Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS

Valentina Guglielmi (DESY), Simone Amoroso, Katerina Lipka

Reweighting with a Machine Learning (ML) classifier:

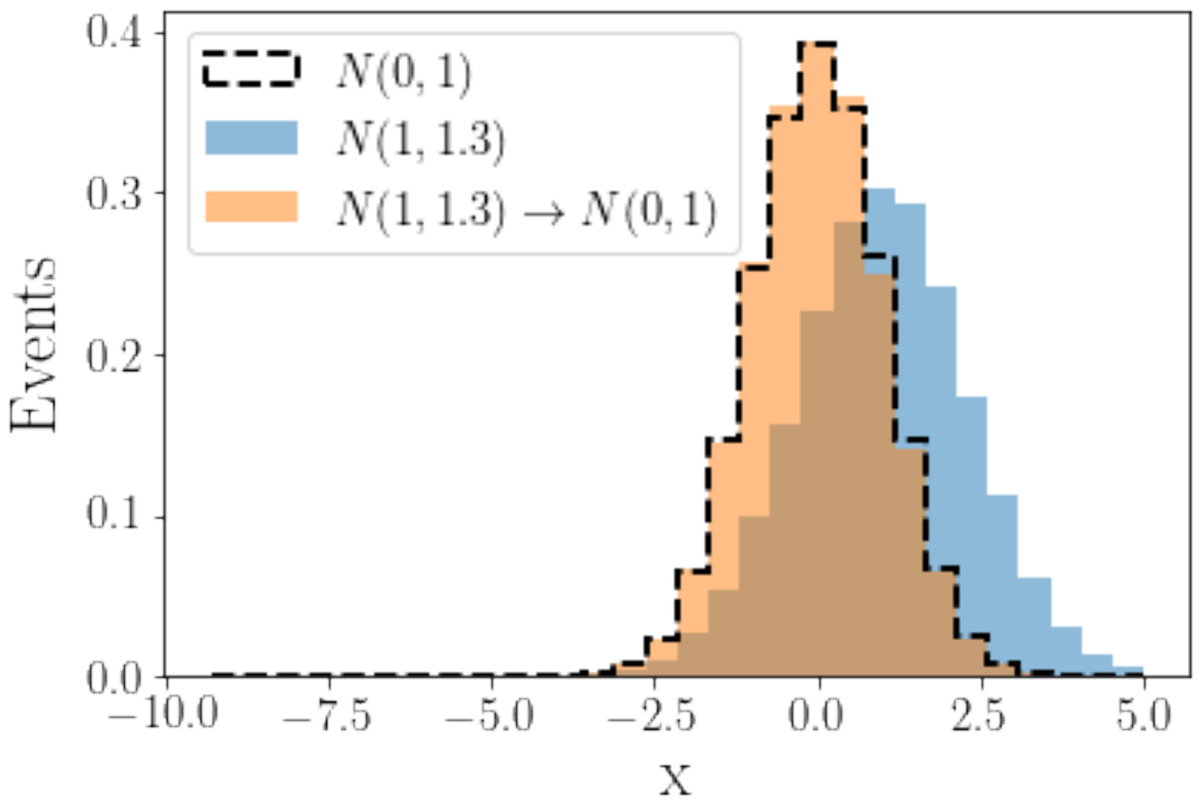
For Large Hadron Collider analyses we need Monte Carlo (MC) samples of simulated events:

- Work-flow process:
  1. Generation of the physics event → cheap (seconds)
  2. Simulation of the detector → expensive (minutes)
- Alternative samples needed to take into account systematic uncertainties

  →High computational cost
- **Reweighting** the nominal sample avoids the need to simulate the detector response multiple times
- A **ML classifier** can be trained to distinguish two simulations and reweight one into another

Benefits ML reweighting (DCTR) [1]:

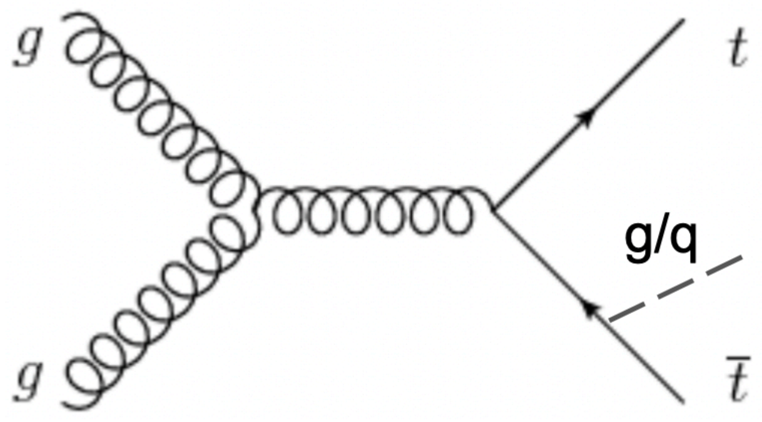
- **Multidimensional** and **unbinned** information
- **Continuous** as function of any MC parameter



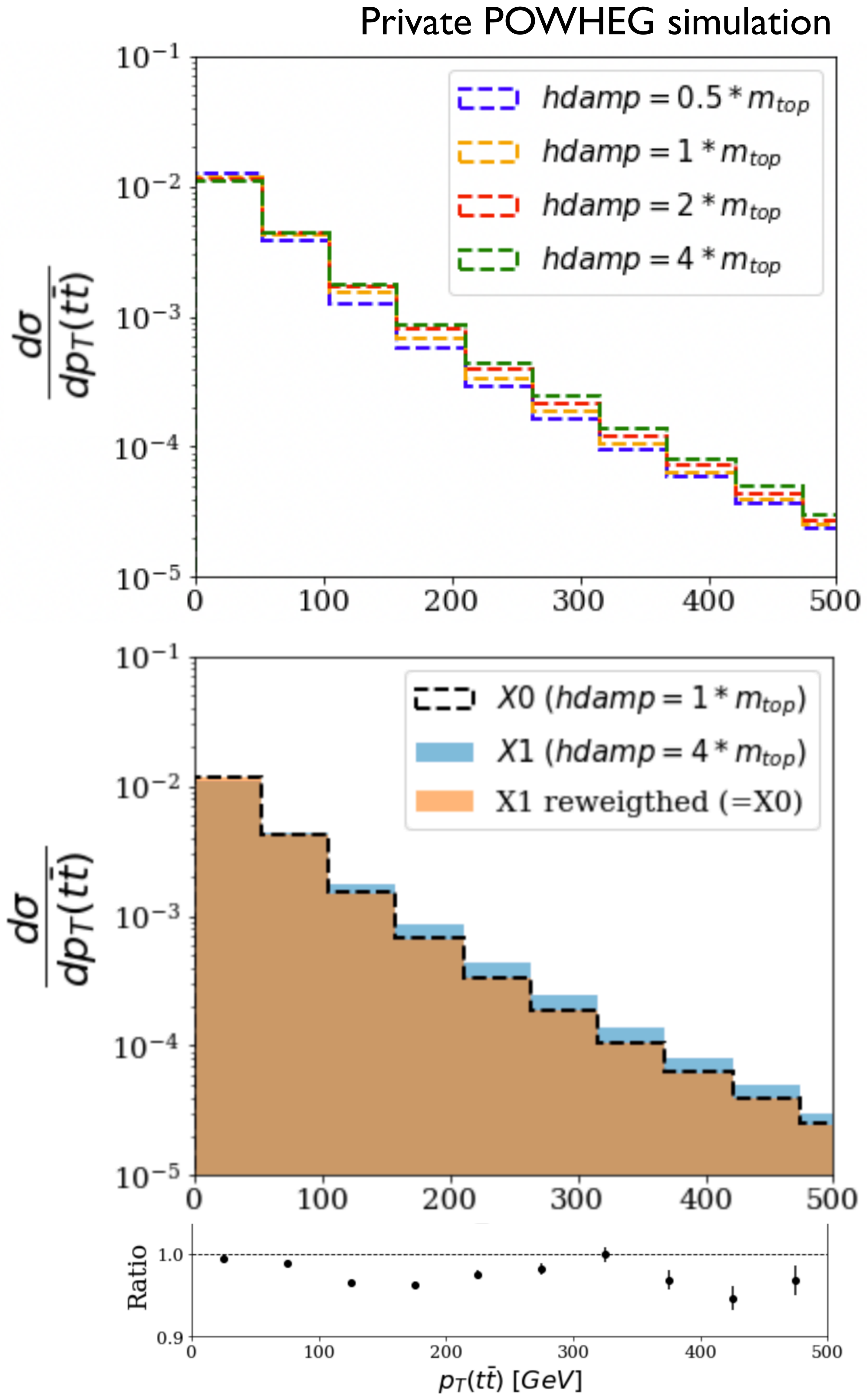
Top analysis application:

Heavy quark process of Powheg MC generator:

- ***Hdamp*** parameter that controls the resummation
- Important systematic in many top quark analyses
- It affects p_T of the $t\bar{t}$ system



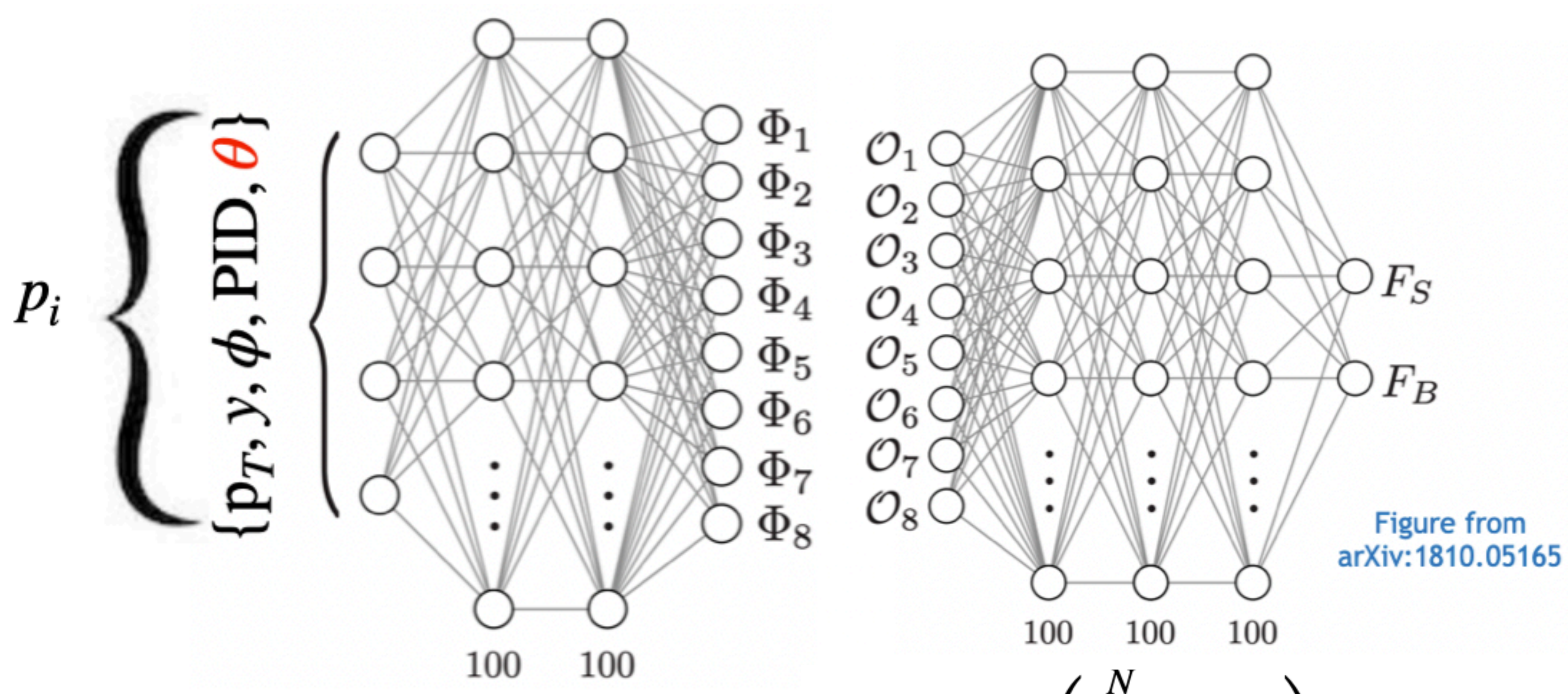
Private POWHEG simulation



Neural Network architecture:

Particle Flow Network [2]:

- Using **parton level information** as inputs: 4-vector (p_T, y, ϕ, m) and particle kind (PID: top, antitop, gluon/quark)



PFN composition of two NN: F and Φ

$$F(\{p_i\}) = \left(\sum_{k=1}^{N} \Phi(p_i)\right)$$

Results and conclusions:

The method found to work very well!

- **Good agreement** up to 1 TeV of $p_T(t\bar{t})$
- **Differences** between original sample and reweighted one of **~5%**

Next steps:

- Extend the study to other parameters and apply **simultaneous reweighting**
- **Integrate** the method in the MC production of the experiment CMS
- **Tuning of MC parameters** with full event information at detector level

References:

1. https://arxiv.org/abs/1907.08209
2. https://arxiv.org/abs/1810.05165