Theoretical physics, Machine Learning and Bioinformatics

Jesús Urtasun Elizari

Milan, March 2021

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 740006.

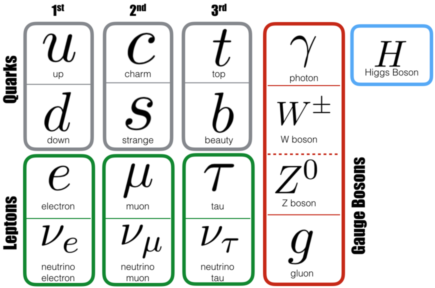
Outline

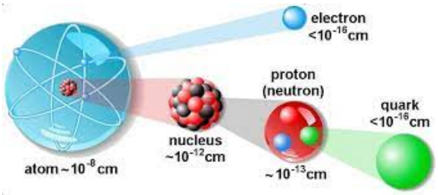
1. QCD in a nutshell
   - The fundamental interactions
   - Exploring matter at the small scales
   - Hadronic physics and the LHC
2. Machine Learning for particle physics
   - Machine Learning algorithms
   - NN for building predictions
3. Bioinformatics
   - Applying data sciences to life sciences
4. Summary

Quantum Chromodynamics in a nutshell

QCD in a nutshell

The Standard Model





Quantum Field Theory describing physics at the TeV scale $\rightarrow$ less than a fermi!

1. Fermions (quarks and leptons) composing matter
2. Bosons mediating interactions
3. Scalar Higgs field generating mass

Quantum Chromodynamics is the theory describing the strong interactions

QCD in a nutshell

Explore the strong interactions

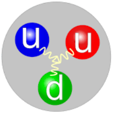
How to explore proton's inner structure?

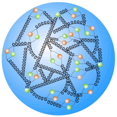
- Point-like projectile on the object $\longrightarrow$ DIS
- Smash the two objects $\longrightarrow$ LHC physics

"A way to analyze high energy collisions is to consider any hadron as a composition of point-like constituents $\longrightarrow$ **partons"** R.Feynman, 1969

QCD in a nutshell

Parton Distribution Functions





- Hadrons made of constantly interacting partons $\longrightarrow$ non perturbative physics
- Parton distribution functions (PDFs) are required for the precision era of the LHC
- PDFs can not be predicted yet not measured $\longrightarrow$ extracted from data via Machine Learning algorithms

Machine Learning for particle physics

Machine learning

What is Machine Learning?

1. Machine Learning algorithms are a subset of Artificial Intelligence (AI) algorithms
2. Used to solve *complex* tasks like classification, regression and pattern recognition
3. Rely on comparison with data $\longrightarrow$ Learning

The N3PDF project

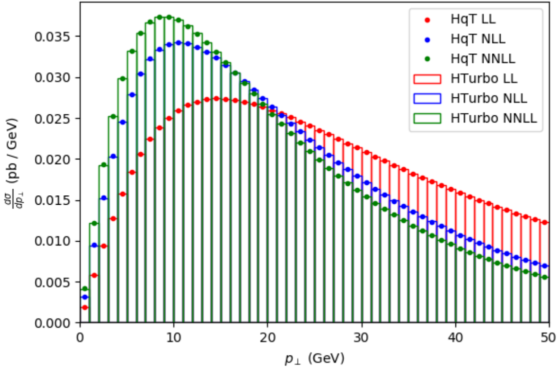
General structure of n3fit

Parton Distribution Functions (PDFs) can not be predicted or measured
PDFs need to be extracted from data!

- Use TensorFlow and Keras to determine the PDFs using neural networks
- Use Stochastic Gradient Descent n3fit replacing primitive fitting algorithms
- See paper by S.Carraza - J.Cruz-Martinez
  "Towards a new generation of parton densities with deep learning models",
  https://arxiv.org/abs/1907.05075
- Operator Implementation in TF - Urtasun-Elizari et al.
  "Towards hardware acceleration for parton densities estimation",
  https://arxiv.org/abs/1909.10547

The HTurbo project

Comparison HRes and HqT - all orders



- Older codes (**HRes**, **HqT**) need 3 days to produce NNLL distribution
- 3 minutes with **HTurbo**! ✓
- Agreement up to NNLL $\longrightarrow$ ready for N^3LL

Bioinformatics
and
data science for life sciences

Bioinformatics

Comoputer sciences

- C++ and computer sciences
  - [https://github.com/JesusUrtasun/CppCourse]
- Machine Learning
  - [https://github.com/JesusUrtasun/MLcourse]
- Data Frames (Numpy, Pandas) and R programming
  - [https://github.com/JesusUrtasun/Bioinformatics]

- NGS data analysis
- Statistics and data analysis of BAM files with R
- **Rsubred** and **Dseq2** packages
- Fast learner!

Bioinformatics

DNA and RNA sequencing data

- NGS data analysis
- Statistics and data analysis of BAM files with R
- **Rsubred** and **Dseq2** packages
- Fast learner!

Summary & Conclusions

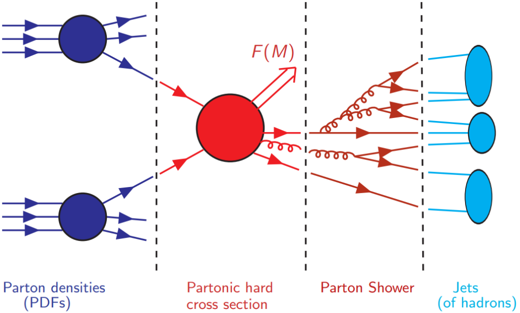
1. Precise knowledge of sub-nuclear interactions are required towards the precision era of the LHC
2. Machine Learning models provide a robust way for PDFs determination optimized through operator implementation in TF
3. We develop a numerical code **HTurbo**, implementing $q_{\perp}$ resummation for Higgs boson production, which is faster than any of the existing codes
4. Experience with Python and R for NGS data, and still looking to improve!

Thank you!

Back up

Hadronic collisions

Hadronic Physics $h_1(p_1) + h_2(p_2) \to F + X$

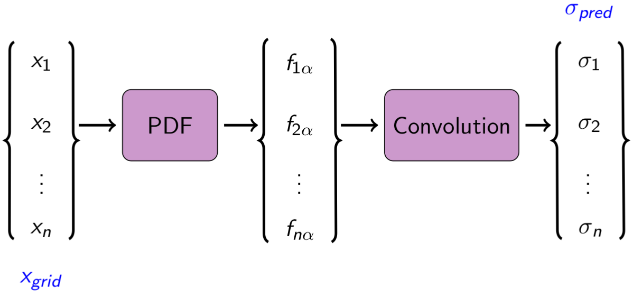


Factorize process as PDFs and partonic (hard) interaction

$$\sigma^{\mathrm{F}}(p_1, p_2) = \sum_{\alpha,\beta} \int_0^1 dx_1 dx_2 \; f_{\alpha/h_1}(x_1, \mu_F^2) * f_{\beta/h_2}(x_2, \mu_F^2) \; * \; \hat{\sigma}^{\mathrm{F}}_{\alpha\beta}(x_1 p_1, x_2 p_2, \alpha_s(\mu_R^2), \mu_F^2)$$

Back up

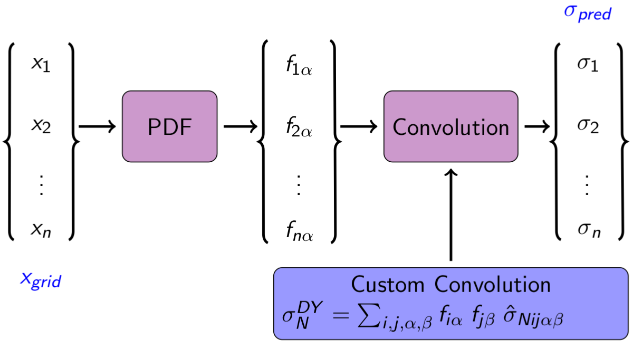
Operator implementation in TF



- Build a NN model to compute σ_{pred} observables from a grid x_i
- Perform χ^2 minimization comparing with data
- Update values of PDF $\longrightarrow$ Fit

The N3PDF project

Operator implementation in TF

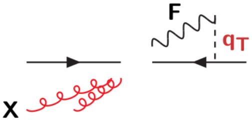


1. TF relies in symbolic computation $\longrightarrow$ High memory usage
2. Implement c++ operator replacing the convolution
3. Further details in Urtasun-Elizari et al.
   "Towards hardware acceleration for parton densities estimation", https://arxiv.org/abs/1909.10547

HTurbo

All order $q_\perp$ resummation

Study the differential $q_\perp$ distribution



$$h_1(p_1) + h_2(p_2) \longrightarrow F(M, {\color{red} q_\perp}) + X$$

$$\int_0^{Q_\perp^2} dq_\perp^2 \frac{d\hat{\sigma}}{dq_\perp^2} \sim c_0 + \alpha_s(c_{12}L^2 + c_{11}L + c_{10}) + ..., \quad \text{being} \quad L = \ln(q_\perp/M^2)$$

$\alpha_S L^2$	$\alpha_S L$	$\cdots$	$\mathcal{O}(\alpha_S)$
$\alpha_S^2 L^4$	$\alpha_S^2 L^3$	$\cdots$	$\mathcal{O}(\alpha_S^2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\alpha_S^n L^{2n}$	$\alpha_S^n L^{2n-1}$	$\cdots$	$\mathcal{O}(\alpha_S^n)$
dominant logs	$\cdots$	$\cdots$	$\cdots$

Truncated fixed order predictions lead to logarithmic enhancement $\alpha_s^n \ln^m(M^2/q_\perp^2)$

All order resummation is needed

HTurbo

Starting point: DYTurbo

$q_\perp$ resummation implemented in numerical codes **HqT** and **HRes** [Catani et al.]
Higher order accuracy require high computation times

Numerical code **DYTurbo** [Camarda et al. ATLAS collaboration, 1910.07049], fast and precise $q_\perp$ resummation and several improvements for Drell-Yan $(h_1 + h_2 \to V + X \to l^+ l^- + X)$

- Goal: set up a numerical code generalizing **DYTurbo** for Higgs boson production
- Goal: extend theoretical accuracy up to N^{3}LL+N^{3}LO and to other processes

Numerical code **HTurbo** [Ferrera, Urtasun-Elizari (in preparation)]
Fast and precise $q_\perp$ resummation and for Higgs boson production

- Old versions as **HqT** and **HRes** in Fortran
- c++ allows for optimization in the integration routines