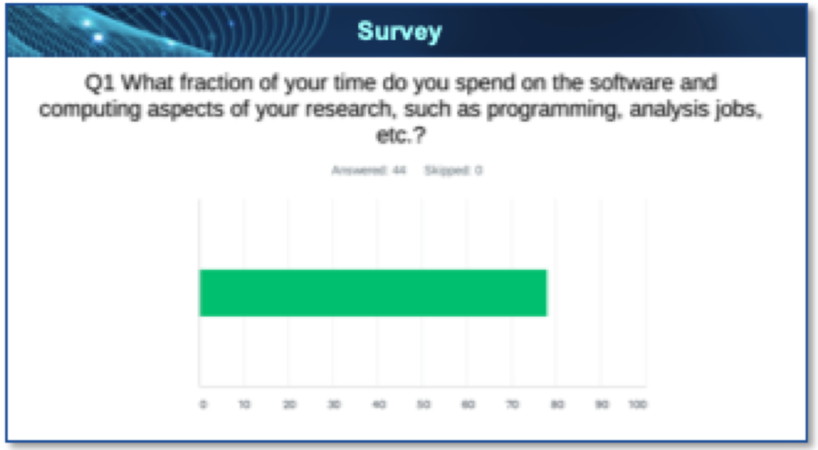
MCEG community

- focus of last two decades: **LHC**
  - **lesson learned** high-precision QCD measurements require high-precision MCEGs
  - MCEG not about tuning but about physics
- ready to work on ep/eA

Explore User-Centered Design



- **Software and computing are an integral part of our research**



- **Goal** All scientists of all levels worldwide should be enabled to participate in EIC simulations and analyses actively.
- **User-Centered Design** To achieve this goal, we must develop simulation and analysis software using modern and advanced technologies while *hiding* that complexity.

Initial Step: State of Software Survey

ELECTRON ION COLLIDER USER GROUP
STATE OF SOFTWARE SURVEY

Survey from February 16 – 23, 2021. Full questions and answers are listed in the appendix.

The Software Working Group collected information on the community's specific software tools and practices during the Yellow Report Initiative. This *software census* will be essential to better understand and quantify software usage throughout the EIC community.

Survey results summarized by Wouter Deconinck (Manitoba), Markus Diefenthaler (JLab), Rebecca Duckett (JLab), Sylvester Joosten (ANL), and Kolja Kauder (BNL).

Download Report

Developing User Stories

Project with BNL and JLab Communication Offices and User Experience Consultant T. Wiggins

Focus Group Discussions

User Stories

Input to software developers as to which users they are writing software for.

Focus Group: Students

Focus Group: Junior Postdocs

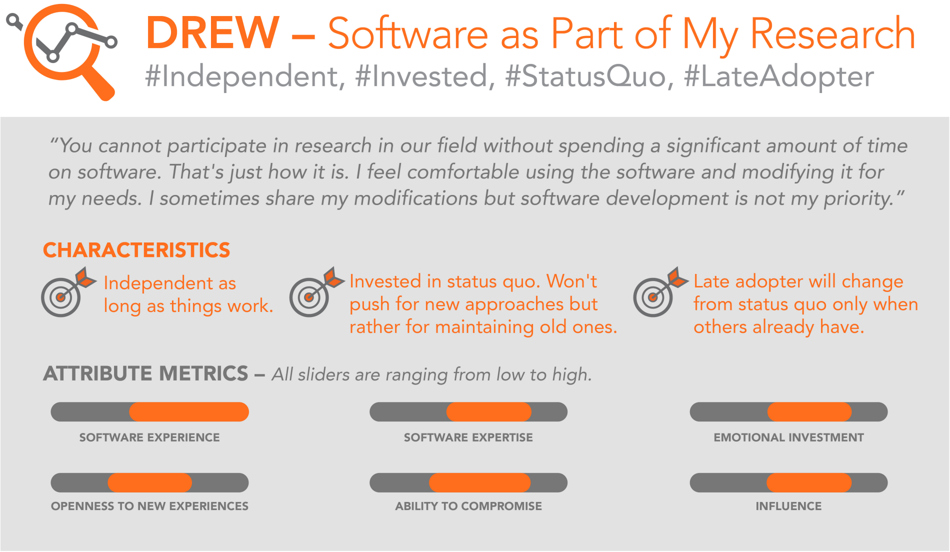
Focus Group: Senior Postdocs

Focus Group: Professors

Focus Group: Staff Scientists

Focus Group: Industry

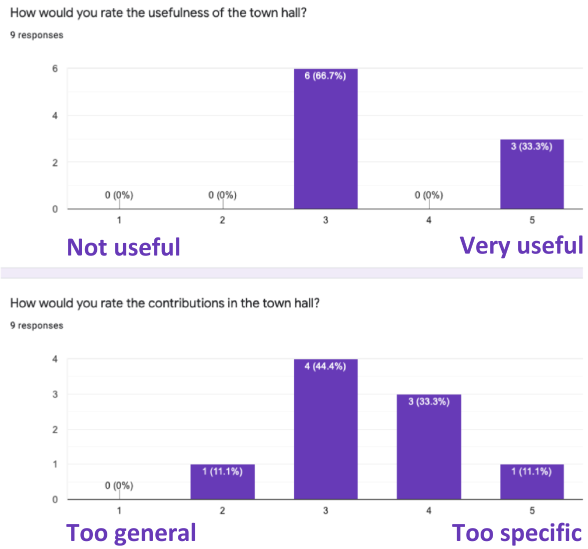
Extrmely valueable feedback, including many suggestions and ideas.



More detailed user profiles with a (partial) focus on software

- *"Important outreach to put a human face on the exciting science we work so hard to do!"* (John Lajoie)
- Rachel Montgomery (University of Glasgow)
- Vaibhavi Gawas (IIT Madras)
- Prabhakar Palni (Goa University)
- Alex Jentsch (BNL)

Software Town Hall



EIC Science and Software

All scientists of all levels worldwide should be enabled to participate actively in the science of the Electron-Ion Collider (EIC). To achieve this goal, we need to understand the requirements of the community on the data analysis software and workflows first and foremost.

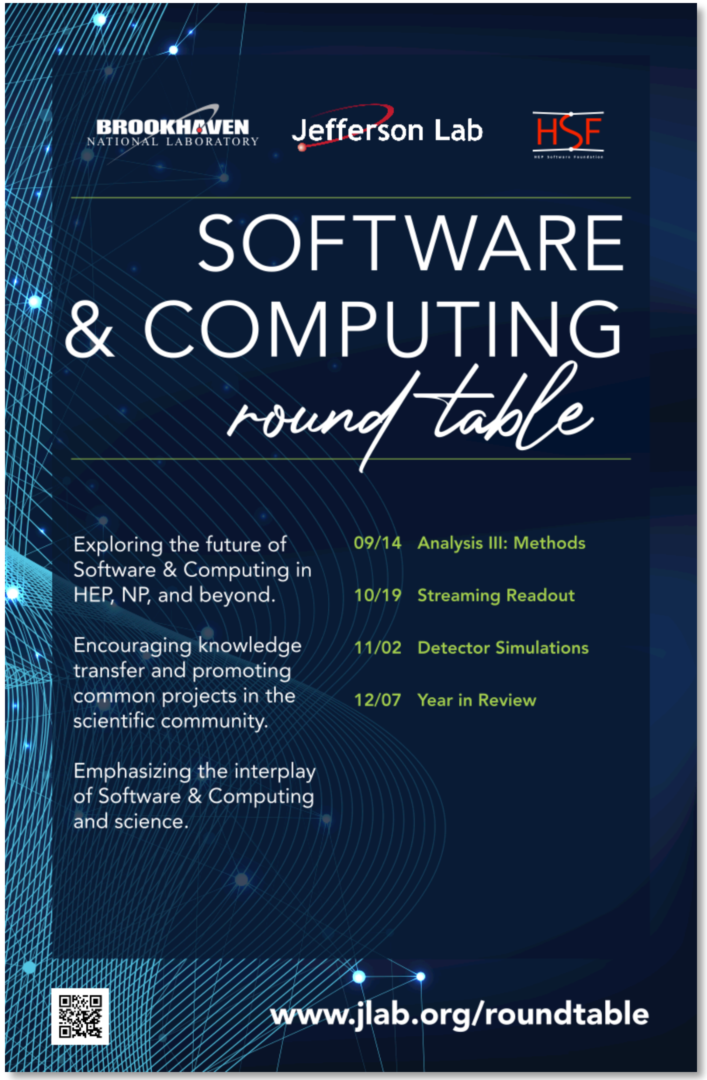
Software Town Hall

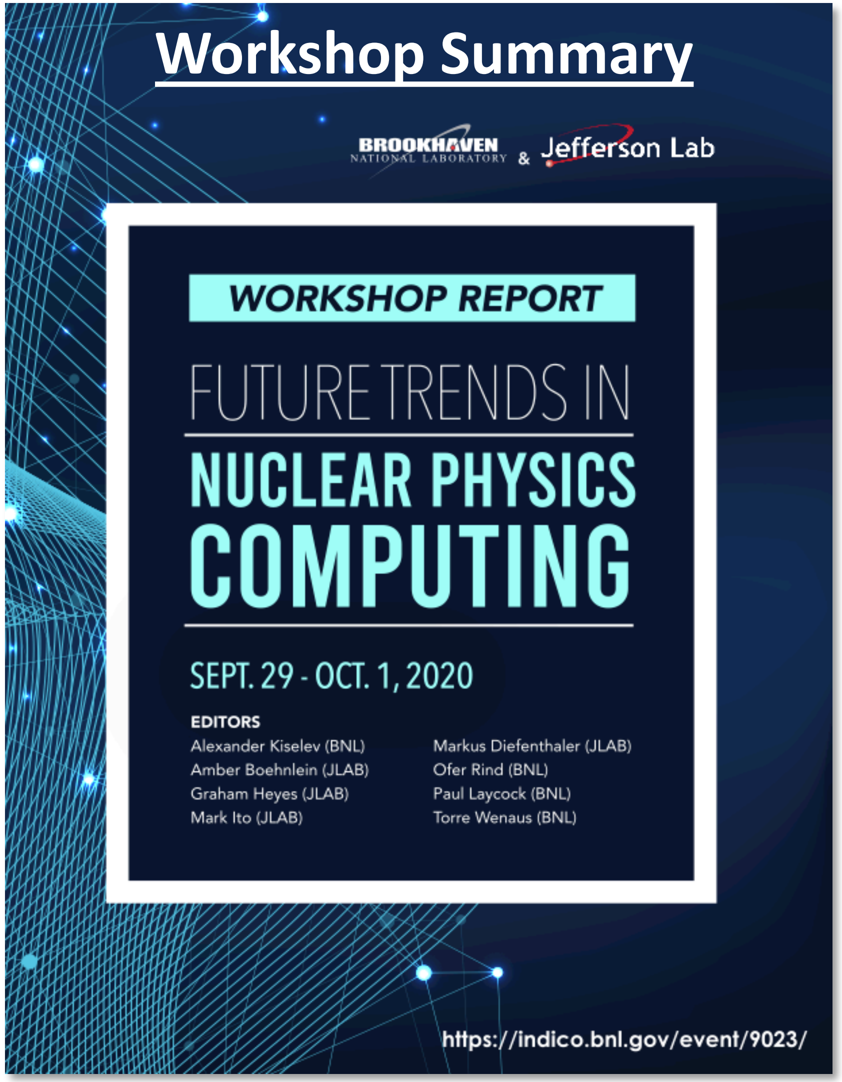
The idea of the event is to allow anyone in the EIC community a chance to share past experiences or suggest requirements for EIC Software in an open environment.

Organizers

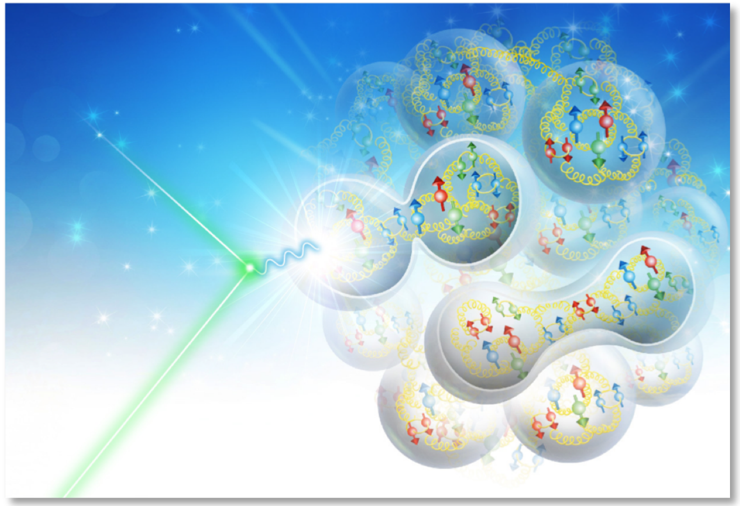
W. Deconinck, A. Deshpande, M. Diefenthaler, O. Evdokimov, T. Hemmick, D. Higinbotham, and K. Kauder.

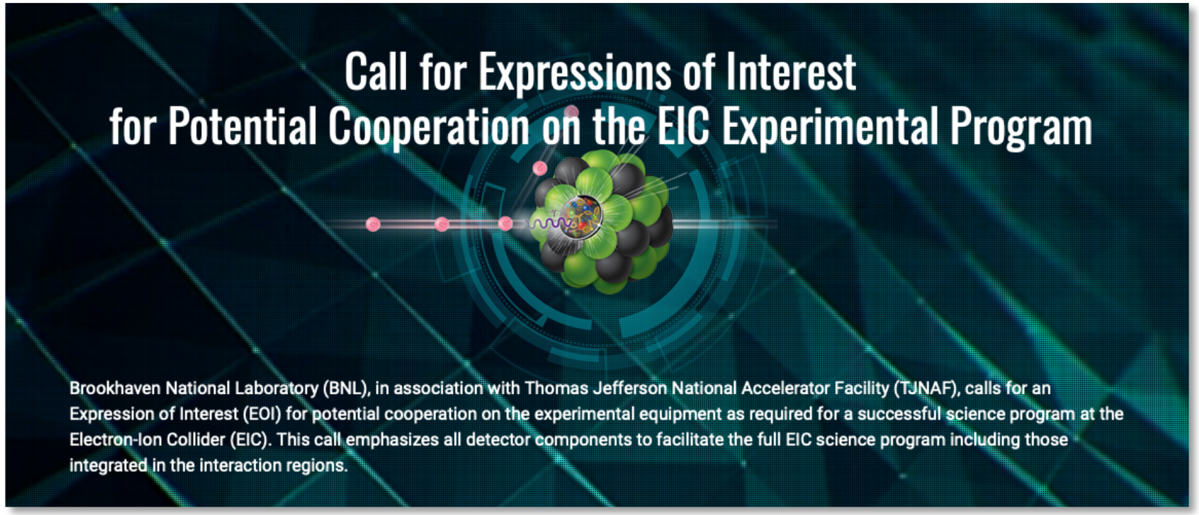
Interplay of Science and Software & Computing





Summary





Develop EIC Science and experiments

working towards the future of NP

Realize Software EoI

sustainable effort, common projects

Weekly Meetings https://indico.bnl.gov/category/301/

Mailing List eicug-software@eicug.org

Get involved