Precision QCD simulations

Frank Krauss

Institute for Particle Physics Phenomenology
Durham University

April 6^{th} 2023 – FNAL – Theory Seminar

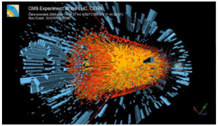
- why precision tools?
- current precision
- better parton showers
- future experiments
- soft QCD as limiting factor
- what is achievable?

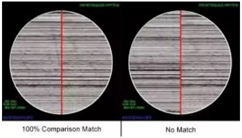
why precision

(carrying coal to Newcastle)

motivation: the need for (more) accurate tools

- to date no discovery of new physics (BSM) (a pity, but that's Nature)
- hope for "simple" discoveries is waning (don't expect anything glaringly obvious)
- push into precision tests of the Standard Model (find it or constrain "subtle"!)
- statistical uncertainties approach zero (because of fantastic work of accelerator, DAQ, etc.)
- systematic exp. uncertainties decrease (because of ingenious experimental work)
- theoretical uncertainties are or become dominant (obstacle to full explitation of LHC)





CSI LHC: need precise & accurate tools for precision physics

how to build an event generator

- paradigm: "divide et impera"
- divide simulation in distinct phases, with (logarithmically) separated scales
- start with signal event

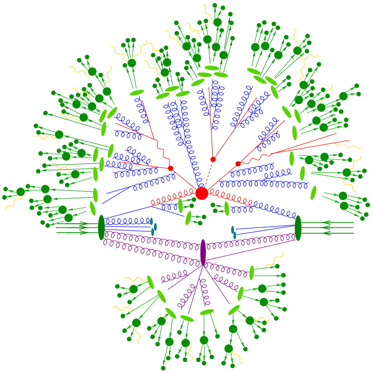
  (fixed order perturbation theory)
- dress partons with parton shower

  (resummed perturbatkon theory)
- add underlying event

  (phenomenological models)
- hadronize partons

  (phenomenological models)
- decay hadrons

  (effective theories, simple symmetries & data)



current precision

(where we are)

status: executive summary

- ✓ NNLO ⊗ parton shower for colour singlet production
  (MINNLO: 1309.4634, 1407.2940, . . . , 2208.12660; UNNLOPS: 1405.4607, 1407.3773)
- ✓ NNLO ⊗ parton shower for heavy quarks
  (MINNLO: 2112.04168 ($t\bar{t}$), 2302.01645 ($b\bar{b}$))
- ✓ MEPS@NLO: NLO multijet merging
  (SHERPA: 1207.5030; MADGRAPH: 1209.6215; PYTHIA: 1211.7278; HERWIG: 1705.06700 plus follow-ups & refinements)
- ✓ all of the above including EW@NLO
  (explicit: 1511.08692, 1705.00598, . . . , 2204.07652; Sudakov approximation: hep-ph/0010201, 2111.13453)
- ✓ (N)NLO ⊗ $\mathrm{N}^{1,2,3}$LL⊗ parton shower
  (GENEVA: 1211.7049, 1508.01475, 2102.08390, . . .)
- ▶ multijet merging with TMDs (2107.01224, 2208.02276 (not covered here))
- ▶ improving parton showers
  ((next-to leading) logarithmic accuracy (see below); amplitude evolution: 1802.08531, . . . (not covered here))

fixed-order accuracy

(apologies for any omissions in active field with ≈ 100 publications/past 5 years)

- N^{3}LO for single-boson production (1503.06056 . . . 1802.00833, 1807.11501)
  for DIS, and for VBF H-production in double DIS (1803.09973; 1606.00840)
- NNLO for practically all $2 \to 2$ (and some $2 \to 3$) processes:
  - jj (1705.10271, 1905.09047, . . .)
  - Vj, γj, Hj (1408.5325, 1504.02131, 1504.07922, 1505.03893, 1705.04664, 1901.11041, 1905.13738, . . .)
  - tt & single top (1303.6254, 1511.00549; 1404.7116, . . .)
  - VV and $\gamma\gamma$ (1408.5243, 1504.01330, 1507.06257, 1604.08576, 1605.02716, 1708.02925, 1711.06631, . . .)
  - VBF (1506.02660, 1802.02445, . . .)
  - dijets in DIS (1804.05663, . . .)
- virtual $2 \to \geq 3$ amplitudes (1511.05409, 1511.09404, 1604.06631, 1712.02229, 1811.11699, . . .)
- relative size argument: $\alpha_s^2 \approx \alpha_W$:
  must include NLO EW corrections for $\mathcal{O}(1-10\%)$ accuracy
  $\Longrightarrow$ automated in OpenLoops, Recola, aMC@Nlo _ MadGraph

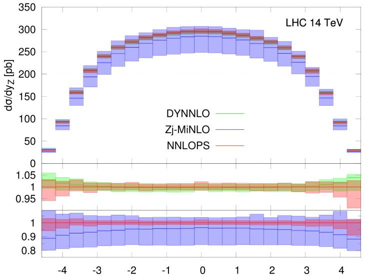
(1705.00598, 1704.05783, 1405.0301)

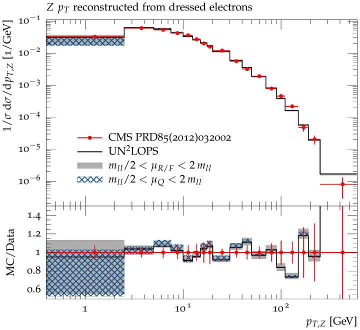
SM precision simulation in a nutshell: Drell-Yan

- current "accuracy standard(s)":
  - fixed-order: N^{3}LO for inclusive, NNLO for Vj
  - matching: NNLOPS for inclusive V
  - merging: MEPS@NLO for $V+\leq 2$ jets at NLO $V+\geq 3$ jets at LO
- dominating QCD effects: $\mathcal{O}$(10-30%)
  - low-$p_\perp$ region dominated by parton shower
  - high-$p_\perp$ region dominated by (multi-) jet topologies
  - higher accuracy in rate (and some shapes) through NNLO matching
- must add EW corrections for %-level precision
  - EW correction at large scales $\mathcal{O}(10\%)$
  - QED FSR + EW for V line shapes at $\mathcal{O}(1\%)$

NNLOPs for Z production: MINNLO & UNNLOPs

(1407.2904, 1405.3607)

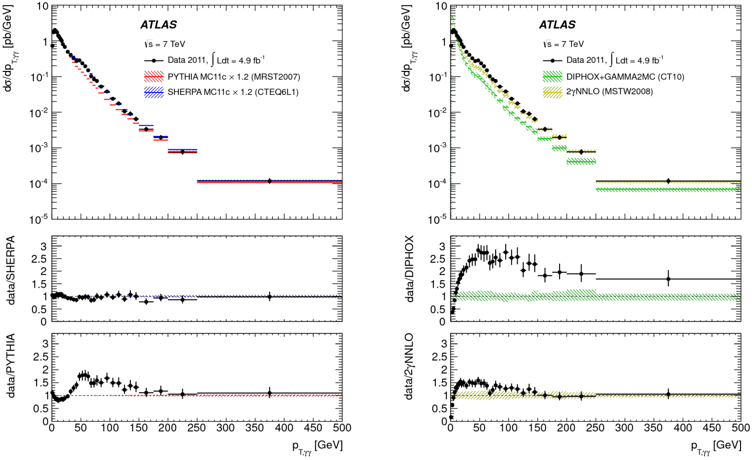




- different logic of achieving NNLO precision
- available for H, V production (both) and VV production (MINNLO)

merging example: $p_{\perp,\gamma\gamma}$ in MEPS@LO vs. NNLO

(arXiv:1211.1913 [hep-ex])



better parton showers?

(the story never gets old)

parton showers, compact notation

- Sudakov form factor (**no-decay** probability)

$$\Delta^{(\mathcal{K})}_{ij,k}(t, t_0) = \exp\left[-\int_{t_0}^{t} \frac{\mathrm{d}t}{t}\frac{\alpha_s}{2\pi}\int \mathrm{d}z\frac{\mathrm{d}\phi}{2\pi}\underbrace{\mathcal{K}_{ij,k}(t, z, \phi)}_{\text{splitting kernel for } (ij)\to ij \text{ (spectator } k)}\right]$$

- evolution parameter t defined by kinematics

generalised angle (HERWIG ++) or transverse momentum (PYTHIA, SHERPA)

- will replace $\frac{\mathrm{d}t}{t}\mathrm{d}z\frac{\mathrm{d}\phi}{2\pi} \longrightarrow \mathrm{d}\Phi$ (subtle differences important for theoretical accuracy (LL vs. NLL)!)
- scale choice for strong coupling: $\alpha_s(k_\perp^2)$ resums classes of higher logarithms
- regularisation through cut-off t_0 scale for onset of non-perturbative effects (hadronization)!

factorisation of amplitudes: colour coherence

- collinear:

$$ {}_n\langle 1,\ldots,n|1,\ldots,n\rangle_n \stackrel{i\|j}{\longrightarrow} \sum_{\lambda,\lambda'=\pm} {}_{n-1}\left\langle 1,\ldots,\not{i}(ij),\ldots,\not{j},\ldots,n\left|\frac{8\pi\alpha_s}{2p_ip_j}P^{\lambda\lambda'}_{(ij)i}(z)\right|1,\ldots,\not{i}(ij),\ldots,\not{j},\ldots,n\right\rangle_{n-1}, $$

  with spin-dependent splitting function $P^{\lambda\lambda'}_{(ij)i}(z)$

- soft:

$$ {}_n\langle 1,\ldots,n|1,\ldots,n\rangle_n \stackrel{p_j\to 0}{\longrightarrow} -8\pi\alpha_s\sum_{i,k\neq j} {}_{n-1}\left\langle 1,\ldots,\not{j},\ldots,n\left|\mathbf{T}_i\mathbf{T}_k\, w_{ik,j}\right|1,\ldots,\not{j},\ldots,n\right\rangle_{n-1} $$

  with colour-insertion operators $\mathbf{T}_{i,k}$ & soft eikonal

$$ w_{ik,j} = \frac{p_ip_k}{(p_ip_j)(p_jp_k)} = \frac{W_{ik,j}}{E_j^2} = \frac{1}{E_j^2}\frac{1-\cos\theta_{ik}}{(1-\cos\theta_{ij})(1-\cos\theta_{jk})} $$

(obviously, frame-dependent when expressed by energies & angles)

soft eikonals, decomposed

- textbook decomposition (pink bible): $W_{ik,j} = \tilde{W}^i_{ik,j} + \tilde{W}^k_{ki,j}$
  with "radiator functions" $\tilde{W}^i_{ik,j}$: (identify "splitters" to combine with collinear terms)

$$\tilde{W}^i_{ik,j} = \frac{1}{2}\left(\frac{1-\cos\theta_{ik}}{(1-\cos\theta_{ij})(1-\cos\theta_{jk})} + \frac{1}{1-\cos\theta_{ij}} - \frac{1}{1-\cos\theta_{jk}}\right)$$

- express θ_{jk} for use in i-splitter term:

$$\cos\theta_{jk} = \cos\theta_{ij}\cos\theta_{ik} + \sin\theta_{ij}\sin\theta_{ik}\cos\phi^i_{jk}\ \ldots$$

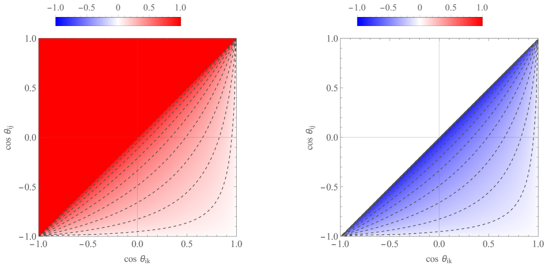
- ...and average over azimuth ϕ^i_{jk}:

$$\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\phi^i_{jk}\,\tilde{W}^i_{ik,j} = \frac{\tilde{I}^i_{ik,j}}{1-\cos\theta^i_j}\,, \qquad \text{where} \qquad \tilde{I}^i_{ik,j} = \begin{cases} 1 & \text{if} \quad \theta^i_j < \theta^i_k \\ 0 & \text{else} \end{cases}$$

(this is the well-known source of angular ordering)

- azimuthally integrated radiator function (normalised to 2π):

positive contributions negative contributions



- need to include azimuth modulation, if observables sensitive to it
- but: naive inclusion bound to fail (MC efficiency $\to 0$)

a new approach: PanScales

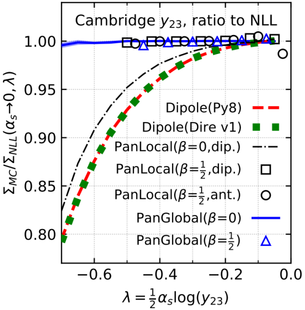
(resolving problems)

issues with existing/frequently used parton showers

- comparison with fixed-order results (PanScales: 1805.09327, 2205.02237)

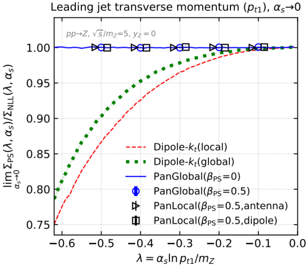
issues with existing/frequently used parton showers

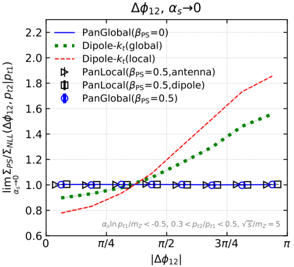
- comparison with fixed-order results (PanScales: 1805.09327, 2205.02237)
- logarithmic accuracy in FS showering (PanScales: 2002.11114, 2011.10054)



issues with existing/frequently used parton showers

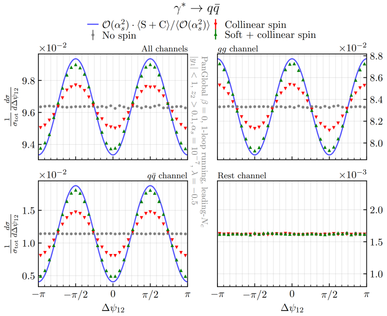
- comparison with fixed-order results (PanScales: 1805.09327, 2205.02237)
- logarithmic accuracy in FS showering (PanScales: 2002.11114, 2011.10054)
- logarithmic accuracy in IS showering (PanScales: 2207.09467)





issues with existing/frequently used parton showers

- comparison with fixed-order results (PanScales: 1805.09327, 2205.02237)
- logarithmic accuracy in FS showering (PanScales: 2002.11114, 2011.10054)
- logarithmic accuracy in IS showering (PanScales: 2207.09467)
- spin correlations (NPB 310 (1988) 571, hep-ph/0110108, PanScales: 2103.16526, 2111.01161)



issues with existing/frequently used parton showers

- comparison with fixed-order results (PanScales: 1805.09327, 2205.02237)
- logarithmic accuracy in FS showering (PanScales: 2002.11114, 2011.10054)
- logarithmic accuracy in IS showering (PanScales: 2207.09467)
- spin correlations (NPB 310 (1988) 571, hep-ph/0110108, PanScales: 2103.16526, 2111.01161)

✓ PanScales encodes massless parton showers at NLL accuracy

(1^{st} matching in e^+e^-: 2301.09645)

a new approach: ALARIC

(progress in SHERPA)

soft eikonals, decomposed again

(2208.06057)

- define **positive definite** radiators:

  (borrowing from Catani & Seymour, Nucl. Phys. **B485** (1997) 291)

$$\bar{W}^i_{ik,j} = \frac{1-\cos\theta_{ik}}{(1-\cos\theta_{ij})(2-\cos\theta_{ij}-\cos\theta_{jk})}$$

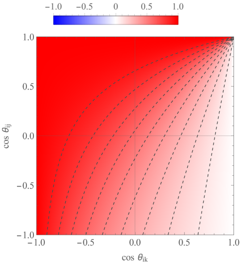
- same result after azimuth averaging, but $\tilde{I}^i_{ik,j} \longrightarrow \bar{I}^i_{ik,j}$ with

$$\bar{I}^i_{ik,j} = \frac{1}{\sqrt{(\bar{A}^i_{ij,k})^2 - (\bar{B}^i_{ij,k})^2}}$$

where

$$\bar{A}^i_{ij,k} = \frac{2-\cos\theta^i_j(1+\cos\theta^i_k)}{1-\cos\theta^i_k}\,,\quad \bar{B}^i_{ij,k} = \frac{\sqrt{(1-\cos^2\theta^i_j)(1-\cos^2\theta^i_k)}}{1-\cos\theta^i_k}$$

integrated radiator, $\bar{I}^i_{ik,j}$



matching with collinear terms

- collinear limit of eikonal factors:

$$w_{ik,j} \stackrel{i||j}{\longrightarrow} w_{ik,j}^{\rm (coll)}(z) = \frac{1}{2p_ip_j}\,\frac{2z}{1-z}\,, \qquad \text{where} \qquad z \stackrel{i||j}{\longrightarrow} \frac{E_i}{E_i+E_j}$$

- compare with leading $(1-z)$-terms of splitting functions

($1/z$ term in $g \longrightarrow gg$ captured with other "dipole")

$$P_{qq}(z) = C_F\left(\frac{2z}{1-z} + (1-z)\right)\,,$$
$$P_{gg}(z) = C_A\left(\frac{2z}{1-z} + z(1-z)\right)\,,$$
$$P_{gq}(z) = T_R\,(1-2z(1-z))\,.$$

$\longrightarrow$ defines "collinear remnant"

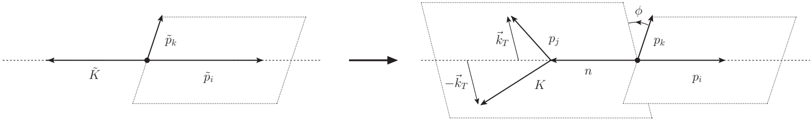
kinematics: birds-eye view

- kinematics as main obstacle to NLL accuracy in dipole showers: recoil of subsequent soft emissions may change "NLL history"
- construct new mapping $\{\tilde{p}_l\} \longrightarrow \{p_l\}$

(inspired by Catani & Seymour's treatment of identified hadrons)

- logic: disentangle colour spectator $\tilde{p}_k$ and recoil partner $\tilde{K}$

(*i.e.* define a **global** recoil scheme, use spectator for eikonal/azimuth)

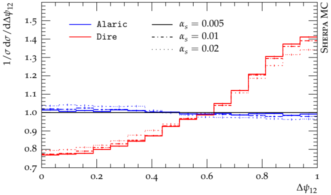


- basis to prove NLL accuracy analytically

set-up of numerical tests

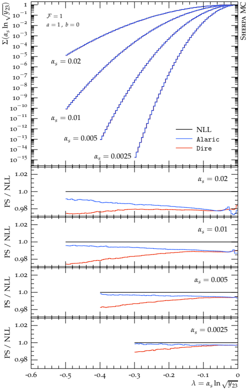
- compare results in $\alpha_S \to 0$ limit with NLL result
- set-up for checks
  - fixed α_S
  - leading colour $C_A = 2C_F = 3$
  - all partons massless
- example: azimuth angle between two leading Lund-plane declusterings

(should be $\Delta\Psi_{12} = 0$)

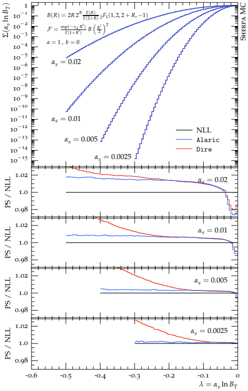


numerics: event shapes

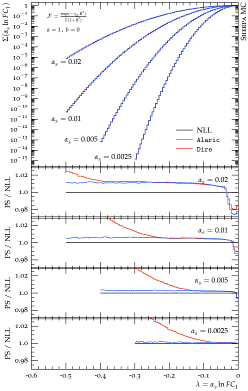
y_{23}



B_T



FC_1



how about data?

(always nice to see practical impact, innit?)

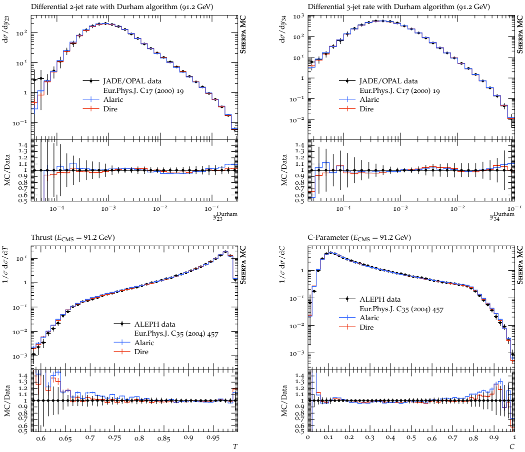
set-up of data comparison

- compare hadron-level results with LEP data
- perturbative set-up
  - no higher orders (no matching or merging)
  - running two-loop α_S with $\alpha_S(M_Z) = 0.118$
  - use CMW scheme for soft eikonal parts
  - all partons massless, masses emulated through simplistic thresholds
  - leading colour $C_A = N_c = 3,\ C_F = \frac{N_c^2-1}{2N_c}$
- non-perturbative set-up
  - need to use PYTHIA hadronization

    (ALARIC not yet ready for heavy hadron decays)
  - default parameters of PYTHIA 6.4, but

`PARJ(21)` $= 0.3$, `PARJ(41)` $= 0.4$, `PARJ(42)` $= 0.36$(ALARIC)/0.45(DIRE)

differential jet rates & event shapes



MC4EIC?

(preparing for the future)

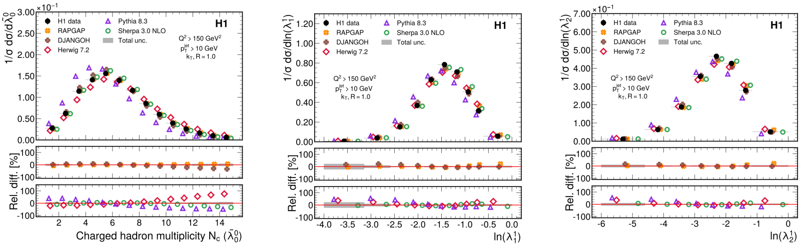
precision QCD studies at HERA

- "old" HERA data and analysis as boot-camp for EIC
- HERA = unique test-bed for (non-)perturbative QCD:
  - large-Q^2 DIS has no MPI $\longrightarrow$ initial state showering "clean"
  - large-Q^2 DIS has no MPI $\longrightarrow$ beam fragmentation "clean"
  - add HERA data to hadronization tunes?
- also: large photo-production cross section:
  - test hadronic structure of photon (relevant for EIC)
  - nota bene: last fits of photon-PDF are 20 years old
  - new fits urgently needed for EIC

(that is, if we want to treat collinear factorisation as limiting case for TMD's etc..)

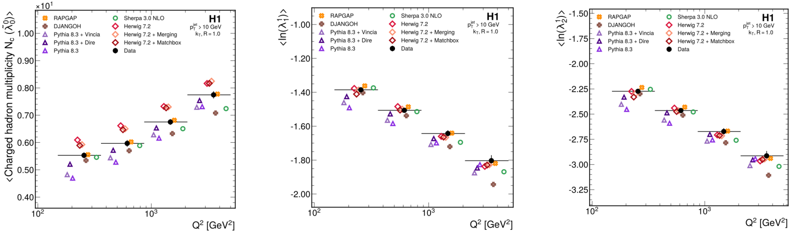
precision QCD studies at HERA

- incidentally, recent paper by H1 2303.13620
- uses modern MC's (HERWIG 7, PYTHIA 8, SHERPA)



precision QCD studies at HERA

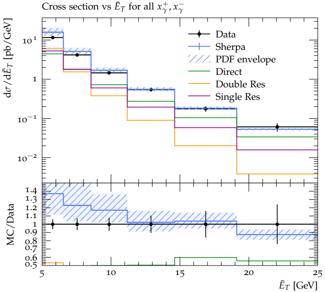
- incidentally, recent paper by H1 2303.13620
- uses modern MC's (HERWIG 7, PYTHIA 8, SHERPA)



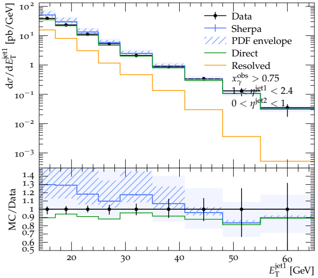
resolved photon processes

- "resolved" photons (i.e. with QCD structure/PDF) at LEP & HERA
- fits date from early 2000's: can supplement with NLO MC machinery
- first steps (SHERPA) below

OPAL, hep-ex/0301013



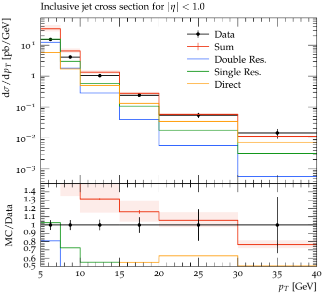
ZEUS, hep-ex/0112029



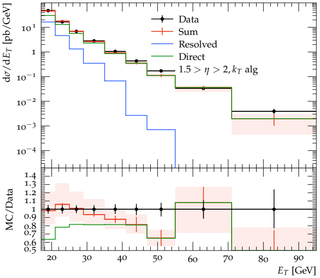
resolved photon processes

- "resolved" photons (i.e. with QCD structure/PDF) at LEP & HERA
- fits date from early 2000's: can supplement with NLO MC machinery
- first steps towards NLO (SHERPA) below

OPAL, hep-ex/0706.4381



ZEUS, hep-ex/1205.6153



soft physics strikes back

(the ultimate frontier?)

$g \to Q\bar{Q}$ — a systematic nightmare

- parton showers geared towards collinear & soft emissions of gluons

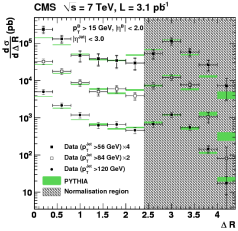
  (double log structure)

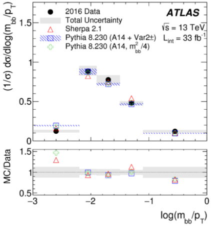
- $g \to q\bar{q}$ only collinear
- old measurements at LEP of $g \to b\bar{b}$ and $g \to c\bar{c}$ rate
- fix this at LHC for modern showers

  (important for $t\bar{t}b\bar{b}$)

- questions: kernel, scale in α_s

  (example: $k_\perp$ vs. m_{bb})

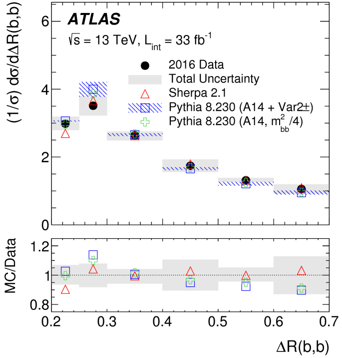


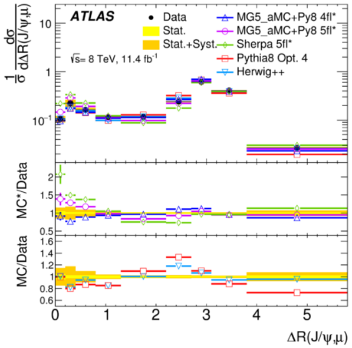


latest ATLAS measurements

(arXiv:1812.09283, 1705.03374 [hep-ex])

- use b-tagged jets with $R = 0.2$ (left)
- use muons in $B \to J/\Psi(\mu\mu) + X$ and $B \to \mu + X$ as proxies (right)





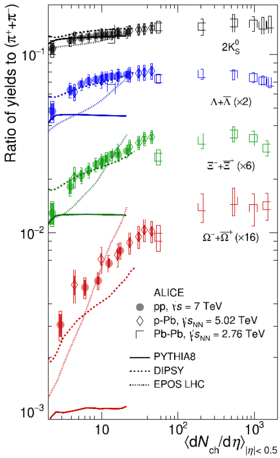
massive quarks are tricky - encore

- heavy quarks also problematic in initial state:
  no PDF support for $Q^2 \leq m_Q^2 \longrightarrow$ quarks stop showering
- possible solutions:
  - naive: ignore and leave for beam remnants (SHERPA)
  - better: enforce splitting in region around m_Q^2 (PYTHIA)
    $\longrightarrow$ effectively produces collinear Q and gluon in IS
- will need to check effect on precision obsevables: $p_\perp^{(W)}/p_\perp^{(Z)}$

soft physics: strange strangeness

- universality of hadronization assumed
- parameters tuned to LEP data
  in particular: strangeness suppression
- for strangeness: flat ratios
  but data do not reproduce this
- looks like $SU(3)$ restoration
  not observed for protons
- needs to be investigated (see next)

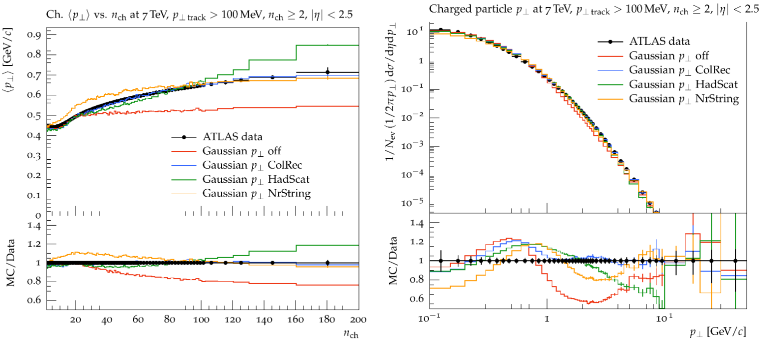
(plot from arXiv:1606.07424 [hep-ex])



hadronization issues

(illustrative plots from arXiv:1610.09818 [hep-ph])

- "ridge" in *pp* collisions:
- HI-like behaviour: unexpected, unexplained
- play with hadronization?



Developments & Challenges

(achievable goals in the next 5)

"predictions" for analytic tools

- expect progress in higher-order calculations
  - (semi-)automated NNLO_{QCD} for multi-particle FS & mixed exact $\text{NLO}_{\text{QCD}} \otimes \text{NLO}_{EW}$ for multi-particle FS
    (I expect this to be driven by constructing efficient basis of master integrals and improving and automating IR subtraction methods)
  - approximate and/or exact $\text{N}^3\text{LO}_{\text{QCD}}$ beyond $2 \to 1$ topologies
    (similar problems as before: master integrals, ...)
  - higher precision in PDF's
- don't know enough about resummation to have a clear picture, will concentrate on parton shower simulations instead

simulation

- reminders/lessons from past decade:
  - NLO matching with parton showers has been trivial

    (parton shower kernels equal NLO subtraction kernels)
  - MEPS@LO and MEPS@NLO usually agree at 10% level, reduced scale uncertainties at NLO

    (this is not yet fully explored when searching for efficient event generation)
- extrapolation to higher-orders in simulation:
  - NNLO will not be quite as simple - will need $\mathcal{O}(\alpha_S^2)$ kernels

    (tricky, think about it as fully automated NNLO subtraction kernels)
  - opens parton shower for systematic uncertainty analysis
    $\longrightarrow$ expect main effect in further reduction of scale uncertainties
  - non-trivial impact of choices (kinematics) on log accuracy

where does it stop?

theory limitations

"The great advances in science usually result from new tools rather than from new doctrines"
(Freeman Dyson)

- technical challenges: speed & stability of fixed-order calculations
  - numerical issues 1: phase-space integration efficiency

    (current approach: multi-channel with process-specific mappings)

    (hard to see how ML can make a massive difference – until now only "blanks")
  - numerical issues 2: special functions in multi-loop master integrals

    (convergence of series expansion, maybe need to go to quadruple precision)
  - numerical issues 3: stability of numerical $(N-1)$-loop results

    (example: NLO inputs to NNLO calculation, supremacy of compact analytic expressions)

$\longrightarrow$ this will necessitate highly technical, barely publishable work

(traditionally this is often a dead-end for careers in theory)

theory limitations

- physics challenges:
  - physics challenge 1: sources of uncertainty
    (simple scale variations by a factor of two may not be sufficient: central scale in multi-scale processes?
    in addition: new sources of uncertainty. example: kinematics scheme in parton showers and impact on log accuracy)

    $\longrightarrow$ community effort: agree on robust procedures
  - physics challenge 2: complexity & control
    (as the calculations become increasingly complex we need more and more independent – algorithmic, implementation, ... – checks. examples: two calculations for fixed order, two implementations of parton showers, etc.)

    $\longrightarrow$ community effort: robust and open (cross-)validation procedures
  - physics challenge 3: input parameters as source of uncertainty
    (my favourite example: PDF's with different values for $\alpha_S(M_Z)$
    and, yes, there's more: PDF's etc.)

    $\longrightarrow$ community effort: harvest "old data"

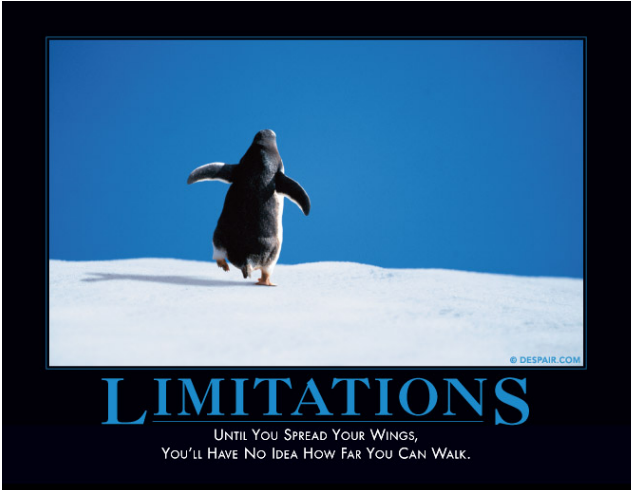
theory limitations

- “philosophical” challenges:
  - philosophical challenge 1: breakdown of factorisation
    (studies suggest that at some high order factorisation breaks down for basic 2 $\longrightarrow$ 2 QCD processes it is not clear how this translates to more complicated processes)

    $\longrightarrow$ don’t know if we should care
  - philosophical challenge 2: impact of higher-twist
    (this is closely related to 1.
    simple back-of-the envelope: at DY $m_P/M_Z \approx 1\% \approx \alpha_s^2$ – is this a problem?)

    $\longrightarrow$ this may need a careful analysis – is it linear or quadratic?
  - philosophical challenge 3: impact of soft physics uncertainties
    (lots of them a a hadron collider: MPI/underlying event, hadronization, . . . –
    quite often models aiming to describe these effects are too similar, and based on identical data)

    $\longrightarrow$ only more data will help here


© DESPAIR.COM
LIMITATIONS
Until You Spread Your Wings,
You'll Have No Idea How Far You Can Walk.