# Event Generators for High-Energy Physics Experiments

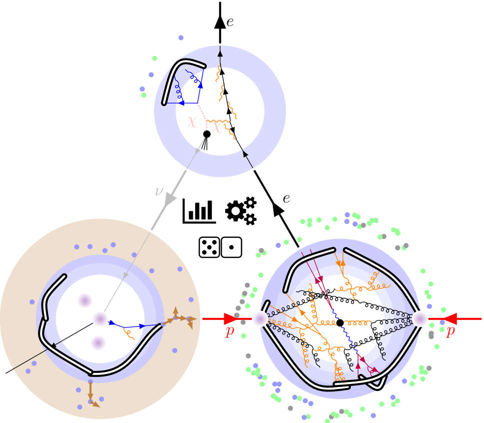
Joshua Isaacson

Based on: Snowmass White Paper (arxiv:2203.11110)

LPC Seminar

23 June 2022

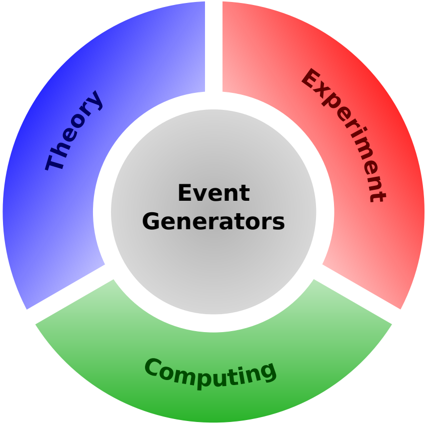
# Introduction



- The success of HEP experiments critically relies on advancements in physics modelling and computational techniques, driven by a close dialogue between large experimental collaborations and small teams of event generator authors.
- Development, validation, and long-term support of event generators requires a vibrant research program at the interface of theory, experiment, and computing

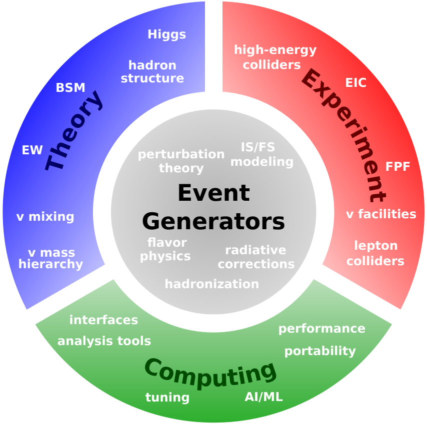
# Introduction

- White paper brought together all event generator communities in HEP for the first time
- Need to continue this collaboration through the creation of a joint theoretical-experimental working group cross-cutting through all experiments



# Why do we need generators?

- Precision understanding of Standard Model
- Ability to model BSM processes
- Essential role in planning and design of future experiments
- Connects the theory and experimental community
- Modelling non-perturbative effects



# Outline

1. Cross-Cutting
    - Physics Components
    - Computing Components
2. Non-LHC
    - Neutrino Experiments
    - Electron-Ion Collider
    - Forward Physics Facility
    - Lepton Colliders
3. LHC
    - Higher-order QCD and EW computations
    - QCD parton and dipole showers
    - Matching fixed-order to parton showers
    - General-purpose resummation tools

# Outline

# Fixed Order Calculations

- Automation has been achieved for tree level and next-to-leading order calculations
- NLO automation required:
    - Real-emission corrections, subtracted by local counterterms
    - Combination of Born, virtual, and integrated subtraction terms
    - Evaluating the one-loop integrals automatically
- Large development in fully-differential NNLO calculations
- A few processes at N$^3$LO accuracy

# Fixed Order Calculations

| Higgs | SM candles | Jets | Other |
|---|---|---|---|
| $H$ | $W^\pm$ | dijets | single top |
| $W^\pm H$ | $Z$ | 3 jets | $t\bar{t}$ |
| $ZH$ | $\gamma\gamma$ | $W^\pm$+jet | $b\bar{b}$ |
| H (VBF) | $W^\pm\gamma$ | $Z$+jet | $H \to b\bar{b}$ |
| HH | $Z\gamma$ | $\gamma$+jet | $t$ decay |
| HHH | $W^+W^-$ | $Z + b$ | $\mathrm{e}^+\mathrm{e}^- \to 3j$ |
| $H$+jet | $WZ$ | $W^\pm c$ | DIS (di-)jets |
| $W^\pm H$+jet | $ZZ$ | $\gamma\gamma$+jet | |
| $ZH$+jet | $\gamma\gamma\gamma$ | | |

Calculations available differentially at NNLO or higher in QCD (for $pp$ initial state). References to the first time the process has been calculated can be found in Table 1 of the white paper.

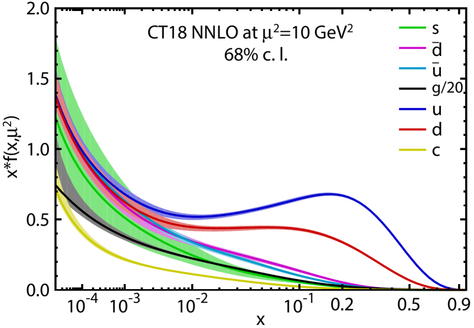
# QCD factorization and parton evolution

- Factorize into short and long distance physics:

$$\sigma[J] \approx \sum_{a,b} \int \mathrm{d}x_a \int \mathrm{d}x_b \; f_{a/A}(x_a, \mu_J^2)\, f_{b/B}(x_b, \mu_J^2)\; \hat{\sigma}[J]$$

- QCD evolution given by:

$$\frac{\mathrm{d}\, x f_{a/A}(x, \mu_J^2)}{\mathrm{d}\ln \mu_J^2} = \sum_{b=q,g} \int_0^1 \mathrm{d}\tau \int_0^1 \mathrm{d}z \, \frac{\alpha_s}{2\pi} \left[ z P_{ab}(z) \right]_+ \tau f_{b/A}(\tau, \mu_J^2)\, \delta(x - \tau z)$$



- PDFs and fragmentation functions are not always consistent
- Improving PDF understanding for neutrino experiments and the EIC are vital
- Work on using lattice to improve PDF accuracy

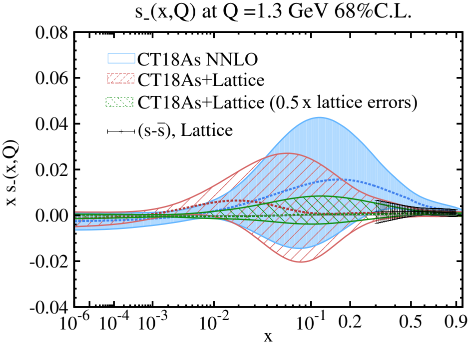
# QCD factorization and parton evolution

- Factorize into short and long distance physics:

$$\sigma[J] \approx \sum_{a,b} \int \mathrm{d}x_a \int \mathrm{d}x_b \; f_{a/A}(x_a, \mu_J^2)\, f_{b/B}(x_b, \mu_J^2)\; \hat{\sigma}[J]$$

- QCD evolution given by:

$$\frac{\mathrm{d}\, x f_{a/A}(x,\mu_J^2)}{\mathrm{d}\ln \mu_J^2} = \sum_{b=q,g} \int_0^1 \mathrm{d}\tau \int_0^1 \mathrm{d}z\, \frac{\alpha_s}{2\pi} \left[z P_{ab}(z)\right]_+ \tau f_{b/A}(\tau, \mu_J^2)\, \delta(x - \tau z)$$



- PDFs and fragmentation functions are not always consistent
- Improving PDF understanding for neutrino experiments and the EIC are vital
- Work on using lattice to improve PDF accuracy

# Photoproduction

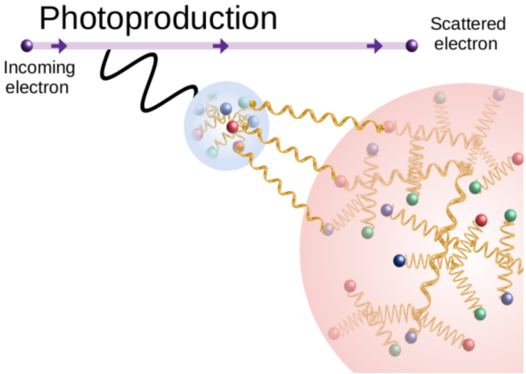


Image reproduced from: [2106.12377]

- Important for lepton-lepton and lepton-hadron collisions
- Low-virtuality photon fluctuates to a hadronic state with partonic content
- Most data comes from LEP and HERA
- Often split into a vector-meson-dominance non-perturbative piece and a point-like perturbative piece
- Can be applied to ultra-peripheral collisions at the LHC experiments

# QED showers

## All-orders decay rate

$$\mathrm{d}\Gamma^{\text{QEDPS}} = \mathrm{d}\Gamma_0 \left\{ 1 + \sum_{c=1}^{n_{\text{ch}}} \sum_{n_\gamma} \frac{(\alpha\, L_c)^{n_\gamma}}{n_\gamma!} \left[ \prod_{i=1}^{n_\gamma} \mathrm{d}x_c^i \right] P_{\epsilon_{\text{cut}}}(x_c^1) \otimes \ldots \otimes P_{\epsilon_{\text{cut}}}(x_c^{n_\gamma}) \right\}$$

- $n_{\text{ch}}$ is charged particles, $L_c$ is the log ratio of the masses, $x_c$ is the retained energy fraction after radiating $n_\gamma$ photons
- Phase space described by Altarelli-Parisi splitting functions $P_{\epsilon_{\text{cut}}}$
- QED showers implemented into Herwig, Photos, Sherpa, and Vincia
- Vincia correctly captures the multipole structure of soft photon emissions

# Soft photon resummation

- Approach developed by Yennie, Frautschi, and Suura (YFS)
- Key feature is the ability to systematically improve by including exact fixed-order expressions
- YFS has been implemented into LEP tools such as KKMC, KoralW/YFSWW
- YFS has also been included in Herwig and Sherpa

## YFS formalism

$$\mathrm{d}\sigma = \sum_{n_\gamma=0}^{\infty} \frac{e^{Y(\Omega)}}{n_\gamma!} \mathrm{d}\Phi_Q \left[\prod_{i=1}^{n_\gamma} \mathrm{d}\Phi_i^\gamma \, \tilde{S}\left(k_i\right)\right] \left(\tilde{\beta}_0 + \sum_{j=1}^{n_\gamma} \frac{\tilde{\beta}_1(k_j)}{\tilde{S}\left(k_j\right)} + \sum_{\substack{j,k=1 \\ j<k}}^{n_\gamma} \frac{\tilde{\beta}_2(k_j, k_k)}{\tilde{S}\left(k_j\right) \tilde{S}\left(k_k\right)} + \cdots\right),$$

# Hadronization

## Lund String Model

- Basic assumption: linear confinement potential approximated by a string stretched between $q\bar{q}$ pairs
- Stored energy in string used to produce new $q\bar{q}$ pairs
- Baryons introduced by splitting to a antidiquark-diquark pair
- Gluons treated as kinks on the string
- Many improvements over the years, but still much work is needed

## Cluster Model

- Guided by local parton-hadron duality and preconfinement
- Evolution based on formation and decay of color-neutral clusters interpreted as resonances of hadrons with a continuous mass spectrum
- Baryons introduced by introduction of diquarks
- Gluons are split into flavor-antiflavor pairs at end of parton shower
- Need to revisit questions of very forward hadronization and color reconnections

# Final-state interactions

- Color Reconnection:
    - Color structure of partons only partially determined by single scattering, modelled by color reconnection
    - Color reconnection assumes that partons close in momentum space will also be close in color space
- Hadronic transport:
    - In interactions with nuclei, the hadrons involved in the hard interaction need to propagate out of the nucleus
    - Important for EIC, FPF, and neutrino experiments
- Hadronic rescattering:
    - Hadrons produced during hadronization can undergo a secondary collision
    - Requires understanding the space-time structure of the hadronization process
    - Baryon number is reduced in rescattering through annihilation processes

# Heavy Flavor

**Major issues**:

- Handling of branching fractions between event generators, the PDG, and experimental collaborations
- Rare production from Hadronization makes generating things like $B_s^0$ nearly impossible (need to generate $\mathcal{O}(10^4)$ LHC events to produce a single $B_s^0$)
- The ability to handle production of heavy-flavor hadrons via NRQCD is desired, and some hadrons are missing completely in hadronization models
- Non-negligible lifetimes requires the development of a method to communicate between event-generators and detector simulation

# New-physics models

| Generator | Representation | | | | | |
|---|---|---|---|---|---|---|
| | singlet | triplet | octet | $\epsilon^{ijk}$ | 6 | 10 |
| MG5aMC | ✓ | ✓ | ✓ | ✓ | ✓ | |
| SHERPA | ✓ | ✓ | ✓ | | | |
| WHIZARD | ✓ | ✓ | ✓ | | | |

| Generator | Representations | | | Lorentz structures | | | Other aspects | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SM | Spin $\frac{3}{2}$ | Spin 2 | Custom | Majorana | 4-Fermi | Propagator | Running | Form factor | Unitarity |
| MG5aMC | ✓ | ✓ | ✓ | ✓ | ✓ | (✓) | ✓ | EFT | ✓ | |
| SHERPA | ✓ | | (✓) | ✓ | ✓ | (✓) | | | | |
| WHIZARD | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |

- FeynRules package allows for the generation of Feynman rules from nearly arbitrary Lagrangians
- UFO file format very successful
- UFO recently extended to be used for BSM in neutrino experiments [2110.15319]

# Outline

1. Cross-Cutting
   - Physics Components
   - Computing Components
2. Non-LHC
   - Neutrino Experiments
   - Electron-Ion Collider
   - Forward Physics Facility
   - Lepton Colliders
3. LHC
   - Higher-order QCD and EW computations
   - QCD parton and dipole showers
   - Matching fixed-order to parton showers
   - General-purpose resummation tools

# Standardized interfaces and analysis tools

**Standardized interfaces**:

- Reduce unnecessary duplication of effort
- Key interfaces:
    - File formats: LHE and HepMC
    - "Afterburners" (*i.e* EvtGen)
    - LHAPDF and TMDlib
- Ultimate goal is to have all experiments rely on a set of common interfaces. For example, all events outputs should use HepMC.

**Analysis tools**

- Rivet:
    - More than 1000 analyses are currently bundled
    - Mainly LHC, but includes other colliders
- Nuisance:
    - Tool for neutrino data/MC comparisons
    - Unofficial unified interface to multiple generator output formats

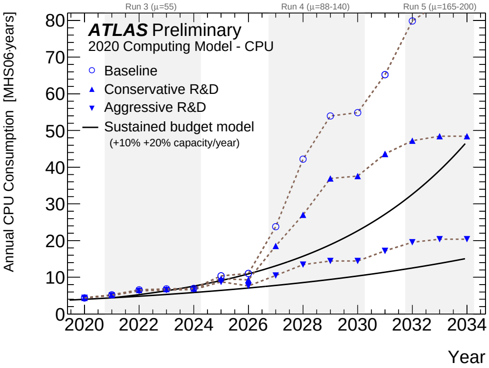
**Data Preservation**:

- Need to develop ability to preserve analyses reliant on machine-learning
- Deserves top-down structural attention in HEP as it profoundly affects the reproducibility and long-term scientific impact

# Tuning and systematics

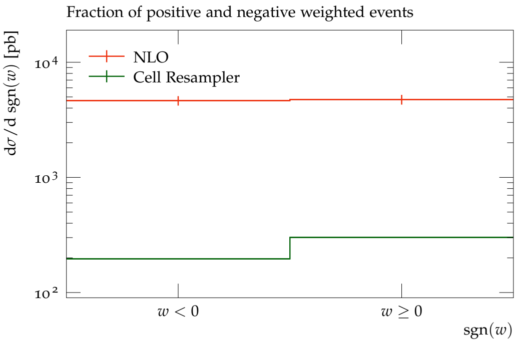
- Uncertainties from the theory can be classified as arising from:
    1. Underlying first-principles SM calculations
    2. Allowed range of parameters of a given phenomenological model, ideally constrained by and derived from a tune to data
    3. Choice of phenomenological model
- All of the above need to be considered before a claim of discrepancy.
- 1 and 2 can be controlled to some extent, 3 is harder to quantify
- Must be careful when tuning to not violate physics constraints

# Computing Bottlenecks



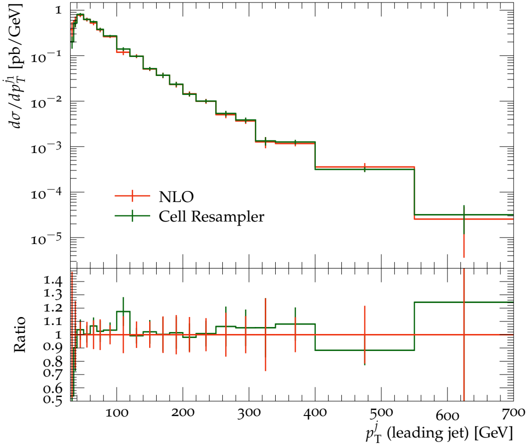
- Unweighting efficiency
- Handling (reducing) negative weights
- Efficiently propagating theory uncertainties (*i.e.* PDFs, scale variations, etc.)
- Modelling Bose-Einstein correlations, hadron rescattering, color reconnections, etc.
- Alternative event weights for parton showers to estimate uncertainties
- Matching / merging schemes have factorial growth problem
- Preliminary GPU implementations

# Computing Bottlenecks



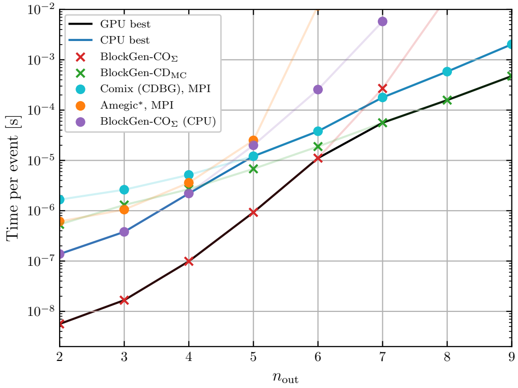
- Unweighting efficiency
- Handling (reducing) negative weights
- Efficiently propagating theory uncertainties (*i.e.* PDFs, scale variations, etc.)
- Modelling Bose-Einstein correlations, hadron rescattering, color reconnections, etc.
- Alternative event weights for parton showers to estimate uncertainties
- Matching / merging schemes have factorial growth problem
- Preliminary GPU implementations

# Computing Bottlenecks



- Unweighting efficiency
- Handling (reducing) negative weights
- Efficiently propagating theory uncertainties (*i.e.* PDFs, scale variations, etc.)
- Modelling Bose-Einstein correlations, hadron rescattering, color reconnections, etc.
- Alternative event weights for parton showers to estimate uncertainties
- Matching / merging schemes have factorial growth problem
- Preliminary GPU implementations

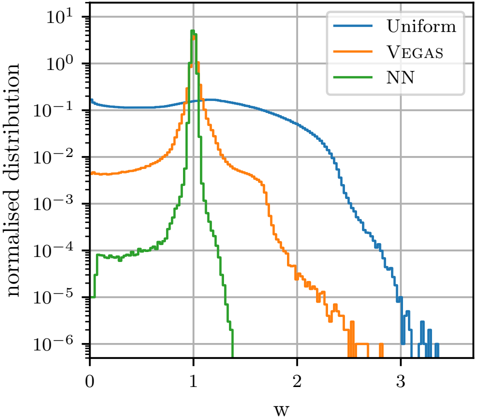
# Computing Bottlenecks



- Unweighting efficiency
- Handling (reducing) negative weights
- Efficiently propagating theory uncertainties (*i.e.* PDFs, scale variations, etc.)
- Modelling Bose-Einstein correlations, hadron rescattering, color reconnections, etc.
- Alternative event weights for parton showers to estimate uncertainties
- Matching / merging schemes have factorial growth problem
- Preliminary GPU implementations

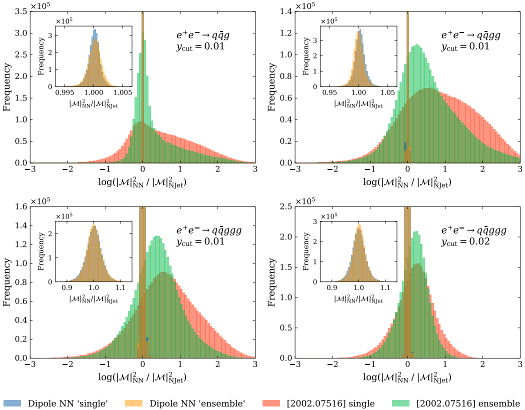
# Role of Machine Learning

- NNPDF parton densities
- Phase space integration
- Matrix element emulation
- Differentiable programming for optimization
- Generative networks
- See arxiv:2203.07460 for more discussion



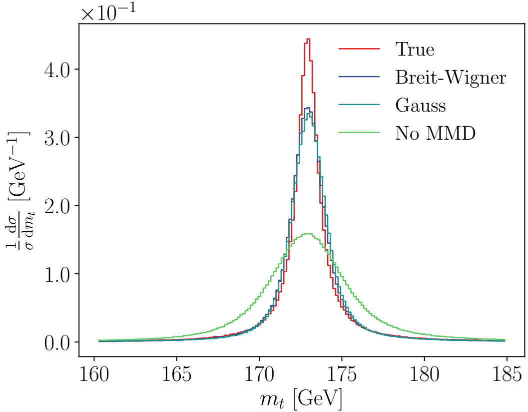
# Role of Machine Learning

- NNPDF parton densities
- Phase space integration
- Matrix element emulation
- Differentiable programming for optimization
- Generative networks
- See arxiv:2203.07460 for more discussion



# Role of Machine Learning

- NNPDF parton densities
- Phase space integration
- Matrix element emulation
- Differentiable programming for optimization
- Generative networks
- See arxiv:2203.07460 for more discussion

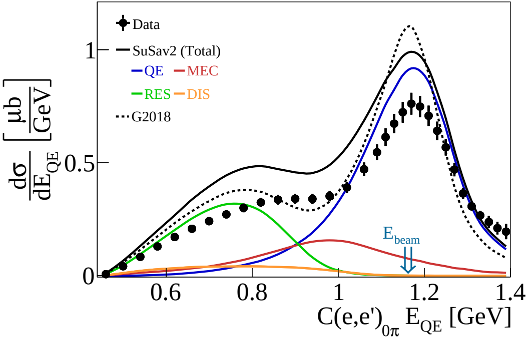


# Outline

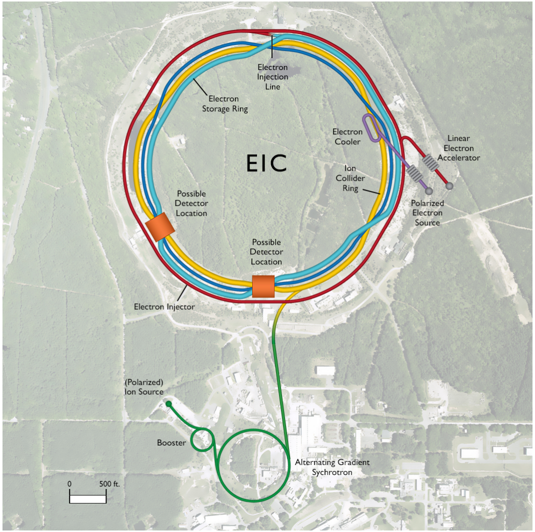
1. Cross-Cutting
   - Physics Components
   - Computing Components
2. Non-LHC
   - Neutrino Experiments
   - Electron-Ion Collider
   - Forward Physics Facility
   - Lepton Colliders
3. LHC
   - Higher-order QCD and EW computations
   - QCD parton and dipole showers
   - Matching fixed-order to parton showers
   - General-purpose resummation tools

# Neutrino Experiments

- DUNE requires a 1% uncertainty on interaction rates
- Excess from MiniBooNE and lack of evidence so far from MicroBooNE
- Main Generators:
  - GENIE
  - GiBUU
  - NEUT
  - NuWro
  - Achilles (in development)
- Significant improvements needed



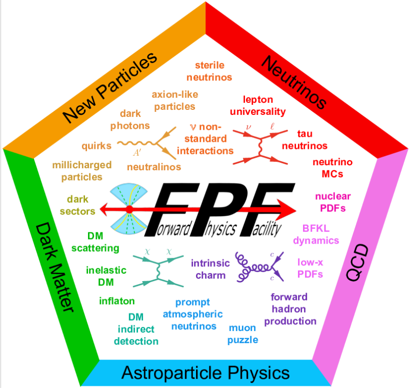
# Electron-Ion Collider



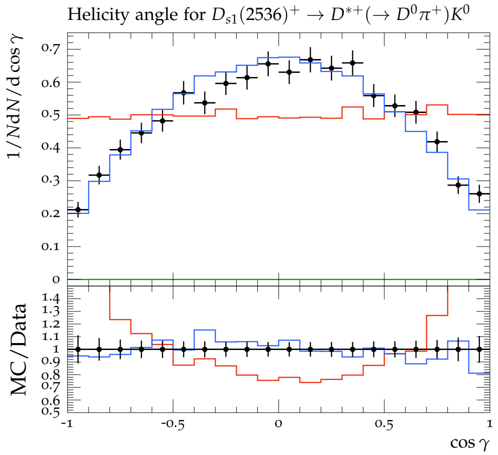
- Goal to investigate how quarks and gluons form nucleons and nuclei
- Provide drastic improvements for PDFs
- Generators:
  - General Purpose: Herwig, Sherpa, Pythia
  - Dedicated: BeaGLE, eSTARlight, IAGER, Sartre
- Requires ability to simulate collisions on spin-polarized electrons with spin-polarized light ions (p, d, $^3$He and unpolarized heavy ions (up to uranium)

# Forward Physics Facility

- Focus on far-forward direction (HL-LHC produce $10^{18}$ pions, $10^{17}$ kaons, and $10^{15}$ D-mesons within 2 mrad of the beam axis)
- Many different tools are required to simulate transport from production at the LHC to the interaction in the detectors
- Forward production of particles (especially charm) is not understood well, vital for modelling
- Use a mixture of LHC generators and cosmic ray generators for simulating forward production
- Use high-energy neutrino generators for hard interactions
- See arxiv: for more details



# Lepton Colliders



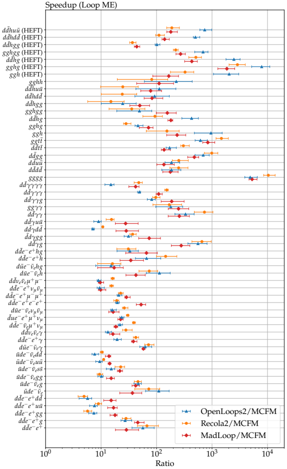
- Tools developed for DORIS/PEP/PETRA, SLC, TRISTAN, and LEP
- Major issues:
  - Collective beam-beam interactions and beam transport
  - Collinear factorization in $e^+e^-$ collisions
  - Soft photon resummation and matching
  - Heavy flavor production and decay
  - Polarization
- Generators: McMule, Sherpa, Whizard

# Outline

1. Cross-Cutting
   - Physics Components
   - Computing Components
2. Non-LHC
   - Neutrino Experiments
   - Electron-Ion Collider
   - Forward Physics Facility
   - Lepton Colliders
3. LHC
   - Higher-order QCD and EW computations
   - QCD parton and dipole showers
   - Matching fixed-order to parton showers
   - General-purpose resummation tools

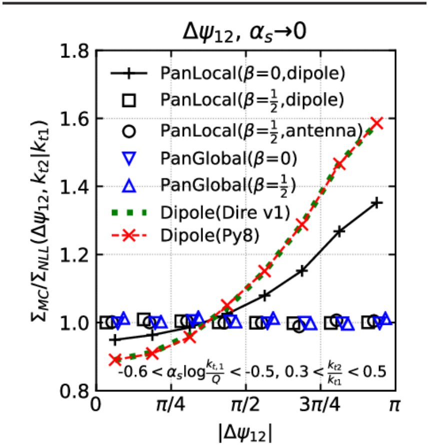
# Higher-order QCD and EW computations

- MADLOOP: One loop automated, work on two-loops
- MATRIX: NNLO accuracy through $q_T$ subtraction, mixed NNLO QCD-EW corrections
- MCFM: NNLO accuracy, recently added resummation using CuTe, work on interface to general purpose generators
- NNLOJet: NNLO accuracy, using antenna subtraction, work towards N$^3$LO
- OPENLOOPS: Automated generator of tree and one-loop amplitudes, stability techniques for one-loop contributions in unresolved regions of phase space for NNLO calculations.
- RECOLA: Automated generator of tree and one-loop amplitudes for full SM and BSM.



[2107.04472]

# QCD parton and dipole showers



[2002.11114]

- Many tools exist for parton showers, but are limited in accuracy
- Drive to formally evaluate accuracy of parton showers is ongoing
- Work on a NLL accurate shower under development
- Work towards including sub-leading color into the shower
- Work towards correct spin-correlations maintained throughout shower
- Major questions on how you handle mass effects in a shower

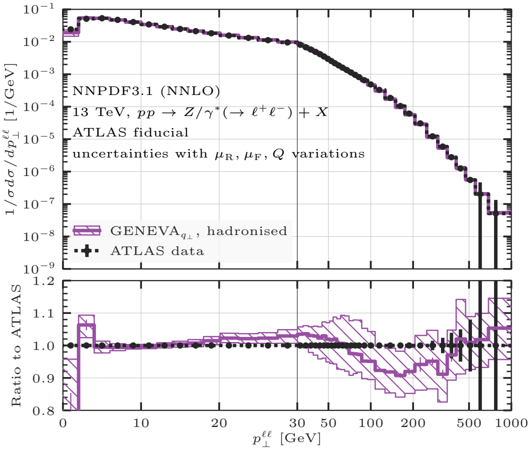
# Matching fixed-order to parton showers

- NLO Matching:
    - MC@NLO: Standard for general purpose generators
    - POWHEG: Combines matrix-element corrected parton showers
    - KRKNLO: Crucial advantage is its simplicity

- (N)LO multijet merging:
    - Combines strengths of matrix element calculations and parton showers
    - Soft and collinear radiation captured by shower
    - Hard radiation captured by higher multiplicity matrix element
    - VINCIAuses sector showers which reduce complexity of matching, merging, and matrix-element correction schemes

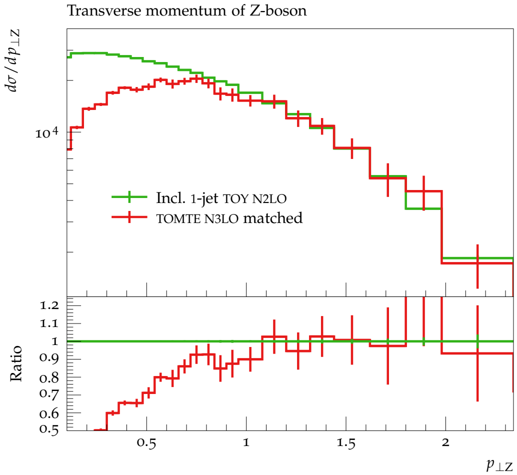
# Matching fixed-order to parton showers

- NNLO Matching
  - GENEVA: Use SCET to match fixed order to parton shower
  - NNLOPS and MiNNLO$_{\mathrm{PS}}$: No reweighting required, and parton shower based on POWHEGmethod
  - Need work in direction of fully differential matching
- TOMTE method for N$^3$LO matching, process independent, and constructed with a simple procedure

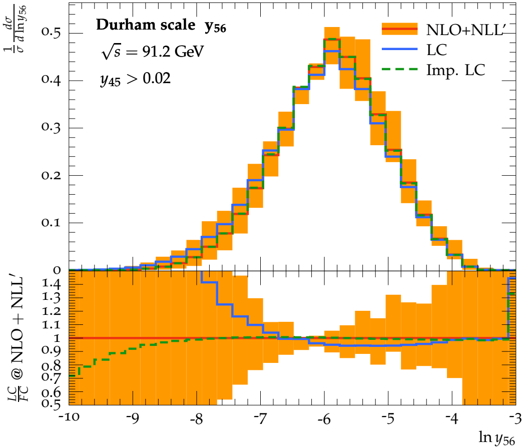


# Matching fixed-order to parton showers

- NNLO Matching
  - GENEVA: Use SCET to match fixed order to parton shower
  - NNLOPS and MiNNLO$_{\text{PS}}$: No reweighting required, and parton shower based on POWHEGmethod
  - Need work in direction of fully differential matching
- TOMTE method for N$^3$LO matching, process independent, and constructed with a simple procedure

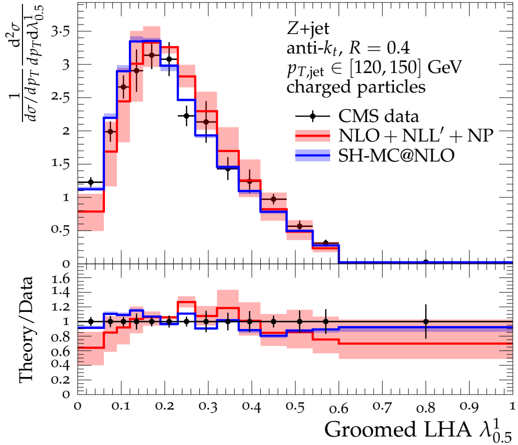


# General-purpose resummation tools



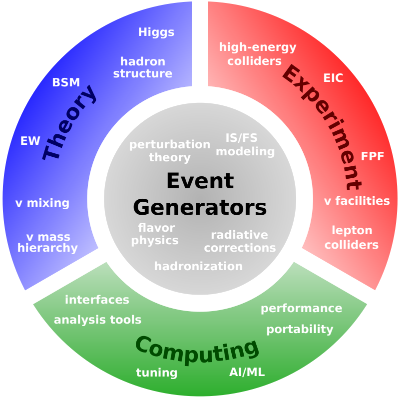
- CAESARFormalism: provides NLL'+NLO accuracy, plugin available to interface with SHERPA
- Possible extensions to NNLL accuracy via the ARESformalism.
- SHERPAhas framework for a $q_T$ resummation for $W$ and $Z$ at N$^3$LL' accuracy based on SCET

# General-purpose resummation tools



- CAESARFormalism: provides NLL'+NLO accuracy, plugin available to interface with SHERPA
- Possible extensions to NNLL accuracy via the ARESformalism.
- SHERPAhas framework for a $q_T$ resummation for $W$ and $Z$ at N$^3$LL' accuracy based on SCET

# Conclusions



**Theory**

**Higgs**

**hadron structure**

**BSM**

**EW**

**ν mixing**

**ν mass hierarchy**

**Experiment**

**high-energy colliders**

**EIC**

**FPF**

**ν facilities**

**lepton colliders**

**Event Generators**

**perturbation theory**

**IS/FS modeling**

**flavor physics**

**radiative corrections**

**hadronization**

**Computing**

**interfaces**

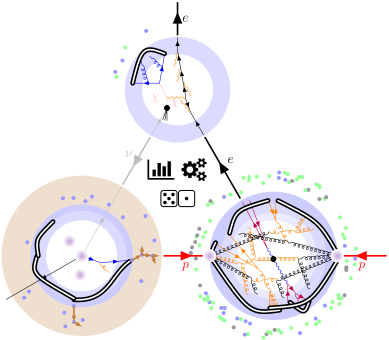
**analysis tools**

**performance**

**portability**

**tuning**

**AI/ML**



- **Event Generators are vital for the success of high energy experiments**
- **Event Generators bridge theory, experiment, and computing**