# A panorama of machine learning in theoretical particle physics

**Benjamin Fuks**

**LPTHE / Sorbonne Université**

**ILP Day on Machine Learning**

**IHP - 16 November 2018**

# Machine learning in theoretical HEP

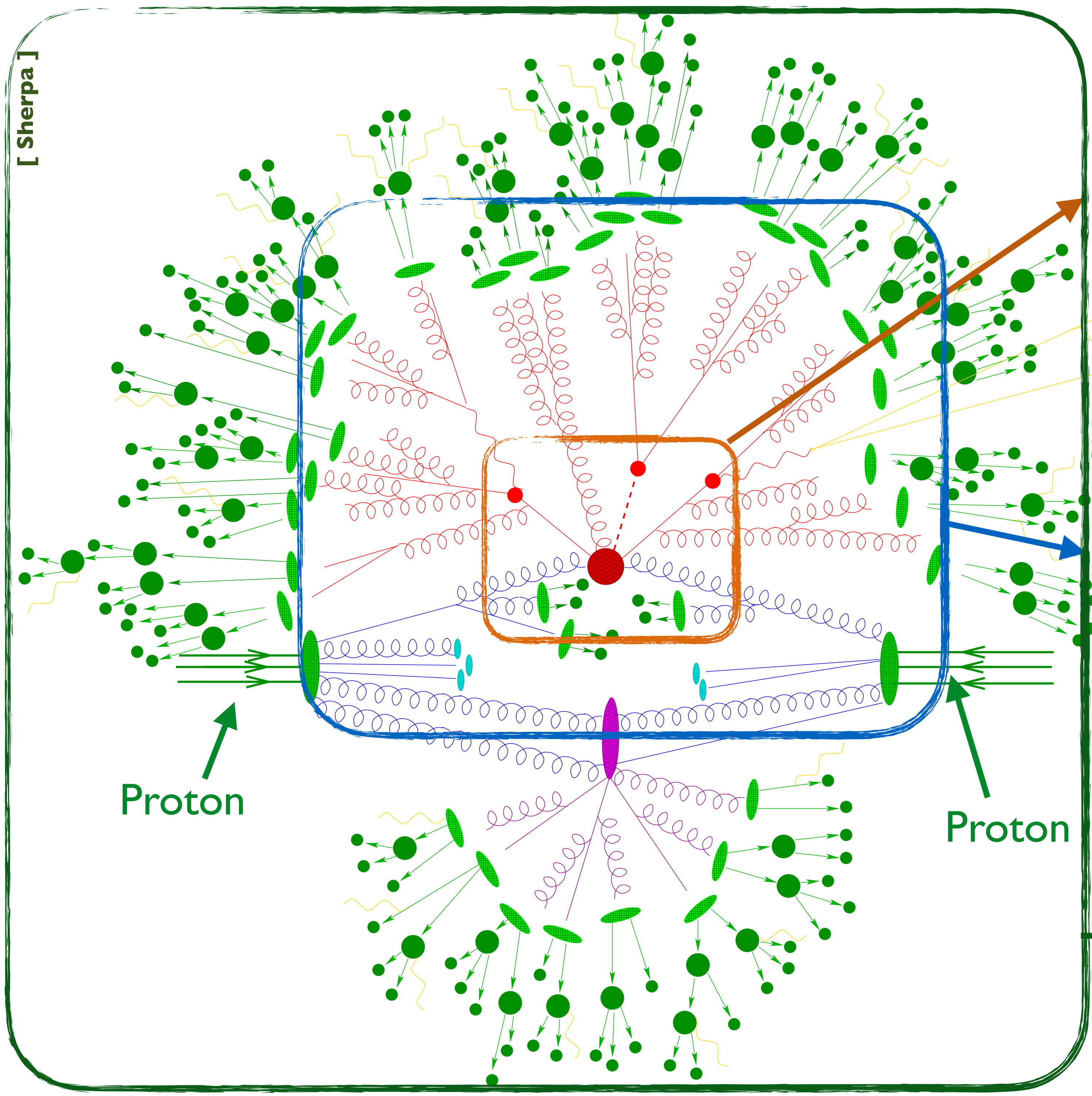
- ✦ Machine learning may seem counterintuitive in theoretical HEP
  - ✤ Theory aims to decode Nature by testing conjectures with data
    - ★Connects observables to the model concepts
  - ✤ Contrasts with a machine-learning black box model
    - ★ Hard (or impossible) to get a physical interpretation

- ✦ Machine learning can however help
  - ✤ When heavy calculations are involved (needs for large computing power)
  - ✤ For the determination of the free parameters of a model

- ✦ Few topics for which machine learning is (or will be) part of the routine
  - ✤ Jet physics
  - ✤ Parton densities
  - ✤ Parameter space scans
  - ✤ Much more… (not covered here): phase space integration, lattice gauge TH

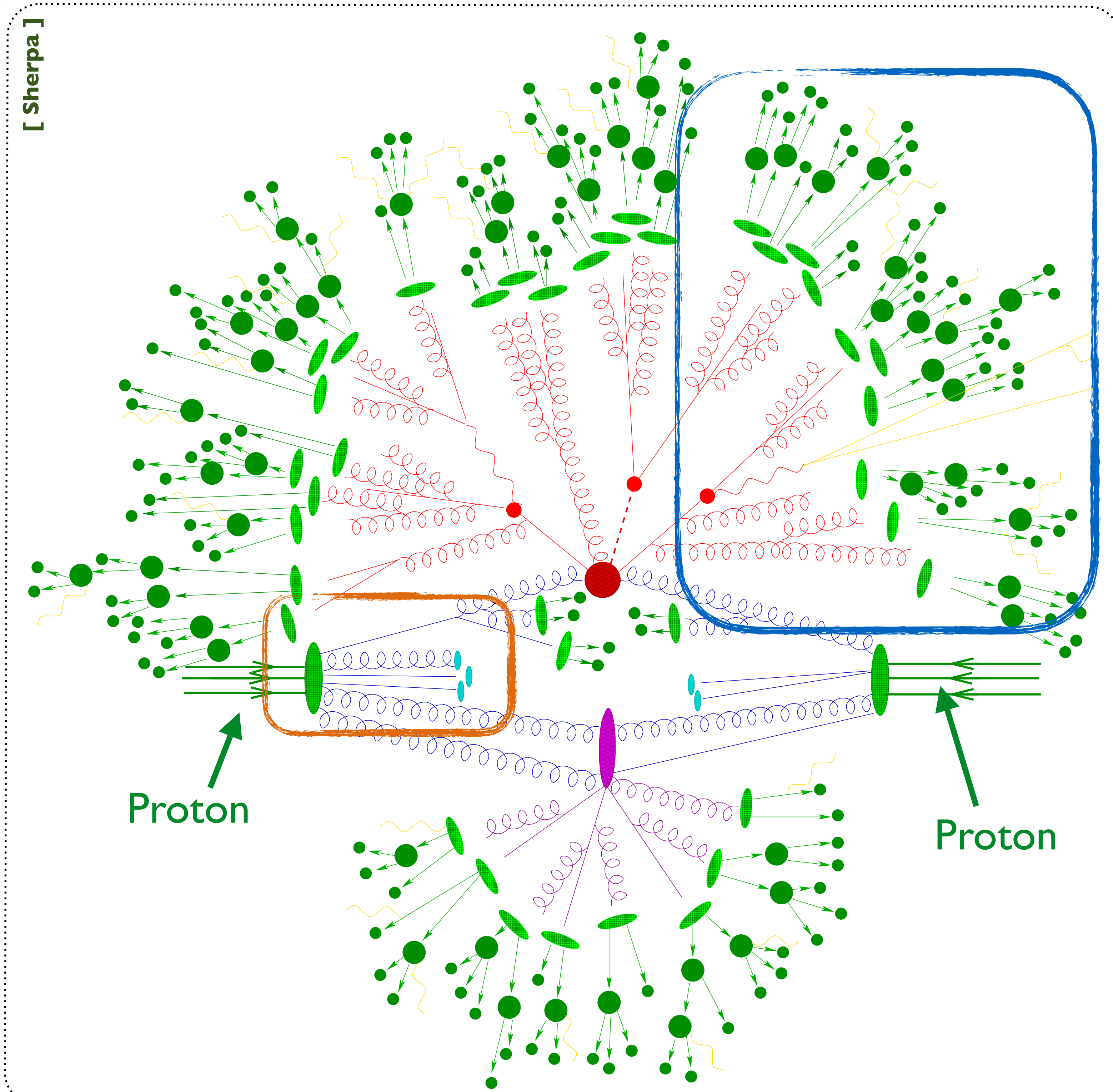
## Jet physics

# Deciphering a proton-proton collision




- ✦ Hard process
  - ✤ Depends on the physics model (SM, BSM,…)
  - ✤ Perturbative QCD
  - ✤ Scale: 0.1−1 TeV
- ✦ Parton showering
  - ✤ Process-independent and universal (QCD)
  - ✤ Scale: 1−100 GeV
- ✦ Hadronization
  - ✤ Model-based, universal
- ✦ Underlying event
  - ✤ Model-based
  - ✤ Energy dependent
- ✦ Scale: sub-GeV

# Deciphering a proton-proton collision




## ✦ QCD processes

- ✤ Complex & image-like
- ✤ Great for ML
- ✤ Jet physics

## ✦ Parton densities

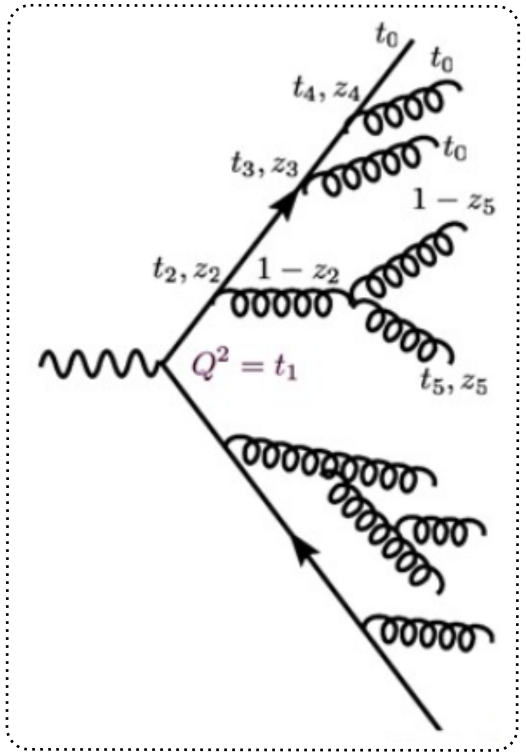
≈ Probability to get a quark or gluon with a given momentum fraction

# Parton showers

✦ Accelerated charges radiate
- ✤ Large momentum transfers ≡ lot of radiation

✦ QED
- ✤ Electrically-charged particles radiate photons
- ✤ Photons can split into a (charged) fermion-antifermion pair

✦ QCD is similar, but from the color-charge standpoint
- ✤ Quarks can radiate gluons
- ✤ Gluons can split into a quark-antiquark pair
- ✤ QCD is non-Abelian: gluons can split into a gluon pair




✦ Highly energetic colored particles radiate
- ✤ Each parton is dressed with an arbitrary number of partons (multiple radiation)
  - ➢ Radiated partons also radiate
- ✤ One ends up with a cascade of radiations ➢ parton showers

✦ Hadronization
- ✤ Free partons do no exist in nature ➢ hadronization

# Hadronization

## ✦ Generalities

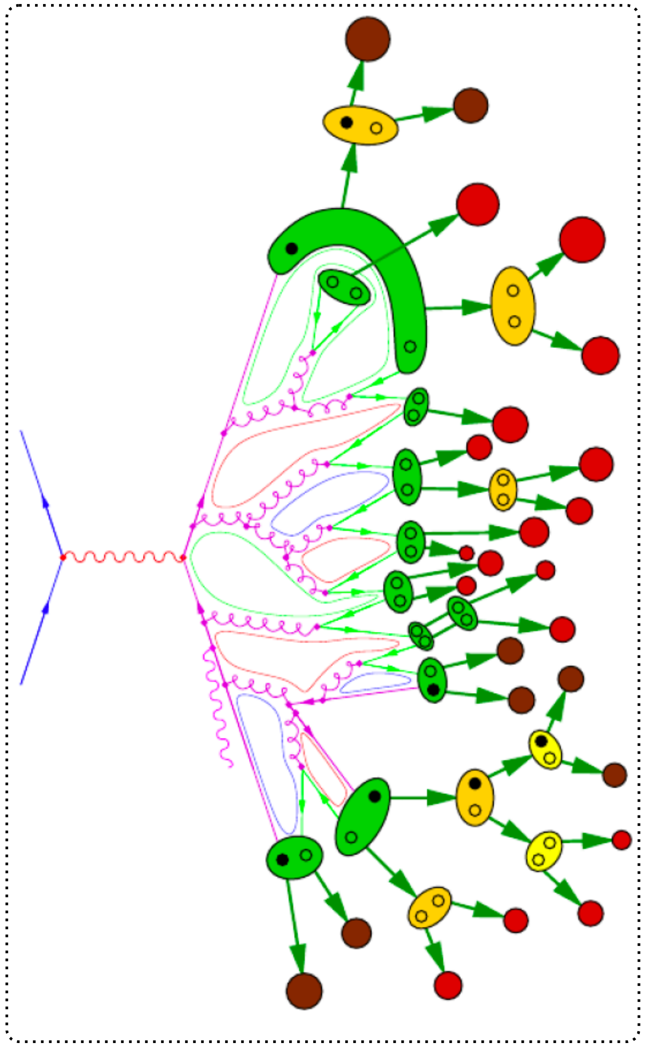
- ✤ Parton showering is not valid at a scale of about 1 GeV
- ✤ Perturbative QCD breaks down
- ✤ Non-perturbative models to get hadrons out of partons
  - ★ Cannot be computed from first principles

## ✦ Two main hadronization models

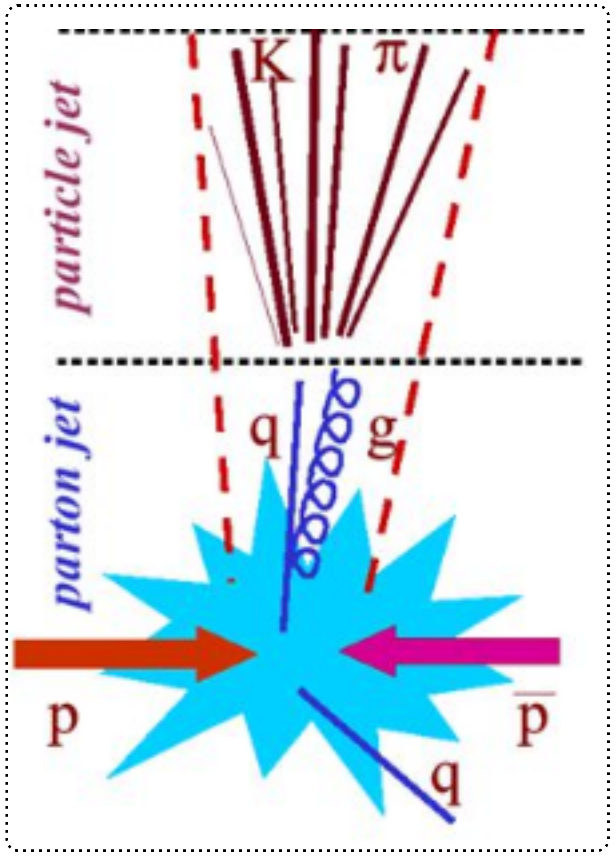
- ✤ The Lund string model [ Andersson, Gustafson, Ingelmanm & Sjöstrand (PR'83) ]
- ✤ The cluster model [ Webber (NPB`84) ]

## ✦ Hadrons then decay

- ✤ Thousands of different channels
- ✤ Relies on form factors
- ✤ Large uncertainties
- ✤ Significant impact on the event shape



# Jet reconstruction




✦ Evolution from one initial parton

✤ Parton showering into many partons

✤ Hundreds of hadrons decaying into each other

✦ Jet reconstruction

✤ Hadrons are clustered into jets

✤ Jets can be matched with the initial partons

✤ Study of the structure of the jet

★ Knowledge on the initial parton giving rise to it

✦ Motivations

✤ Probing the dynamics of the Standard Model in the high-energy regime

★ Where radiation is more collimated

★ Jet (sub)structure can be used to get to initial W-bosons or top-quarks

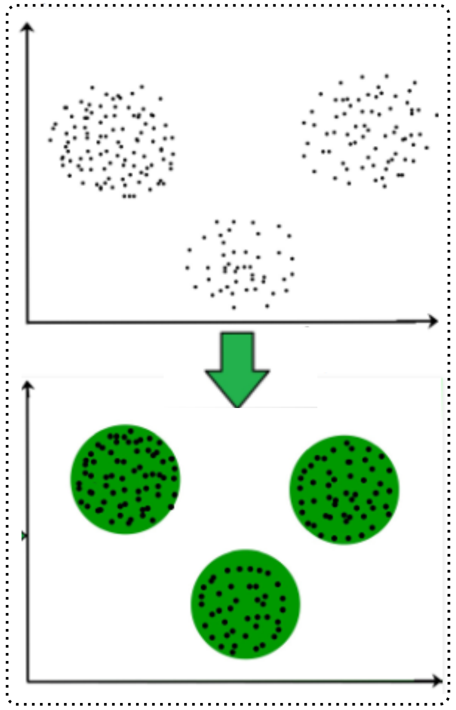
✤ Getting a hand on new phenomena (expected to occur at high energies)

# Machine learning and jet reconstruction

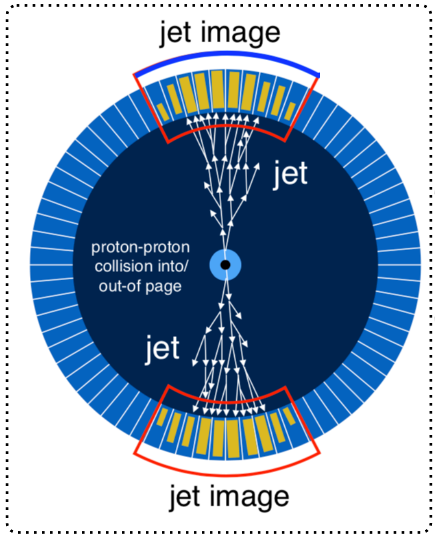
[ Larkoski, Moult & Nachman (2017) ]

- ✦ Jets are defined by a clustering algorithm
- ✦ Clustering ⇔ unsupervised machine learning
  - ✤ The HEP community uses homemade methods
  - ✤ Standard clustering algorithms are not reliable
    - ★ Jets have a specific physical meaning
    - ★ The algorithm must satisfy few key (physical) properties

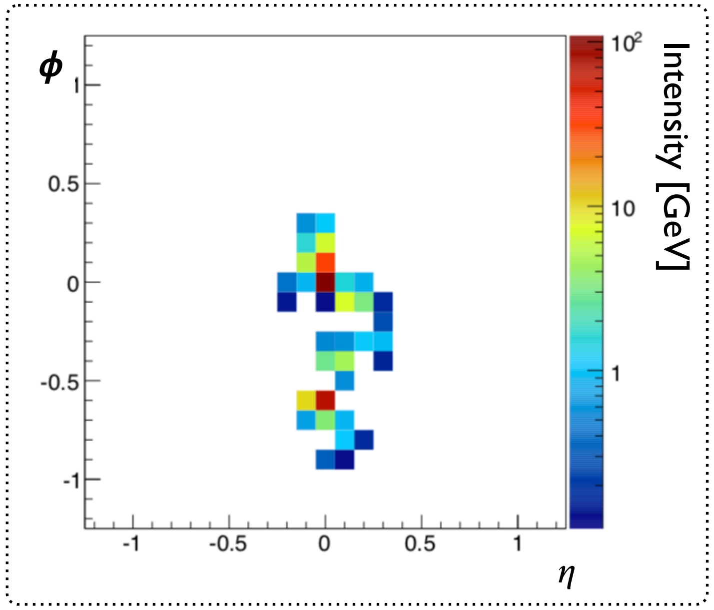
**Jet clustering does not rely on ML**



- ✦ The jet image can be seen as a 2D representation of the radiation pattern




- ✤ Azimuthal angle: $\phi$
- ✤ Pseudorapidity: $\eta = -\log\tan\frac{\theta}{2}$
- ✤ Advantages
  - ★ Direct visualization of physics
  - ★ Benefits from image processing

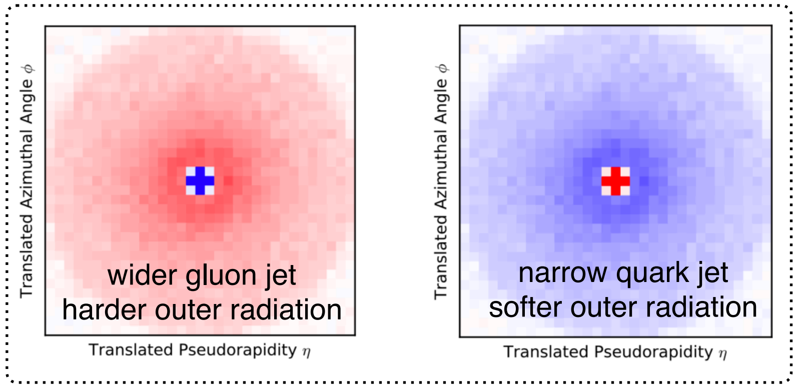


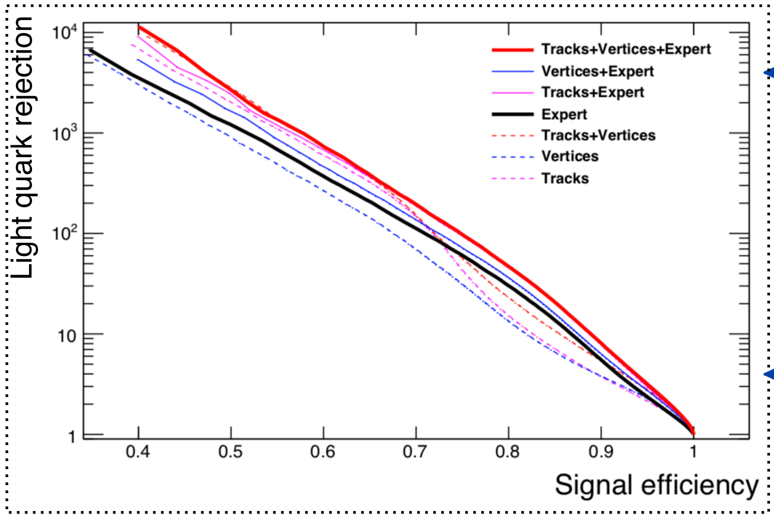
# Jet classification

[ Larkoski, Moult & Nachman (2017) ]

- Jet classification consists in tagging a jet a originating from a given particle
- Perfect playground for supervised learning
  - One can simulate large samples of simulated data (*i.e.* particle collisions)
  - Known process nature and final state particle content
    - ➢ Inputs: the jet constituents

- Fixed size representation ML: image classification
  - ★ *e.g.* Convolutional networks ➢ jet images
  - ★ The particle energy is mapped to a pixel intensity
    - ➢ 10-15% occupancy: contrasts with usual methods
    - ➢ Pseudorapidity-azimuthal maps
  - ★ W-boson or top-quark jets, quark/gluon separation



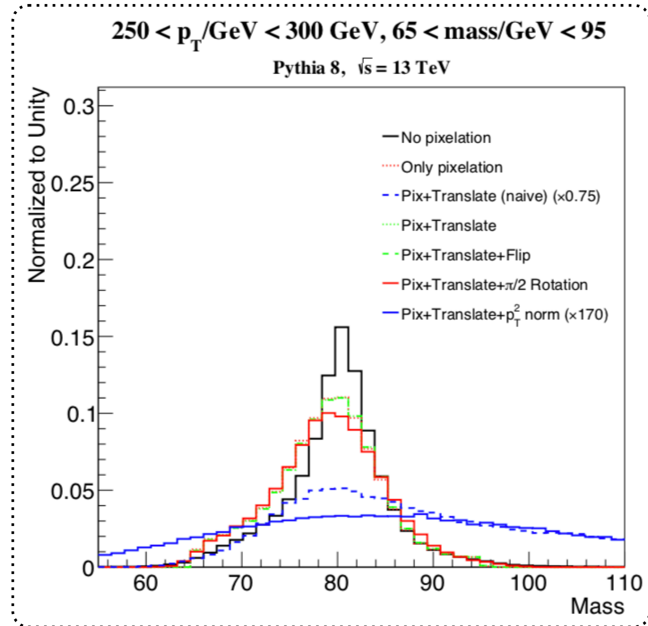

- Variable size representation ML: language processing
  - ★ *e.g.* Recurrent networks ➢ jet radiation pattern (sentences)
  - ★ No unique ordering of the words ($p_T$, $k_T$, *etc.*)
  - ★ Heavy flavor-tagging, quark/gluon separation, *etc.*
- Combining may be the key
  - ★ Tracking (much finer) and calorimetry (naturally pixelized)

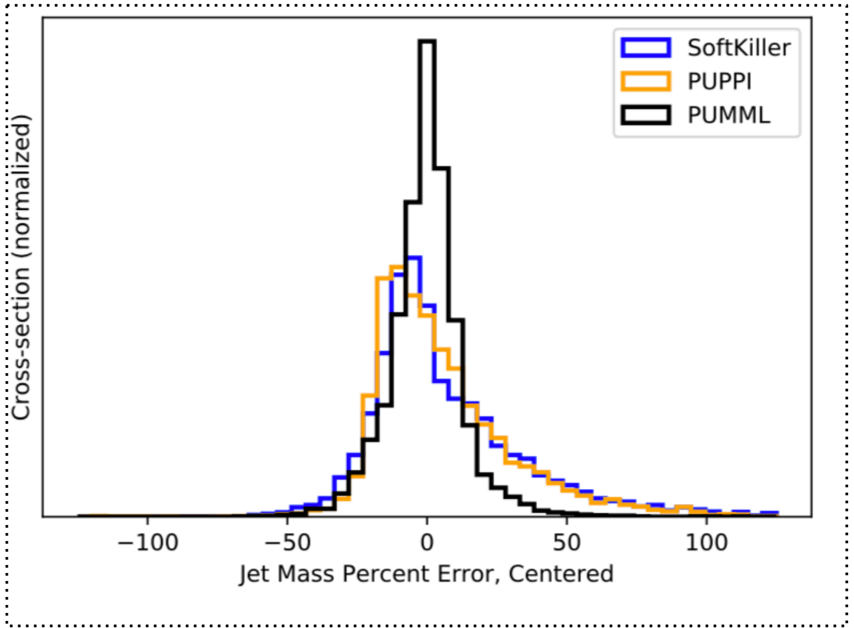
# Some other applications and issues

[ Larkoski, Moult & Nachman (2017) ]

## ✦ Standard issue: preprocessing

- ✤ Can alter the physics and thus the conclusions
  - ★ Centering/rotating the image on the leading subjet (an energy-based pixel intensity is not boost-invariant)
  - ★ Normalization add random noise
  - ★ The loss of information can alter observable spectra
- ✤ Preprocessing included in the architecture




## ✦ Pile-up mitigation

- ★ Pile-up events are mostly diffuse noise
- ★ Convolutional neural networks to get rid of it
- ★ Various methods to remove the pile-up

## ✦ Jet generation (to speed up event simulation)

- ★ Use of generative adversarial networks
- ★ Generation of jets of a given type
- ★ Usually faithfully reproduce the properties
- ★ Hard to populate all possible configurations

## Parton densities

# Predictions at the LHC (using QCD)

✦ Distribution of an observable $\omega$: the QCD factorization theorem

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\omega} = \sum_{ab} \int \mathrm{d}x_a\, \mathrm{d}x_b\, \mathbf{f_{a/p_1}}(x_a;\mu_F)\, \mathbf{f_{b/p_2}}(x_b;\mu_F)\, \frac{\mathrm{d}\sigma_{\mathbf{ab}}}{\mathrm{d}\omega}(\ldots,\mu_F)$$

- Long distance physics: the parton densities
- Short distance physics: the differential parton cross section $d\sigma_{ab}$
- Separation of both regimes through the factorization scale $\mu_F$
  - ★ Choice of the scale ➢ theoretical uncertainties

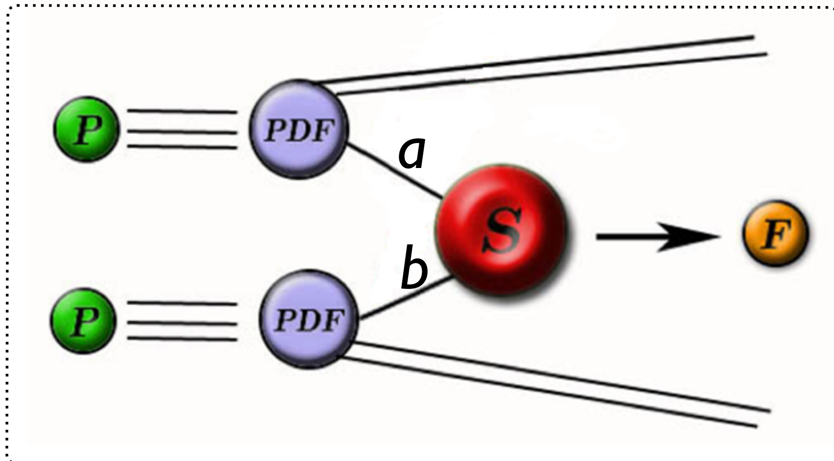
✦ Short distance physics: the partonic cross section

- Calculated order by order in perturbative QCD: $d\sigma = d\sigma^{(0)} + \alpha_s\, d\sigma^{(1)} + \ldots$
  - ★ The more orders included, the more precise the predictions
  - ★ Truncation of the series and $\alpha_s$ ➢ theoretical uncertainties

# Parton densities

✦ Long distance physics: parton densities

✤ Relate the hadrons to their content



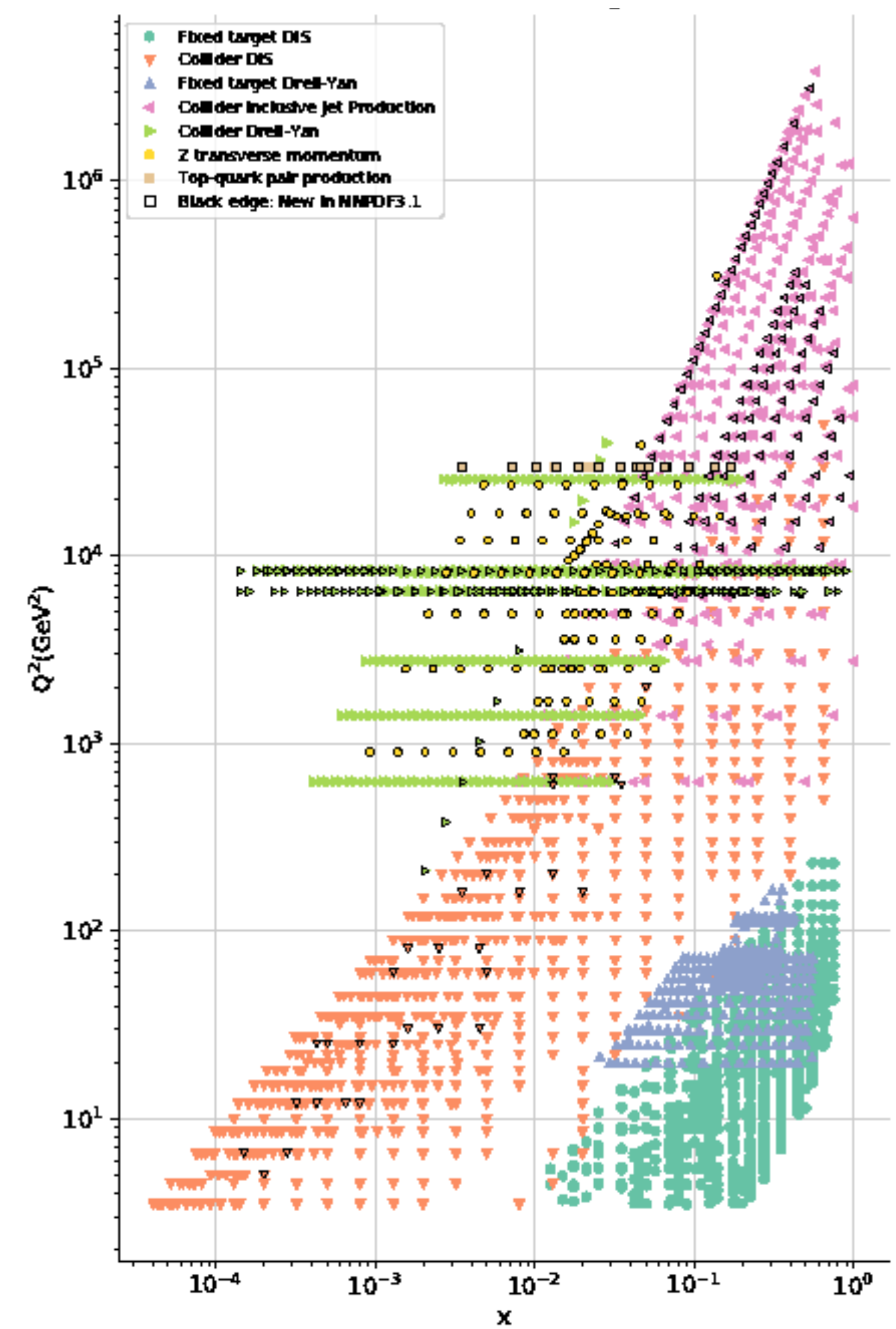

✤ Depend on the momentum fraction *x* of the parton in the proton

✤ Depend on a scale Q

✤ Fitted from experimental data [in some kinematical regimes (x,Q)]

✤ Evolution driven by QCD (DGLAP/BFKL)

[ NNPDF 3.0 (EPJC'17) ]

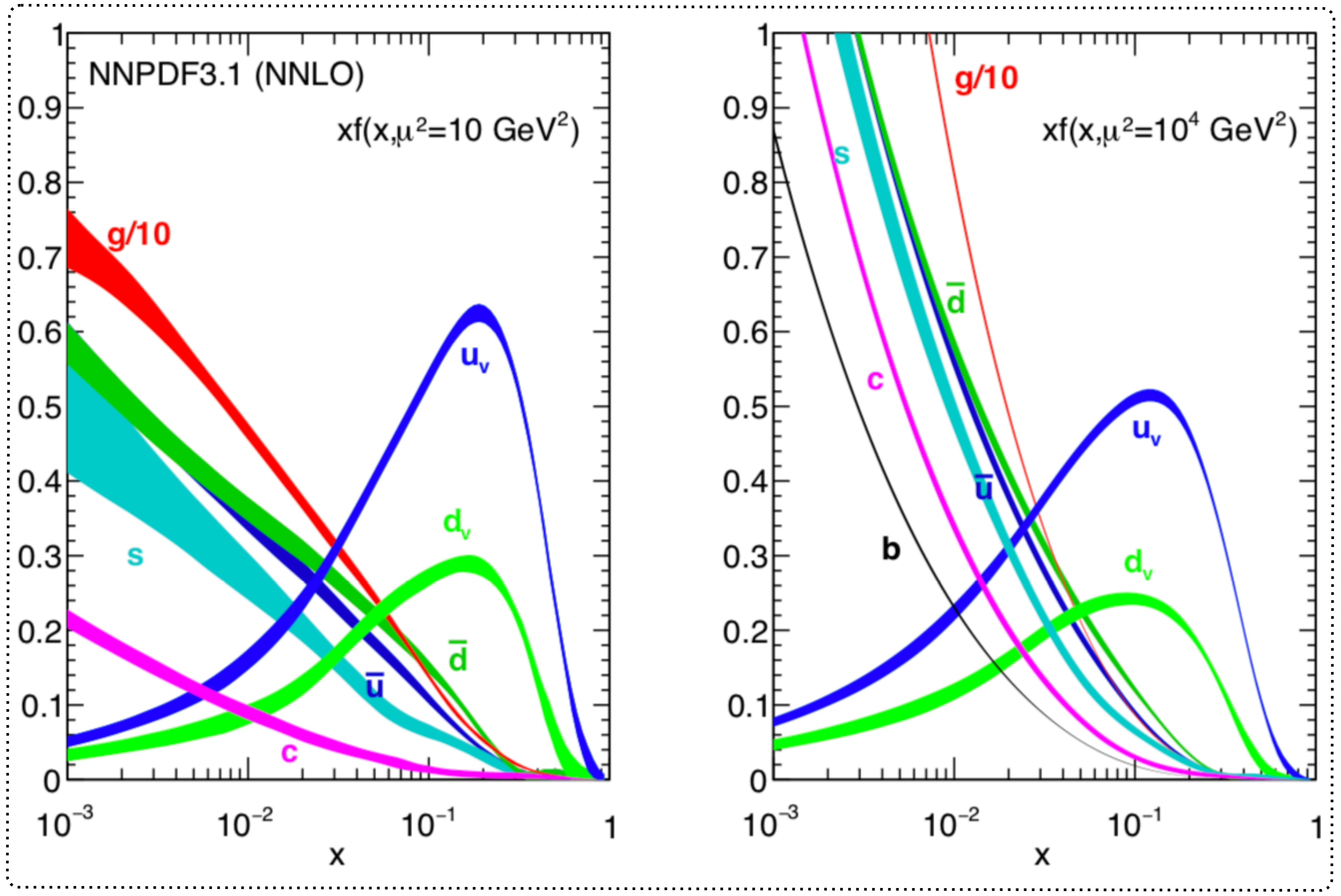


# NNPDF: parton densities with neural nets

[ Ball *et al.* (JHEP`14) ]

✦ Two-fold goals: obtain the best fit of data together with the uncertainties

- ✤ Generation of artificial data
  - ★ From experimental measurements and their (correlated) errors
  - ★ Many sets are generated, consistent with the covariance matrix
- ✤ Fit of generated data with a NN
  - ★ Most accurate theoretical predictions are used
  - ★ The NN ≡ the parton density
  - ★ A genetic algorithm is used
  - ★ New replicas are generated from one generation to the next
- ✤ Predictions obtained after getting statistical estimators (means, quantiles, …)
- ✤ Complexity: data taken over several decades

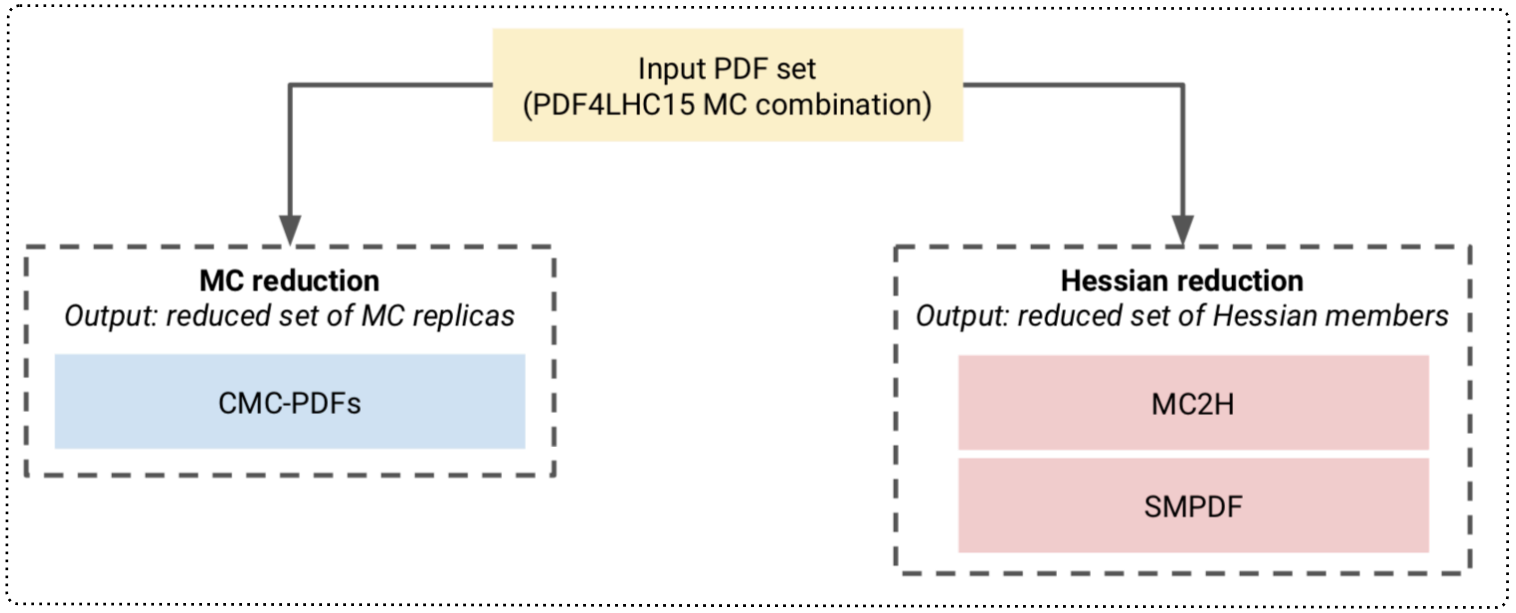


# PDF4LHC: parton densities for the LHC

[ Butterworth *et al.* (JPhysG`16) ]

## ✦ Which set of parton densities to use for calculations

- Many groups perform global fits
- How to combine individual sets?
  - ➢The PDF4LHC recommendations




## ✦ Methodology

- Input parton densities
  - ★ MC determination from existing densities
  - ★ Same dataset
  - ★ Same QCD properties
- Reduction of the number of replicas
  - ★ 900 to 100
  - ★ e.g. based on NN and genetic algorithms

## Parameter space scans

# Reinterpreting LHC physics analyses

- ✦ Exploit the full potential of the LHC (for new physics)
  - ✤ Priority #1 of the European strategy for particle physics
  - ✤ Designing new analyses to probe new ideas — Prospectives (based on MC simulations)
  - ✤ Recasting LHC analyses to study models not considered — The LHC legacy

- ✦ LHC data has been collected with significant human and financial efforts
  - ✤ Important for on-going analyses (within popular theoretical contexts of today)
  - ✤ Important for future opportunities (within future scientific contexts)
- ✦ Data preservation in high-energy physics is mandatory [ Kogler, South & Steder (JPCS'12) ]

- ✦ Related tools need to be supported by the entire community [ Kraml *et al.* (EPJC'12) ]
  - ✤ Both theorists and experimentalists
  - ✤ Allowing for the reinterpretation of the LHC analysis results

# LHC Recasting

✦ There are plethora of new physics realizations that deserve to be studied

✤ Experimentalists cannot study all the options

✤ Simulating the detector response of ATLAS and CMS

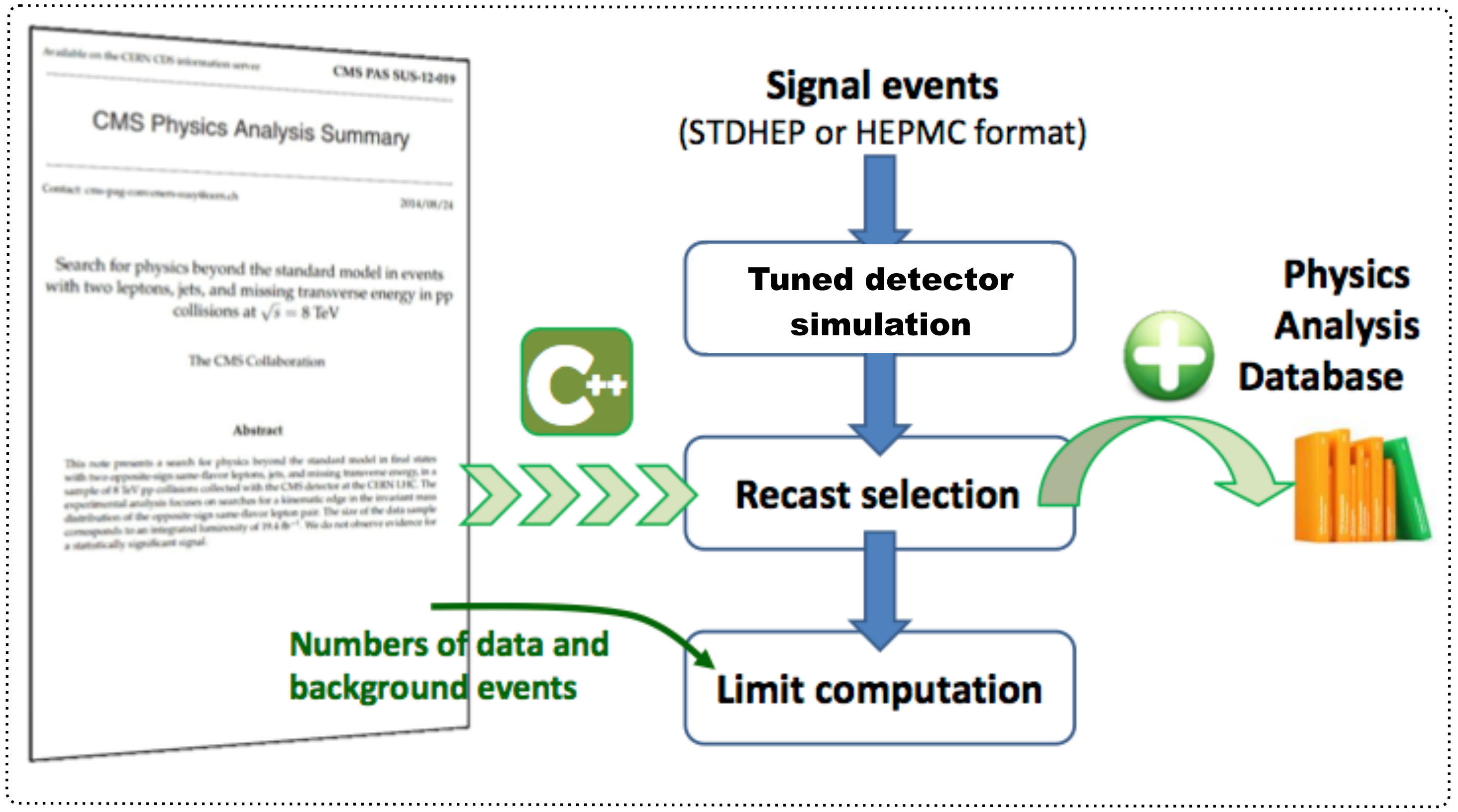
✤ Relying on public frameworks where LHC analyses can be easily implemented

✦ LHC simulations take time

✤ Efficiency ⇔ ML
No need for LHC simulations

✤ Including other constraints

✤ Heavy parameter space scans



# Making the situation better with ML

✦ A simple example: probing dark matter with cosmology and colliders

✤ Simplified models are very powerful in particle physics

★ Bottom-up frameworks with few additional particles and parameters

★ Characterization of all relevant processes for DM production, scattering and annihilation

[ Bertone *et al.* (JCAP'18) ]

✦ Constraints on the models

✤ LHC searches for all new particles

✤ DM relic abundance

✤ DM indirect detection bounds (searches for the product of DM annihilations)

✤ DM direct detection bounds (searches for DM in observatories on Earth)

✦ Interpretation of the results

✤ Bottleneck: LHC likelihoods for a given parameter set

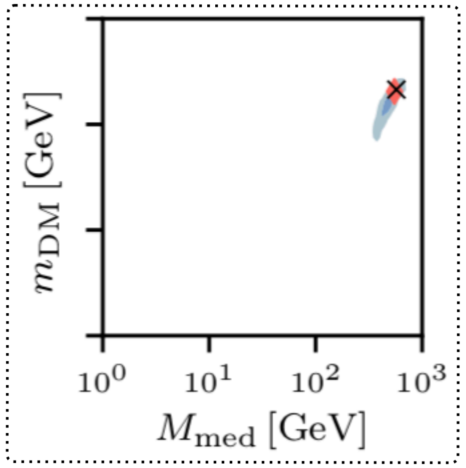
★ Computationally expensive

★ To be evaluated 10k – 100k times even for a four-parameter model

✤ Using (supervised) machine learning to accelerate it

★ Mapping of parameter space points to signal region populations

★ No need for simulations anymore




✦ Other works exist, on simplified supersymmetric models (SCYNet, SUSY-AI)

[ Bechtle *et al.* (EPJC'17); Caron *et al.* (EPJC`17) ]

# Summary

# Machine learning in theoretical HEP

- ✦ Theory aims to decode Nature by testing conjectures with data
  - ✤ May seem to contrast with machine-learning
- ✦ Machine learning can however help
  - ✤ When heavy calculations are involved (need to computing power)
  - ✤ For the determination of the free parameters of a model

- ✦ Few topics for which machine learning is (or will be) part of the routine
  - ✤ Jets physics has seen intense activities involving machine learning
  - ✤ Parton density fits can be evaluated thanks to machine learning methods
  - ✤ New physics parameter space scans
  - ✤ Much more…

- ✦ Discussions
  - ✤ Any other TH problem that could be tackled with machine learning
    - ★ See also the next talk (many EXP issues are common to the TH community)
  - ✤ Any better way to address the currently-addressed problems
  - ✤ Any wrong approach?