Falcon Simulator

Sergei V. Gleyzer, University of Florida
Harrison B. Prosper, Florida State University
Omar Zapata Mesa, University of Antioquia

Outline

- Motivation
- What is Falcon
- Results
- Plans and Outlook

Motivation

- LHC is taking data again
  - in Discovery mode
  - until compelling evidence of new physics is found
    - Compare thousands of experimental results with theoretical predictions of thousands of models
      - Make quantitative statements about their validity

Experiment & Theory

Compare experimental results with theoretical predictions

- **Two approaches:**
  - **Fold theoretical predictions with detector effects and compare folded predictions with experimental results**
  - **Unfold detector effects from experimental results and compare unfolded results directly with predictions**

Folding vs. Unfolding

- Both have pros and cons
  - Folding is typically preferred
    - technically easier when experimental results are multi-dimensional
  - Price to pay:
    - Computation time
    - Codes generally not publically available (or require knowledge and expertise not available outside an experimental collaboration)

Folding

Approximate Multidimensional Function:

$$p(rparticles \mid \theta) = \int R(rparticles \mid particles) H(particles \mid partons) P(partons \mid \theta)$$

- **Probability density** **to observe collection of reconstructed particles (rparticles) for point θ in given model's parameter space**

Folding

Approximate Multidimensional Function:

$$p(rparticles \mid \theta) = \int R(rparticles \mid particles) H(particles \mid partons) P(partons \mid \theta)$$

- **P(partons | θ): parton level theory prediction**
- **H(particles | partons): parton to particle level mapping or hadronization**
- **R(rparticles | particles): detector response**

Matrix Element Methods

Approximate this function with matrix element methods

$$p(rparticles \mid \theta) = \int R(rparticles \mid particles) H(particles \mid partons) P(partons \mid \theta)$$

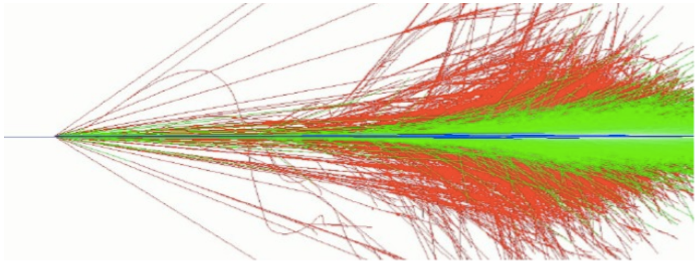
- **Requires highly parallel computing**
- **Empirical functions that approximate detector response don't fully capture non-Gaussian effects**
  - **For ex. CMS jet response function is non-Gaussian**
- **For accurate detector effects:**
  - **Monte Carlo simulation**

Monte Carlo Simulation

If you need to simulate hundreds of thousands events for 1k-100k points in a model's parameter space

- **You are generally out of luck (computationally)**

- **Or perhaps, not?**



Single hadronic shower

Fast Simulation

These difficulties spurred development of fast detector simulators which, like the matrix element method, approximate the detector response (R) parametrically

- **Delphes, J. de Favereau et al, JHEP 02 2014 057, a good example of state of the art**
  - **Parametrize detector response**

Delphes

- Start with simulated events at particle level
- Replace Monte Carlo simulation of detector with random sampling from R

Example: Jets

- – Apply resolution or smearing function to genjets
    - jets reconstructed at particle level
- – Cluster on calo-towers
- – Speed of event generation
    - Goes up 3 orders of magnitude for HL-HLC type pileup (from 1ms/event@PU0 to 1s/event @PU150)

Parametric FastSim

- Need to **hand-code the form** of detector response function
  - Detectors **change**
  - Experimental conditions **change**
  - non-Gaussian effects become important
  - Must **re-code** the response function
- Response function form is **different** from experiment to experiment
- Fast enough?

Non-Parametric FastSim

- **Not a NEW IDEA**
  - **Successfully applied at CDF, D0, HERA and others**
    - **First by Rajendran Raja in late 80s-1990, then B. Knuteson et al. (implementation called Turbosim)**
- **Falcon is a modern implementation of its non-parametric predecessors**
  - **Using novel ideas and programming methods**

Non-Parametric FastSim

- Simulator that **does not require** hand-coded detector response
- Learns the **response function**
  - Extracts from millions of fully-simulated events already available
  - Extremely fast to apply
    - generate events at the order > 10^4/s
      - Independent of pileup

Speed vs. Accuracy

- **Typical dilemma**
  - **Extremely fast → sacrifice accuracy**
  - **Extremely accurate → sacrifice speed**
    - **Find a good balance of both**

Falcon's Goal

Substantially increase the rate at which events are simulated

- **Faster than current fast simulators by reasonable factor**
- **At least as accurate as these hand-built simulators**
  - **For example DELPHES**
  - **Eliminate need to implement mapping by hand**
  - **Without sacrificing accuracy**

Falcon

Builder

- **Abstracts detector response from fully-simulated events**
- **Creates non-parametric representation database (one per flavor)**

Simulator

- **Uses this database to simulate events at reconstruction level from parton level**

Falcon

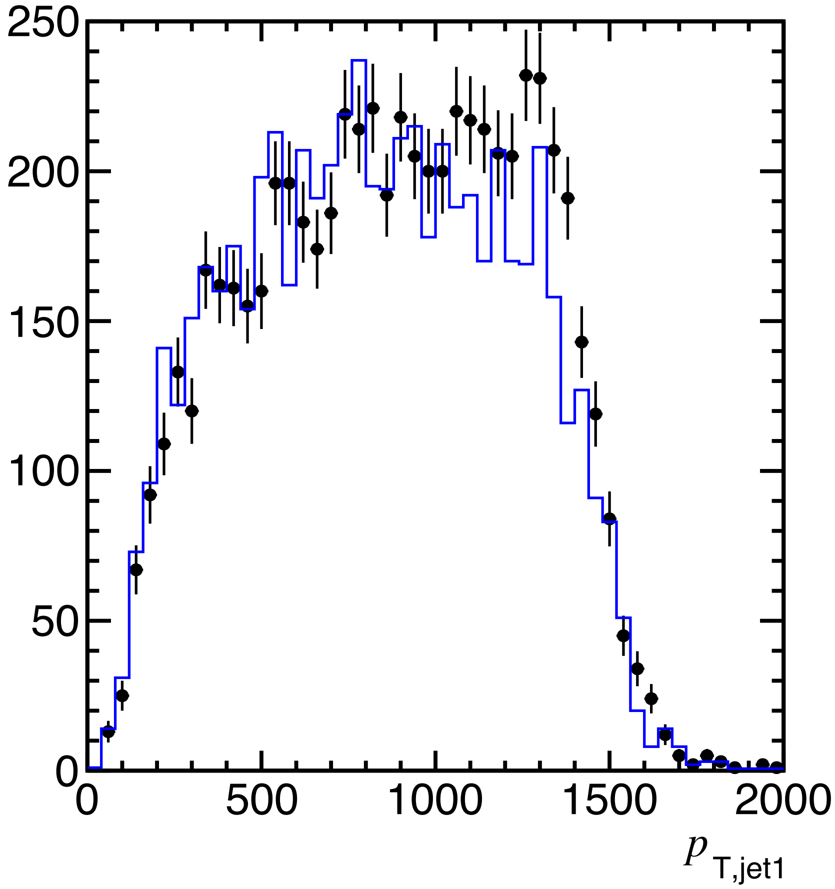
- **Maps partons to reco-level objects automatically taking into account inefficiencies**
- **Extremely fast**
  - **look-up techniques**
- **Tool with as few knobs to tune as possible**
- **The more events in the database, finer granularity the better it performs**

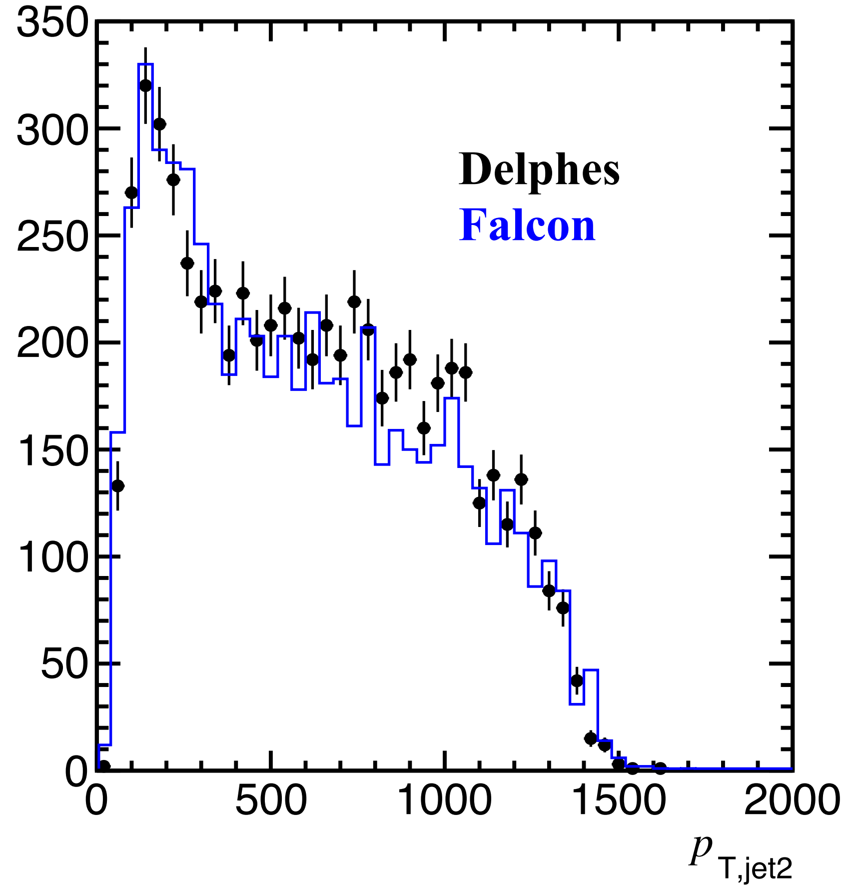
Proof of Principle

- **Leptons are relatively easy to accurately simulate**
  - **Jets are hard(er)**
    - **Start with Jets**
  - **Consider a Model of Heavy Neutral (2.9TeV) Higgs at 13TeV LHC**
    - **Decays to bottom quarks (50%)**
    - **Decays to taus (12%)**

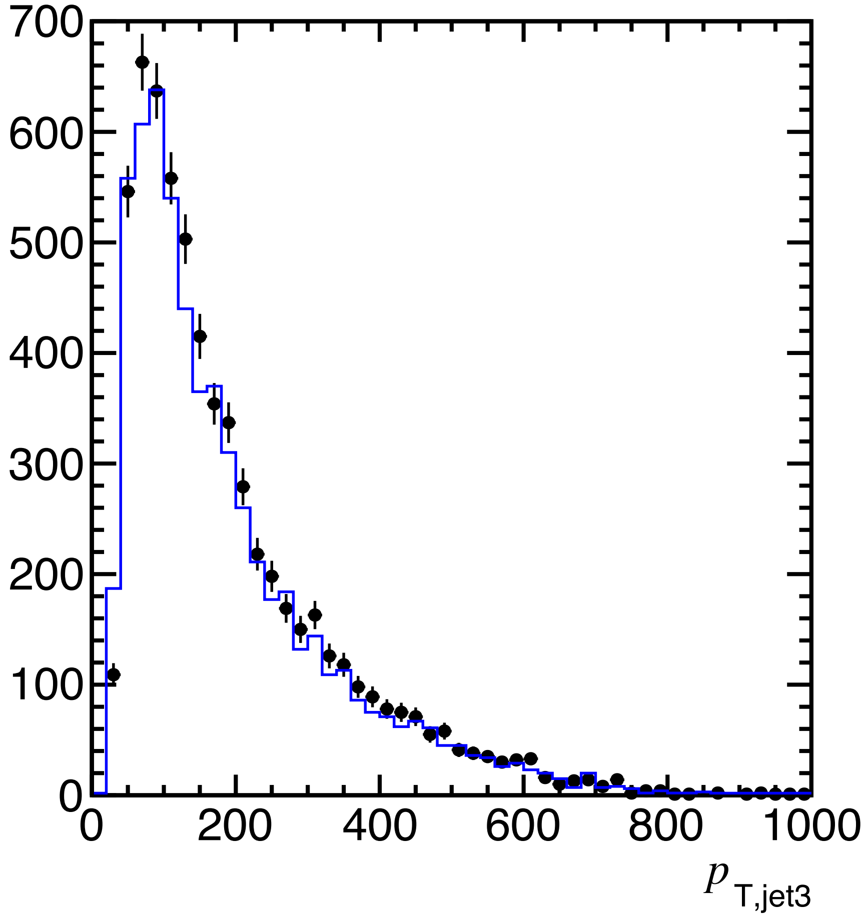
Goal: reproduce P_T spectra of 3 highest P_T jets using Falcon

Results


250
200
150
100
50
0
0
500
1000
1500
2000
$p_{T,jet1}$


350
300
250
200
150
100
50
0
0
500
1000
1500
2000
Delphes
Falcon
$p_{T,jet2}$

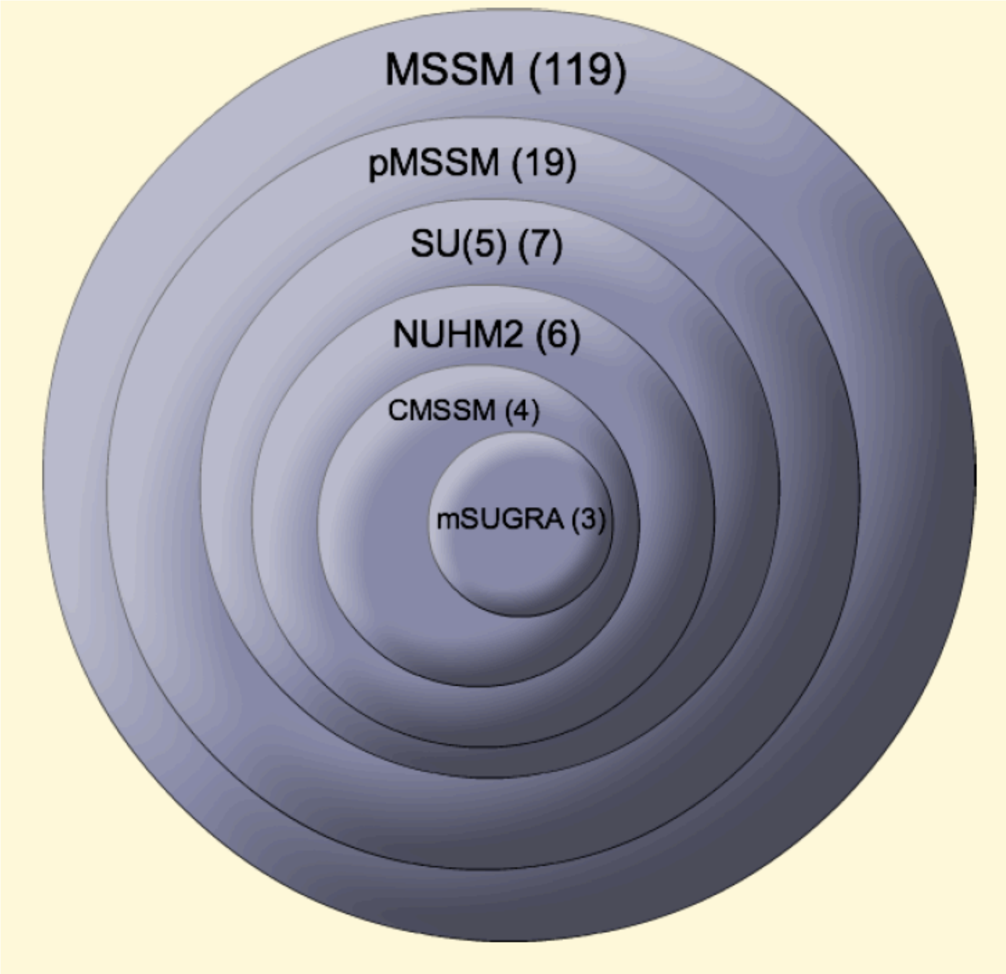
Results



Delphes
Falcon

Very good agreement

Example



pMSSM

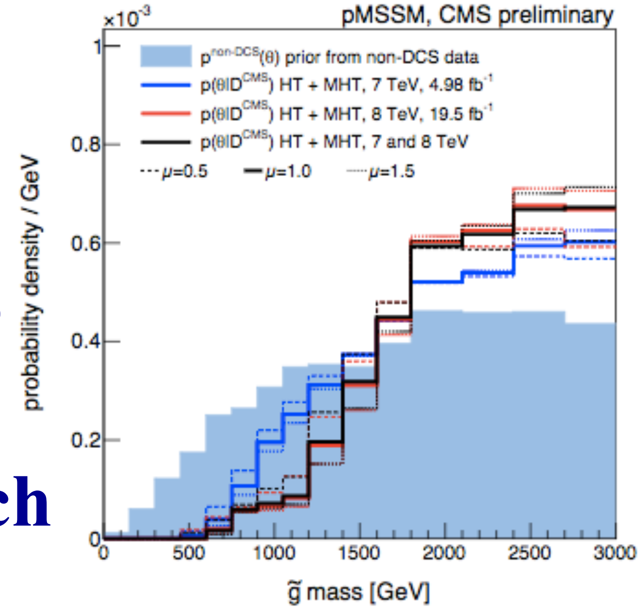
- phenomenological Minimal Supersymmetric Standard Model **pMSSM**
- 19 dimensional space
- No GUT scale assumptions or simplifications
  - as few assumptions as possible at electroweak scale
- Want to take these **high-dimensional** models seriously
  - and make **realistic statements** about them based on data

pMSSM

- [arXiv:1606.03577]
- **CMS Run 1 Analysis**
- **19 dimensional space**
  - **– enough density of points to make a realistic statement**
  - **– 1000 points: 10k events each**

  That's too low.

  O(1-10b) to make a good scan
  - **– 400ev/s → 1b event/month**



Outlook

Accuracy

- Make use of existing simulated events
- Increase density of what is stored
- Finer granularity of binning in parton space - higher fidelity

Speed

- Parallel look-up

Outlook (2)

Mapping

- **Best way to perform the mapping**
  - Naïve
    - ΔR threshold (per species) resulting in a 1 to N map
    - implemented
  - Alternative
    - from a single parton a tree of objects with probability quantifying a degree of match
      - User can choose a cut on this probability (or none)

Summary

- Falcon ultrafast non-parametric detector simulation
- A modern take on an old idea
- Our studies suggest it works well
- As 750GeV bump has shown, low hanging fruit is elusive. Will have to deal with more difficult regions and multi-parameter models
  - Having fast and accurate simulation at hand is critical to the evaluation of multi-parametric models with data from LHC
- More about Falcon: arXiv:1605.02684
- Inputs and suggestions are welcome

Backups

KDTreeBinning

KDTree is a data structure

- **k-dimensional tree**
- **special case of binary partitioning**
- **resulting bins have same number of events**

FALCON uses KDTreeBinning in ROOT to store and partition data

- **Added more functionality to the class**