Qualification Task AFT 455:

Optimization of Inputs for High Level Discriminants (DL1 and MV2) to Improve Performance of B-Tagging in Heavy Ion Collisions

Xiaoning Wang

University of Illinois-Urbana Champaign

Oct 17, 2019

Task Description and Summary of Plan

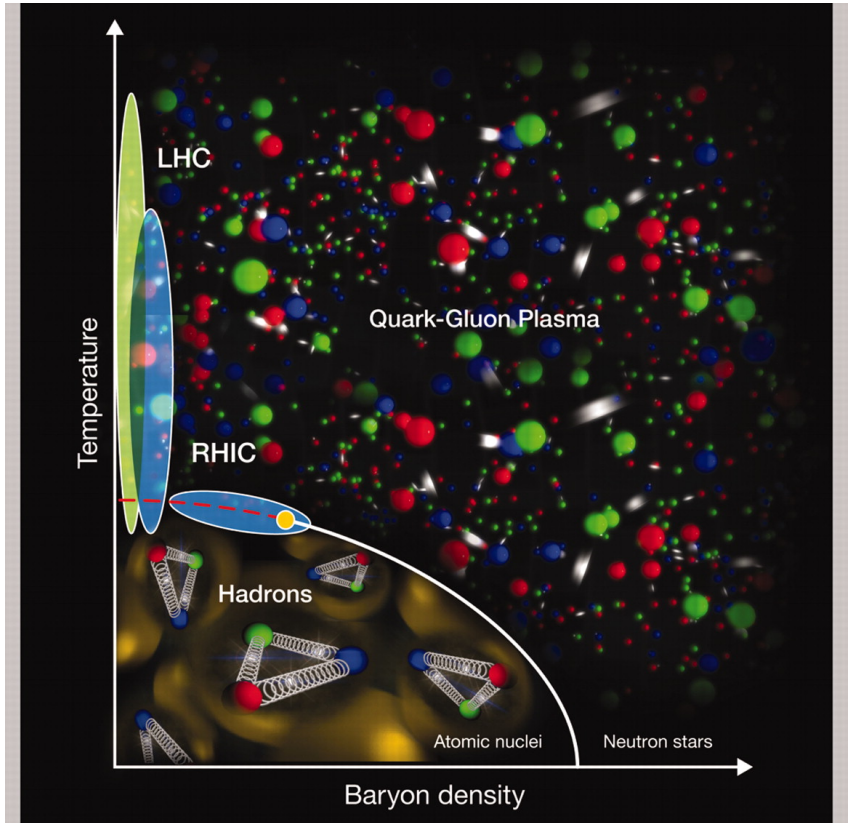
- Goal: optimize the inputs of high-level discriminates (DL1 & MV2) for b-tagging in HI collisions.
- Problem: HI collisions have large number of Underlying Event (UE) tracks that modify some inputs.

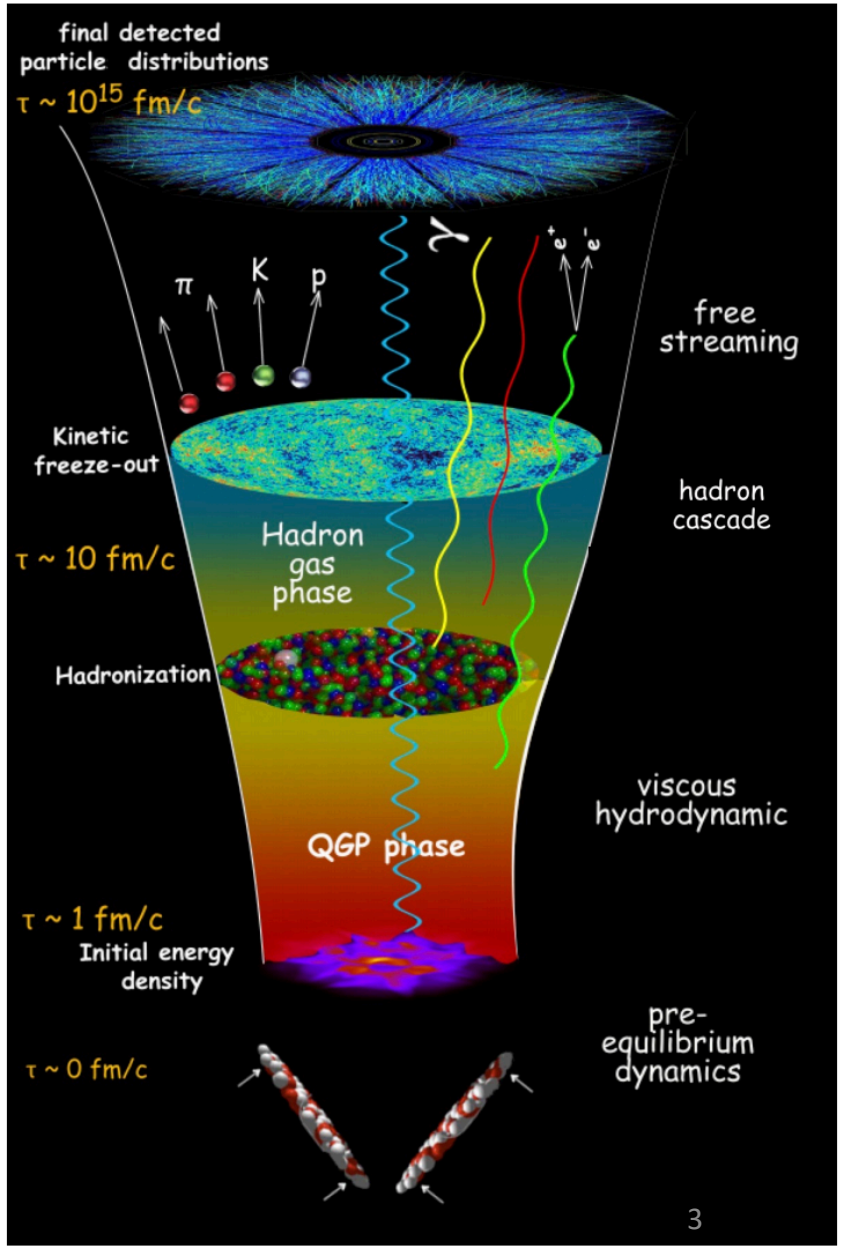
- First step plans: apply selections on tracks in HI MC to see whether performance improves
  - Impose cuts on p_T.
  - Apply “cone method” to subtract UE tracks effect.

“The optimization of the inputs of high-level taggers(DL1 and MV2) for b-tagging in heavy ion collisions, following the work done in a previous QT described in AFT-233. It is known that some inputs for the taggers training are affected (like ipxd probabilities and jet fitter and sv1 energy fraction) by the large number of tracks coming from the HI collision underlying event (UE). This degrades the performance for central collisions and induces a strong centrality dependence. This effect can be reduced by implementing tighter tracking selections or an UE subtraction at the tracking level prior the calculation of the tagger inputs. If time permits, following the optimization, the calibration of the taggers will be done using HI data control samples that have a specific flavor composition e.g. jets with a muon from a heavy flavor semi-leptonic decay. This study will be documented in an internal note and the analysis recommendations will be described on a twiki.”

Background: Heavy Ion (PbPb) Collisions

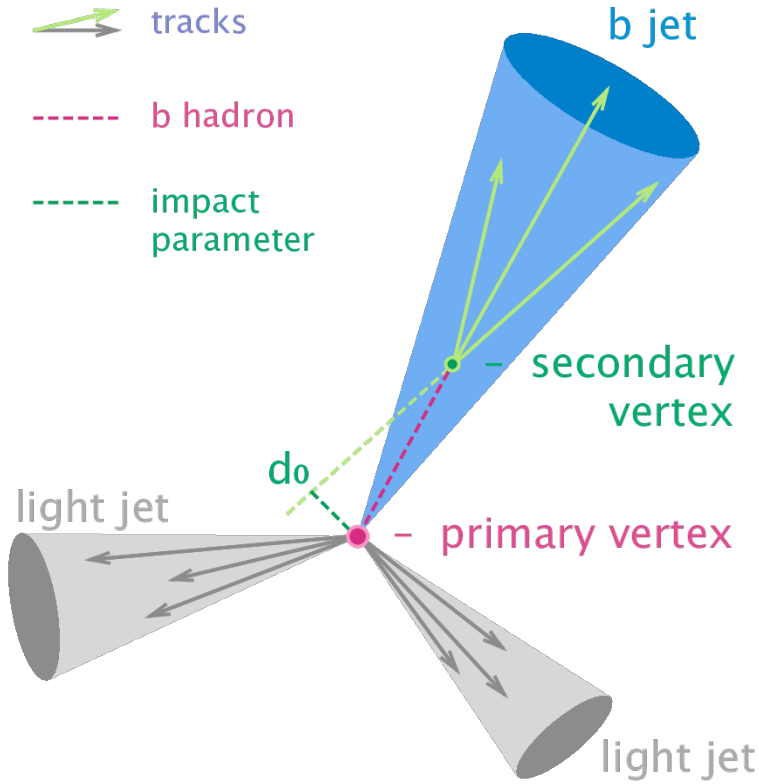
- Quark Gluon Plasma (QGP) are formed in heavy ion collisions.
- Partons are confined in normal matter, free in QGP.
- Jets are used as probes to study QGP.
- For central collision, about 1600 interactions from a single vertex.





Description of Task Background (AFT-455)

- B-tagging: the identification of jets containing b-hadrons
  - Several dedicated algorithms exploiting specific proprieties like long lifetime, high mass and decay multiplicity of b-hadrons and the hard b-quark fragmentation.
  - Low Level Discriminants:
    - IP3: Impact Parameter (d0, z0) based
    - SV1: Secondary Vertex based
    - JetFitter: Secondary Vertex based
  - As input of High Level Discriminants:
    - MV2 and DL1



- Algorithms are trained on *pp* simulations.
- B-tagging algorithms can now run heavy ion data from AFT-233.

Scope of This Study

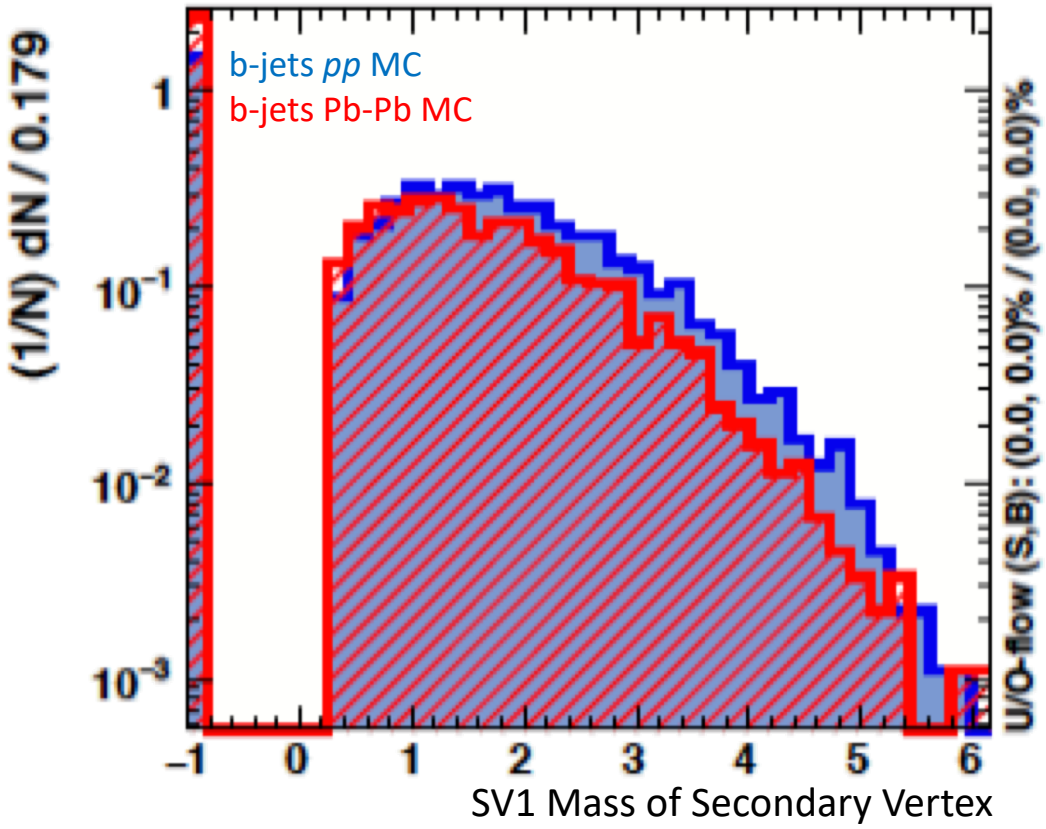
- PbPb 2018 5.02 TeV data with integrated luminosity 1.76 nb^{-1}.
- We estimate to have order of 100 b-jets at p_T 200-300 GeV, and order of 10 b-jets at p_T 300-400 GeV, and the latter will be the upper limit of this optimization.
  - The estimation was made for all centrality using the $L_{int} \times \sigma^{pp}_{b-jets} \times \langle N_{coll} \rangle \times \frac{\sigma^{tot}_{AA}}{\sigma^{pp}_{tot}}$, where $\langle N_{coll} \rangle \times \frac{\sigma^{tot}_{AA}}{\sigma^{pp}_{tot}}$ was estimated as 40,000. (N_{coll} is the expected number of binary nucleon-nucleon collisions)
  - The theoretical calculation for b-jets cross section is from this paper by Hai Tao Li and Ivan Vitev: https://arxiv.org/abs/1811.07905

- There are ongoing measurements within the HI group of b-jet cross-sections in PbPb and *pp* at 5.02 TeV using muon based tagging, which can provide cross checking reference.
  - The link to this internal note is here: https://cds.cern.ch/record/2683608?#

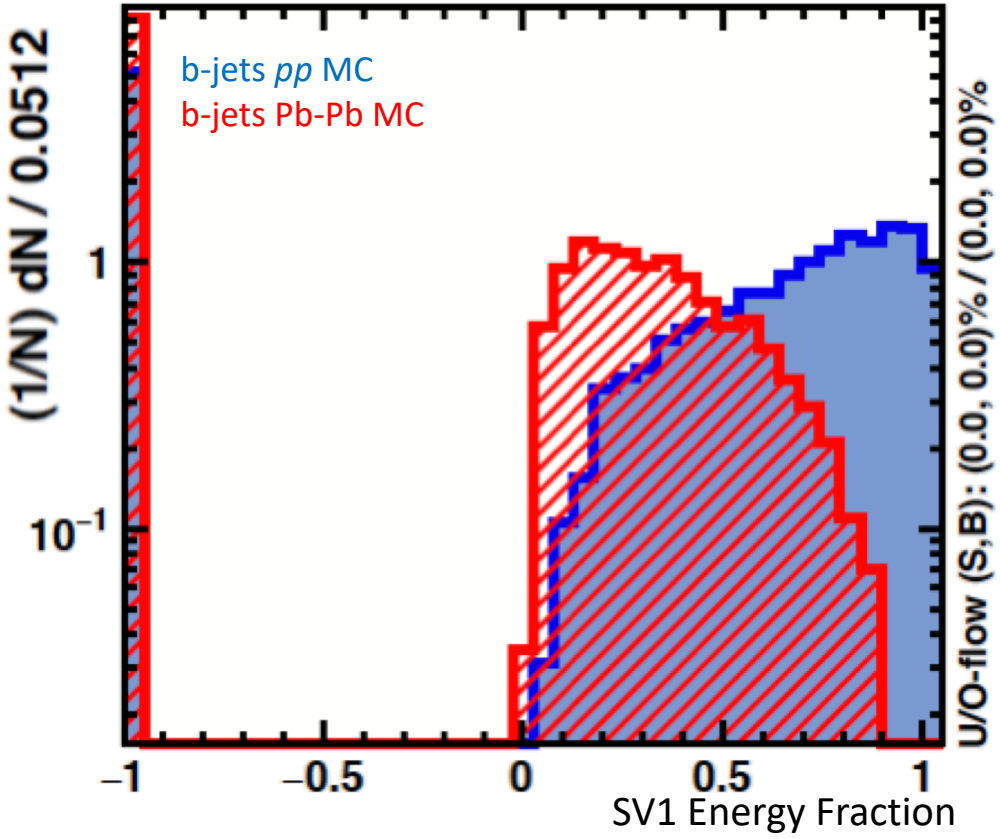
Description of Qualification Task (AFT-455)

- Optimization of inputs for high level discriminants (DL1 and MV2) to improve performance of B-tagging in heavy ion collisions.
- More underlying events (UE) in HI collisions, and some inputs are affected.

Mass of secondary vertex is similar in pp and Pb-Pb



Energy Fraction peaks differently in pp and Pb-Pb



Pb-Pb MC Samples

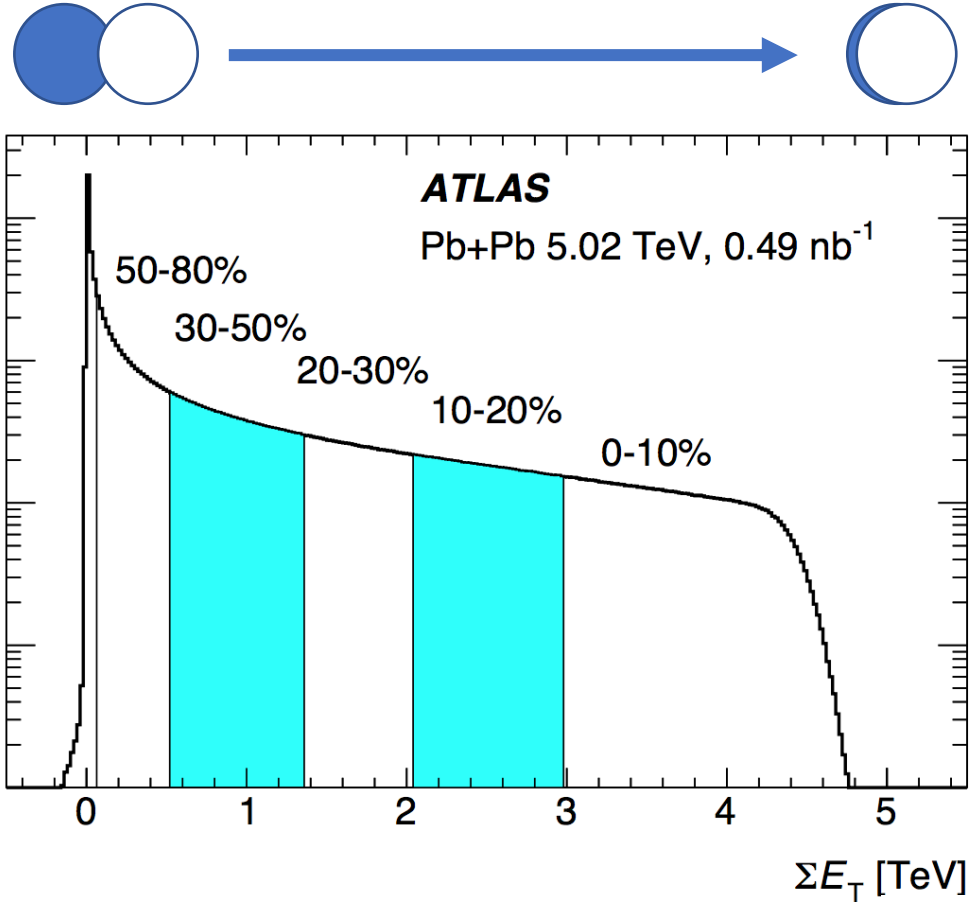
- Pythia MC Overlay as Heavy Ion simulation.
  - Pythia MC events embedded into minimum-bias data from Heavy Ion collisions .

- As of now, we have 50k events of bbar pythia dijets embedded in PbPb 2018 MinBias data as a start.
  - https://its.cern.ch/jira/browse/ATLHI-240
  - Release 21

- We will validate these MC and request more.

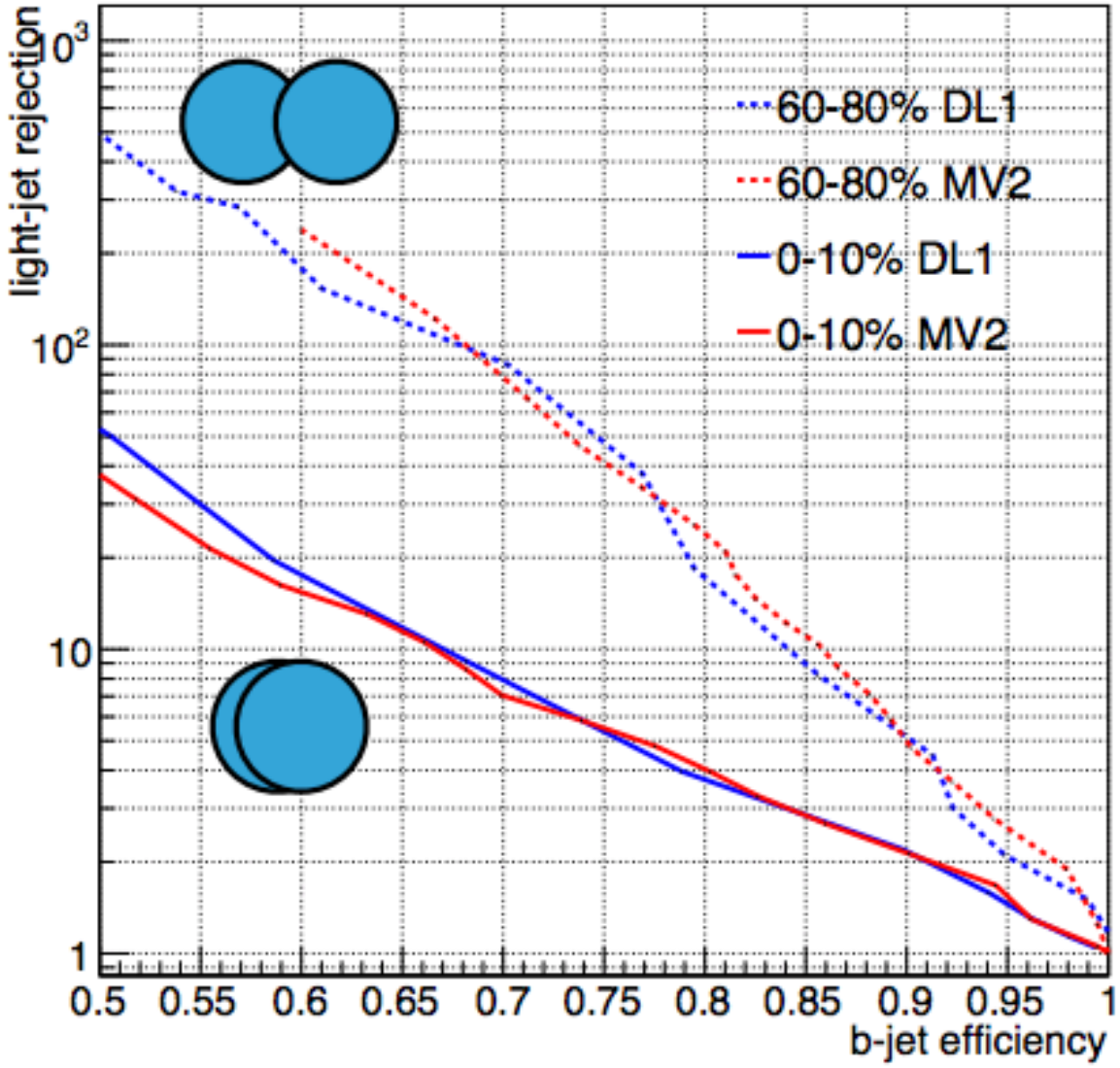
Challenge in Heavy Ion Collisions

- Centrality Dependence
  - Centrality:
    - Whether the collision is central (“head-on”) or peripheral (“glancing”)
    - Estimated using the total transverse energy measured in the ATLAS Forward Calorimeter (ΣE_T)
  - Central collisions have high occupancy (~1600 nucleon-nucleon interactions / collision)



B-Tagging Performance in Overlay MC

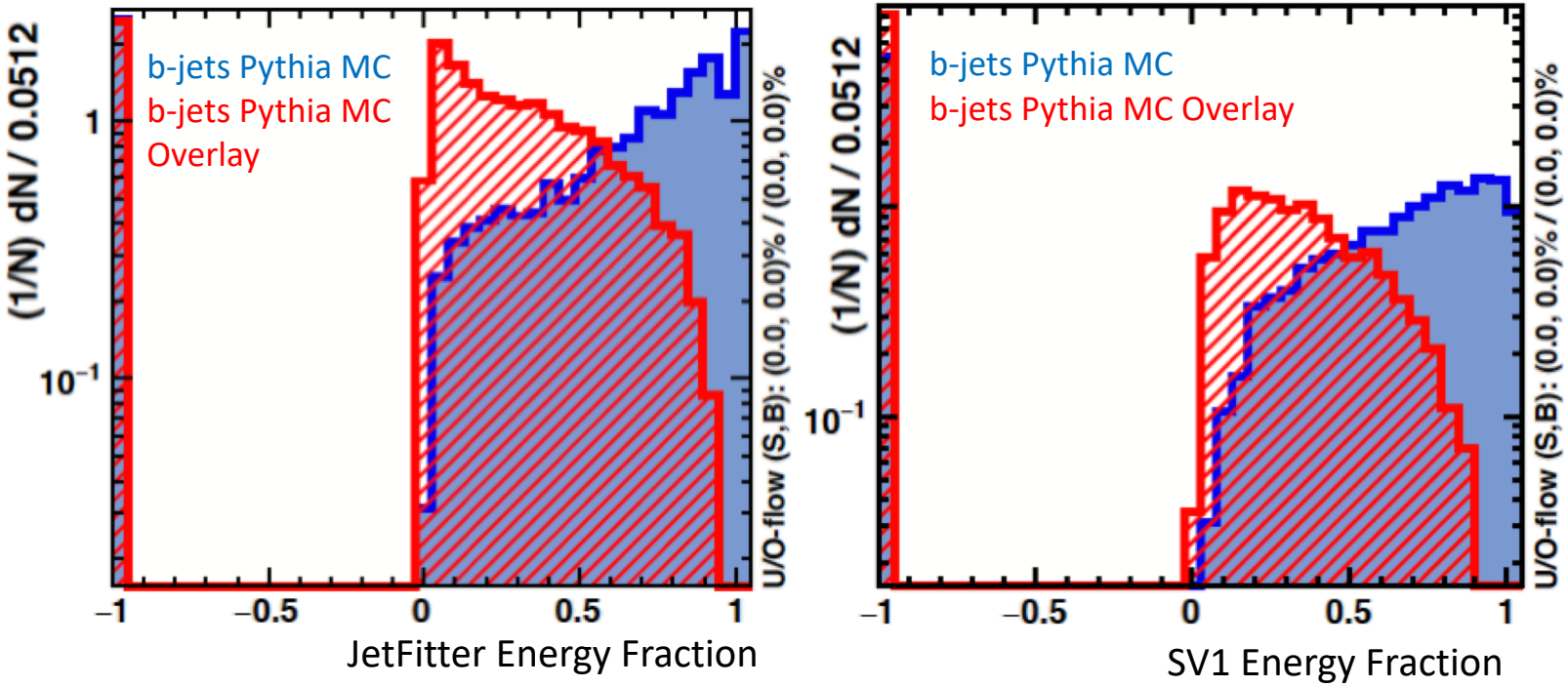
Inclusive Jet-MC

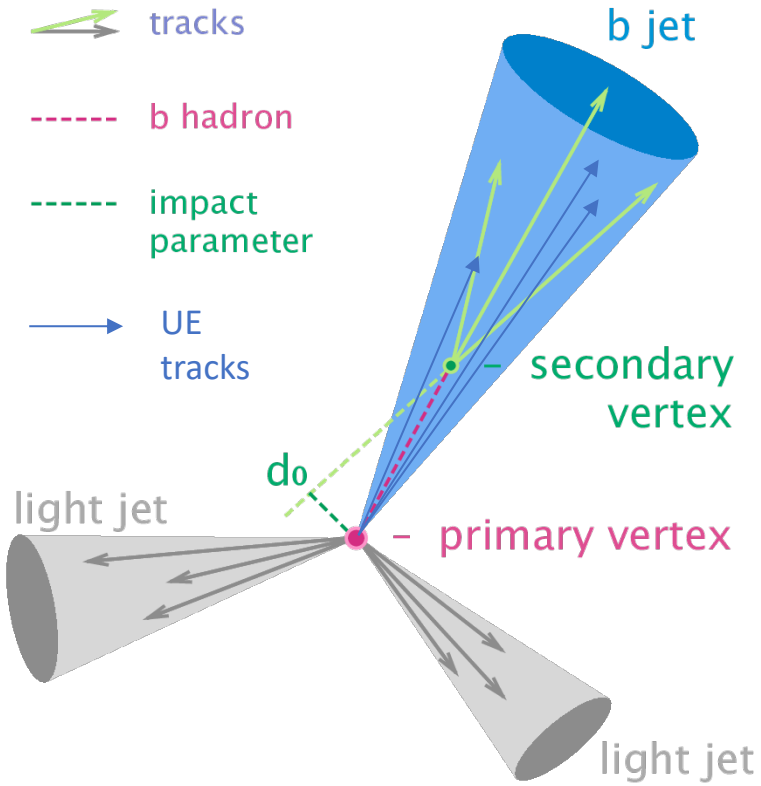


For the same efficiency in b-jet efficiency, central collisions have an order of magnitude lower of light-jet rejection than peripheral collisions.

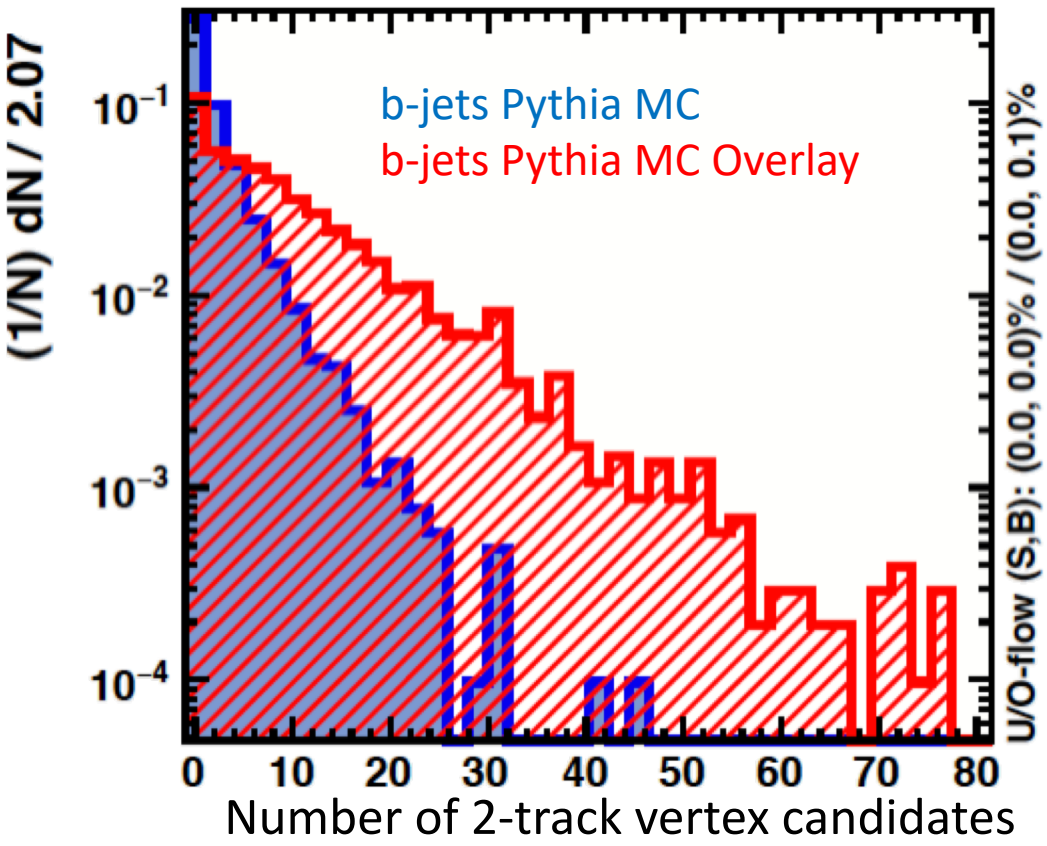
Inputs Significantly Modified by Underlying Events (UE)

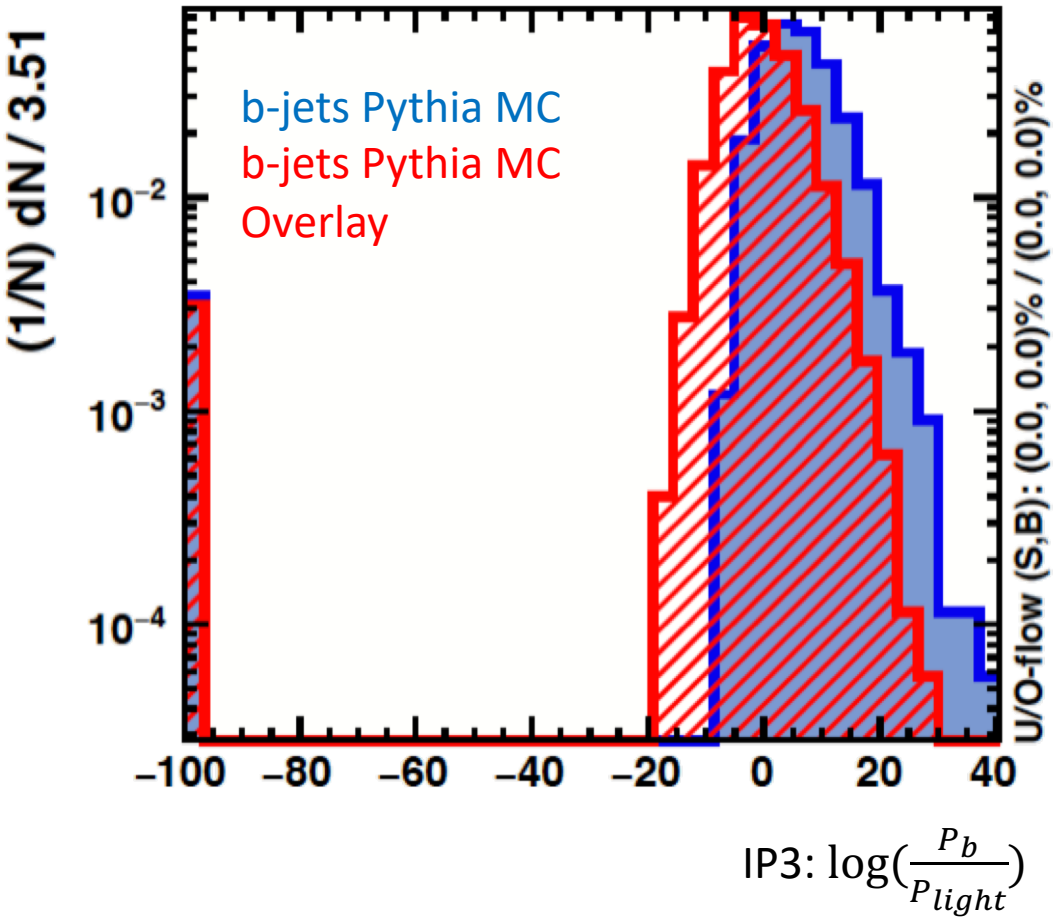
- Tracks from Underlying Events (UE) are mistaken as part of jet tracks.
  - Wrong energy fraction (efc=$E_{SV.trks}/E_{All\ trks}$), UE are included in denominator.





Inputs modified Heavily by Underlying Events (UE) (Continued)





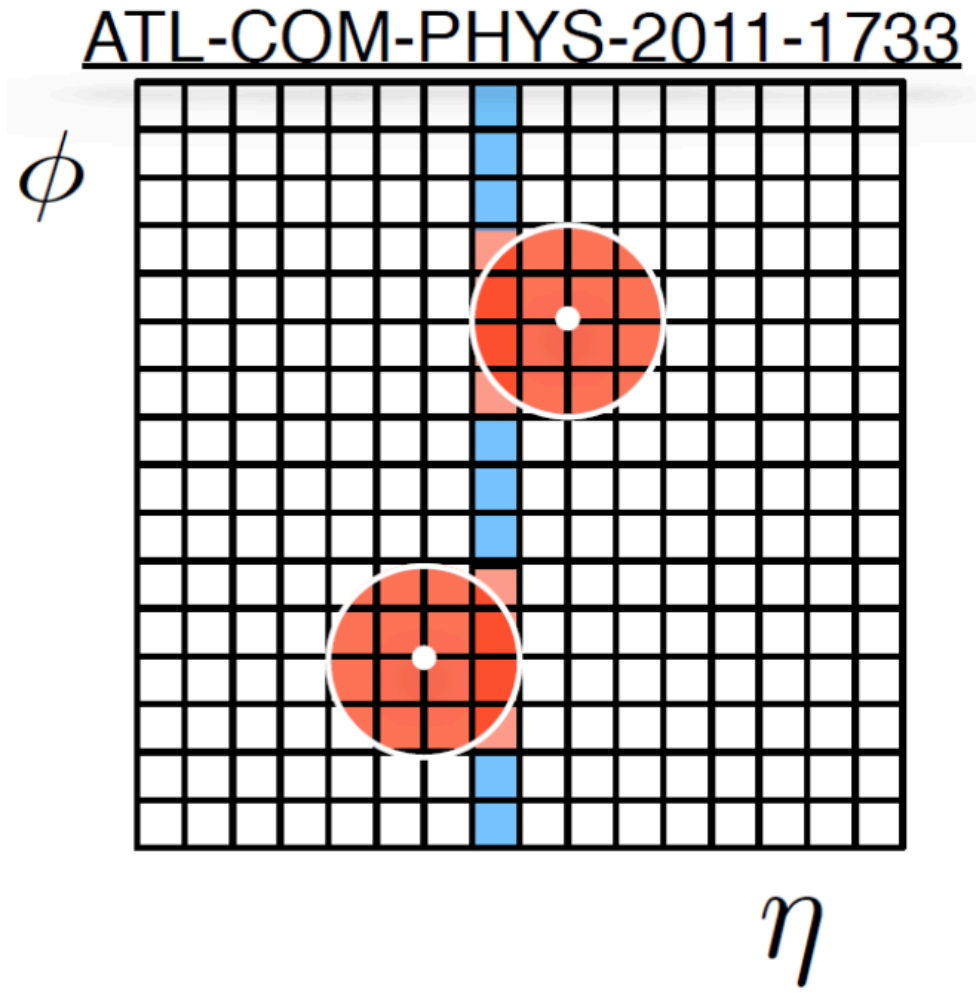
Question: What selections are applied?

$\log(\frac{P_b}{P_{light}})$: Likelihood ratio between the b-jet and light-jet hypotheses

First Step: Apply Selections to Reduce UE Effects

- Develop selections for heavy ion collisions on HI MC (Pythia MC Overlay) to reject UE tracks before calculating inputs.

- Possible methods:
  - Tighter track selections
    - For example, most UE tracks have lower p_T.
  - Cone method
    - Exclude regions of jets and use remaining tracks to model UE tracks distribution to energy. Will help shift Energy Fraction to the right.
    - A method used by current HI groups to model and subtract UE tracks effect from jets.
    - https://arxiv.org/pdf/1805.05424.pdf



Plans and Progresses (Dec 9, 2019)

- Complied the btagging tool out-of-box and tried on one MC sample from https://its.cern.ch/jira/browse/ATLHI-240
  - Program ran, is looking at debugging for calibration tool.
    - Meanwhile understanding the details of b-tagging algorithms.
  - Add FCal information into the algorithm.
  - MC needs validation (waiting for people's work on inner detector track issues)
    - Need to request more MC samples

- Experiments with cutting on small samples locally (without re-training the network)
  - Applying pt cut in tracks used for tagging.
  - Get more ideas from comparing the distributions between pp and PbPb.

Summary

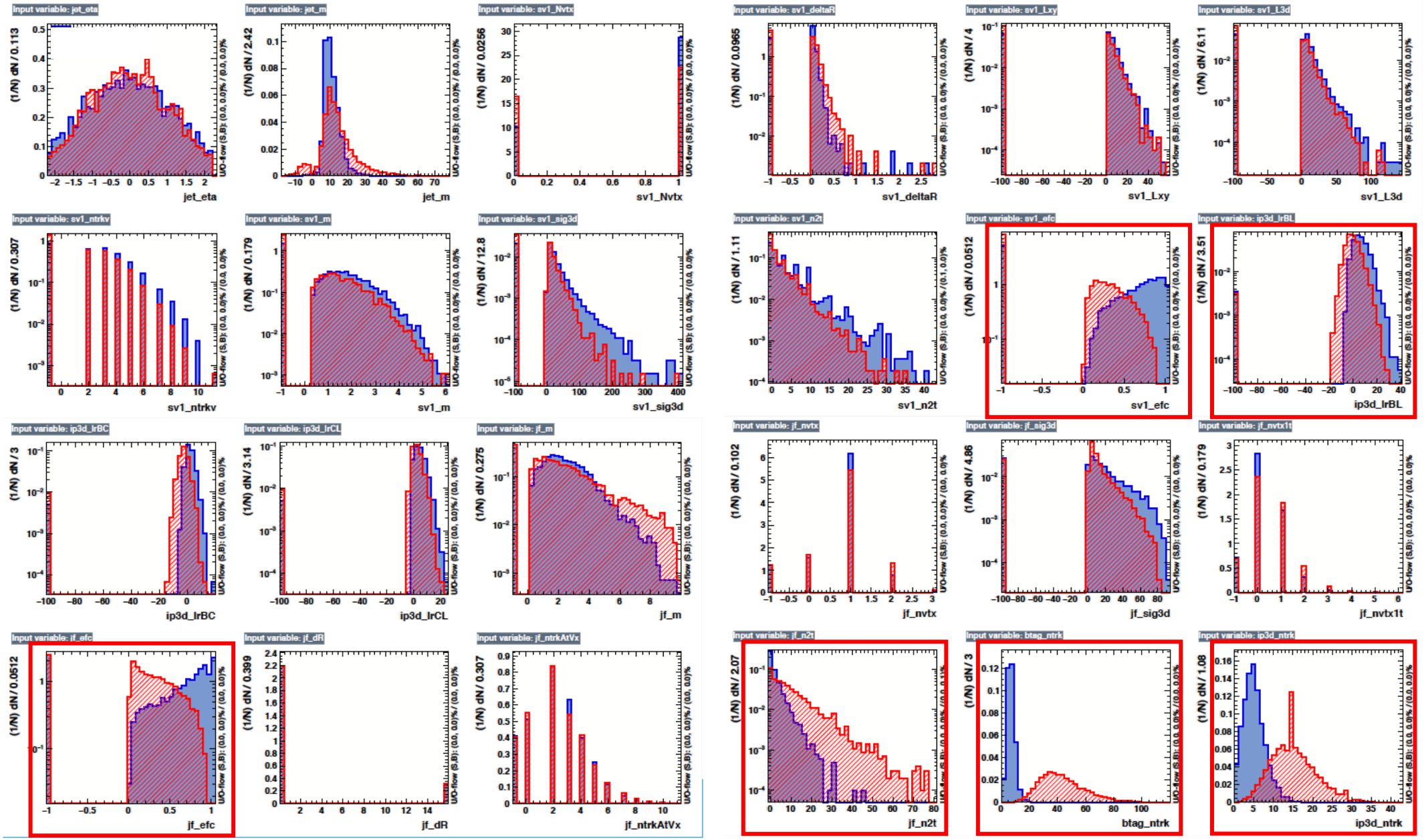
- This task will optimize the inputs for the high level discriminants (MV2 and DL1) in order to improve the B-tagging performance in heavy ion collisions.
  - Performance of high level discriminants on HI collisions have a strong centrality dependence.
  - More underlying events are present in central collision, lowering the performance.
  - It is found that some inputs are significantly modified in heavy ion collisions.

- First steps planned are to develop selections to reduce effects of underlying event (UE) tracks using Pythia dijets overlay.

- The p_T range of this study will overlap with the ongoing measurements for b-jets using muon-based tagging within the HI group, which can provide reference for result comparison.

Back-up Slides

Inputs of b-jets in pp and Pb-Pb simulations



b-jets Pythia MC

b-jets Pythia Overlay

Significantly modified inputs.

Notes for myself (question slide)

- Is there a difference between the 2 algorithms DL1 and MV2? (Should we optimize for each or the optimization should yield same improvements in both)
- Where can we access current algorithms of IP3, SV1, JF. to check what selections have been applied?
- Can the selections/method of calculating inputs be changed in these algorithms?
- Selections on calculating n2t

List of Inputs for Low Level Discriminants

Input	Variable	Description
Kinematics	p_T (jet)	Jet transverse momentum
	$\eta(jet)$	Jet pseudorapidity
IP2D, IP3D	$\log(P_b/P_{light})$	Likelihood ratio between the b- and light-jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{light})$	Likelihood ratio between the c- and light-jet hypotheses
SV	$m(SV)$	Invariant mass of tracks at the SV assuming π masses
	$f_E(SV)$	Fraction of the charged jet energy in the SV
	$N_{TrkAtVtx}(SV)$	Number of tracks used in the SV
	$N_{2TrkVtx}(SV)$	Number of two track vertex candidates
	$L_{xy}(SV)$	Transverse distance between the PV and the SVs
	$L_{xyz}(SV)$	Distance between the PV and the SVs
	$S_{xyz}(SV)$	Distance between the PV and SVs divided by its uncertainty
	$\Delta R(jet, SV)$	ΔR between the jet axis and the direction of the SV relative to the PV

Jet Fitter	$N_{2TrkVtx}(JF)$	Number of 2-track vertex candidates
	$m(JF)$	Invariant mass of tracks from displaced vertices assuming π masses
	$S_{xyz}(JF)$	Significance of the average distance between the PV and displaced vertices
	$f_E(JF)$	Fraction of the charged jet energy in the SVs
	$N_{1-trkvertices}(JF)$	Number of displaced vertices with one track
	$N_{\geq 2-trkvertices}(JF)$	Number of displaced vertices with more than one track
	$N_{TrkAtVtx}(JF)$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

Bayesian Method for Calculating Efficiency Uncertainty

Xiaoning Wang

Dec 9, 2019

Calculating Efficiency Uncertainty

- Problem: use regular propagation of errors cannot account for that the total number of events N and number of passed events k are correlated.
- Two commonly used methods:

- Binomial errors $\frac{\sqrt{k(\frac{1-k}{N})}}{N}$

$$\begin{aligned}\sigma_k &= \sqrt{\mathrm{var}(k)} \\ &= \sqrt{\epsilon(1-\epsilon)N}.\end{aligned}$$

$$\delta\epsilon' = (1/N)\sqrt{k(1-k/N)}$$

$\varepsilon' = k/N$: estimate of true efficiency
$\delta\varepsilon' = \sigma_k/N$: uncertainty of estimated efficiency

Problem: gives symmetric uncertainty, but k=N can have nonzero lower uncertainty, and k=0 can have nonzero upper uncertainty.
Source:
https://home.fnal.gov/~paterno/images/effic.pdf

- Bayesian errors

$$P(\epsilon|k;N) = \frac{1}{norm} \times P(k|\epsilon;N) \times Prior(\epsilon)$$

$$P(k|\epsilon;N) = Binomial(N,k) \times \epsilon^k \times (1-\epsilon)^{N-k} \ldots \text{ binomial distribution}$$

$$Prior(\epsilon) = \frac{1}{B(\alpha,\beta)} \times \epsilon^{\alpha-1} \times (1-\epsilon)^{\beta-1} \equiv Beta(\epsilon;\alpha,\beta)$$

$$\Rightarrow P(\epsilon|k;N) = \frac{1}{norm'} \times \epsilon^{k+\alpha-1} \times (1-\epsilon)^{N-k+\beta-1} \equiv Beta(\epsilon;k+\alpha,N-k+\beta)$$

https://root.cern.ch/doc/master/classTEfficiency.html
Step by step explanation see next slide

Bayesian Method

$$P(\epsilon|k;N) = \frac{1}{norm} \times P(k|\epsilon;N) \times Prior(\epsilon)$$

Bayes' Theorem

$$P(k|\epsilon;N) = Binomial(N,k) \times \epsilon^k \times (1-\epsilon)^{N-k} \ldots \ binomial\ distribution$$

$$Prior(\epsilon) = \frac{1}{B(\alpha,\beta)} \times \epsilon^{\alpha-1} \times (1-\epsilon)^{\beta-1} \equiv Beta(\epsilon;\alpha,\beta)$$

In our case where we have no prior knowledge of true efficiency, α and β are set to 1, that is uniform distribution between 0 and 1. And we set it 0 probability outside of this range for normalization.

$$\Rightarrow P(\epsilon|k;N) = \frac{1}{norm'} \times \epsilon^{k+\alpha-1} \times (1-\epsilon)^{N-k+\beta-1} \equiv Beta(\epsilon;k+\alpha,N-k+\beta)$$

- ROOT code used:

```
TGraphAsymmErrors *new_eff2 = new TGraphAsymmErrors(hMatched, hTotal,"cl=0.683 b(1,1) mode");
```

Confidence level: 0.683

b(1,1): prior probability distribution of ϵ $Prior(\epsilon)$ is set as uniform from 0 to 1 $Beta(\epsilon; 1,1)$.

mode: Use mode of the posterior probability distribution of ϵ $P(\epsilon|k;N)$ for estimate of centroid ϵ value.

Mode of a beta distribution $Beta(\varepsilon;x,y)$ for x, y >1 is given by the formula $\frac{x-1}{x+y-2}$ (https://en.wikipedia.org/wiki/Beta_distribution#Definitions). And in this case it is $\hat{\varepsilon} = \frac{k+\alpha-1}{N+\alpha+\beta-2}$

- With the probability distribution given by last step, ROOT numerically finds the shortest interval boundaries that integrates to the confidence level and subtract the mode value as upper and lower efficiency uncertainty.

- Source code implemented in TGraphAsymmErrors and TEfficiency
- https://root.cern.ch/doc/master/TEfficiency_8cxx_source.html#l01249

How uncertainty in hMatched and hTotal is used (Part 1)

```
tw = total->GetBinContent(b);
tw2 = (total->GetSumw2()->fN > 0) ? total->GetSumw2()->At(b) : tw;
pw = pass->GetBinContent(b);
pw2 = (pass->GetSumw2()->fN > 0) ? pass->GetSumw2()->At(b) : pw;
```

tw: number of total events.

pw: number of passed/matched events.

tw2 and pw2 as squares of weight for total and passed/matched.

Note:

In case of unweighted events where sumW2 are set before filling, tw2 = tw = bin content.

In case the error bar of binning is set manually instead of by filling, ***tw2 = error^2***

The latter case was used in case of extracting passed and total using fitting to data, where error is the error of fitting.

How uncertainty in hMatched and hTotal is used (Part 2)

```
if ((bEffective && !bPoissonRatio) && tw2 <= 0) {
   // case of bins with zero errors
   eff = pw/tw;
   low = eff; upper = eff;
}
```

In case of tw2 being 0, that is, error bars are manually set to be 0, efficiency also has no error bar.
bEffective is the true and bPoissonRatio is false when we are trying to calculate efficiency instead of ratio of two Poisson means (option "pois").

```
if (bEffective && !bPoissonRatio) {
   // tw/tw2 re-normalize the weights
   double norm = tw/tw2;  // case of tw2 = 0 is treated above
   aa =  pw * norm + alpha;
   bb =  (tw - pw) * norm + beta;
}
if (usePosteriorMode)
   eff = TEfficiency::BetaMode(aa,bb);

if (useShortestInterval) {
   TEfficiency::BetaShortestInterval(conf,aa,bb,low,upper);
}
```

alpha and beta is given by "b(alpha, beta)" in the option

usePosteriorMode is set by "mode" in the option
userShortestInterval is by default set to true when using "mode" option, see slide 9 for the author (Marc Paterno)'s reason for recommending using shortest confidence interval method.

$$P(\epsilon|k;N) = \frac{1}{norm'} \times \epsilon^{k+\alpha-1} \times (1-\epsilon)^{N-k+\beta-1} \equiv Beta(\epsilon; k+\alpha, N-k+\beta)$$

Recall this is our formula for posterior probability distribution

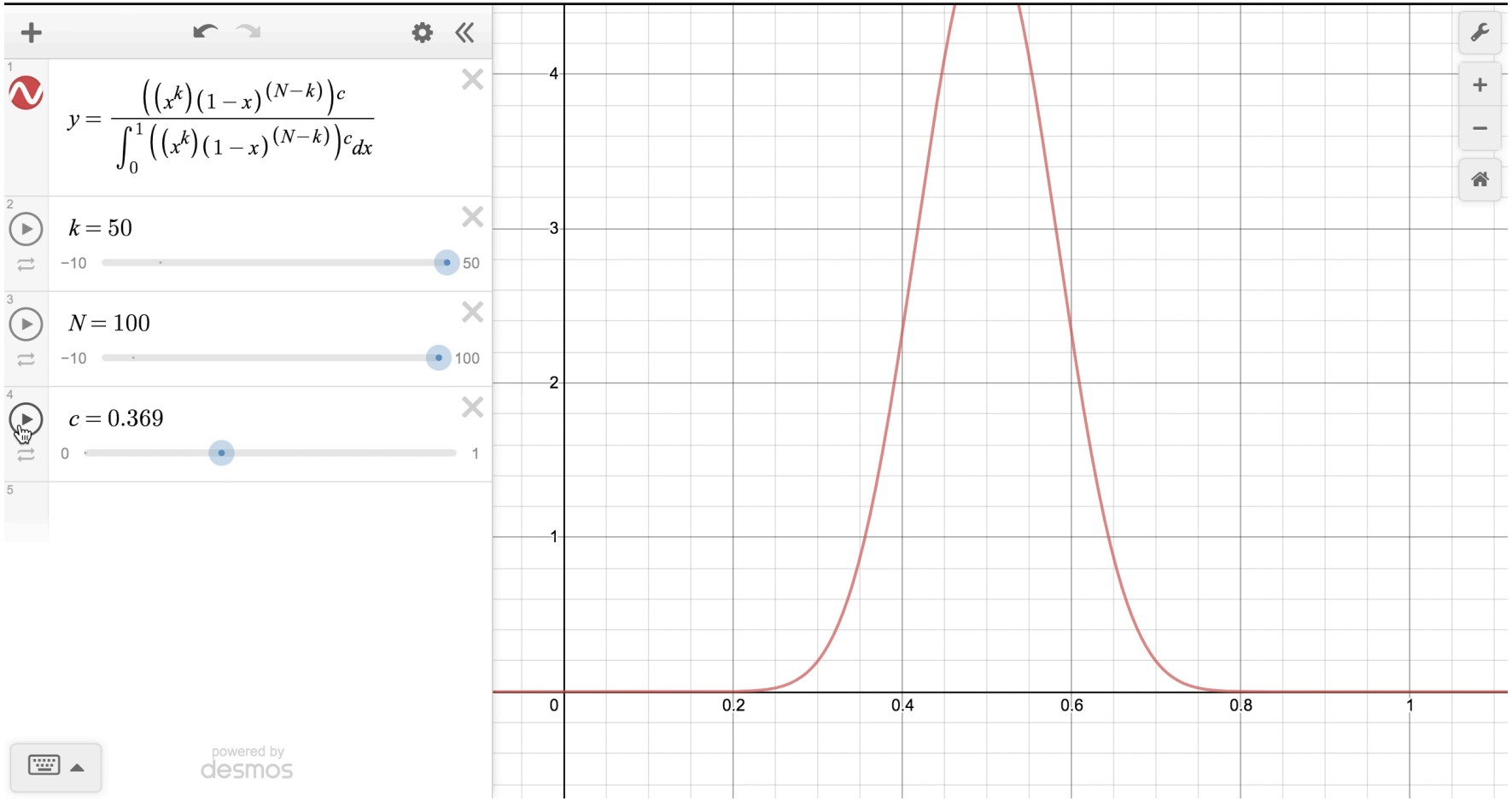
In case of tw2 being set manually, a factor (tw/tw2), $\frac{N}{\sigma_N^2}$, is applied to k and $N-k$.

$k \to k \times \frac{N}{\sigma_N^2}$, $(N-k) \to (N-k) \times \frac{N}{\sigma_N^2}$

Centroid value is not affected by this renormalization for $\alpha = \beta = 1$, for its calculated using $\hat{\varepsilon} = \frac{k+\alpha-1}{N+\alpha+\beta-2}$

How uncertainty in hMatched and hTotal is used (Part 3)

- The renormalized posterior distribution $P' \propto P^{\frac{N}{\sigma_N^2}}$, with larger σ_N^2, the distribution becomes flatter and efficiency uncertainty increases.



C = $\frac{N}{\sigma_N^2}$, as c becomes smaller (fitting uncertainty of hTotal increases), the distribution is flatter.

As seen in the formula, this error is not affected by fitting errors in hMatched, and the following simple code verified the case.

```
void matched_noerror(){
    TFile* f = TFile::Open("output/files/eff_ID_ECpt_datapp_HI_dr0.050_exp.root");
    //hMacthed, histogram of number of ID tracks with a matched MS track, uncertainty is set to be fitting uncertainty
    //hTotal, histogram of number of ID tracks in total, uncertainty is set to be fitting uncertainty
    //new_eff2, efficiency graph calculated vs pT, using,
    //TGraphAsymmErrors *new_eff2 = new TGraphAsymmErrors(hMatched, hTotal,"cl=0.683 b(1,1) mode");
    TGraphAsymmErrors* new_eff2 = (TGraphAsymmErrors*)f->Get("new_eff2");
    TH1F* hMatched = (TH1F*)f->Get("hMatched");
    TH1F* hTotal = (TH1F*)f->Get("hTotal");
    //define a new hMatched, and set its error bar to 0
    TH1F* hMatched_n = (TH1F*)hMatched->Clone();
    for (int i = 0; i < 6; i++){
        hMatched_n->SetBinError(i+1,0);
    }
    TGraphAsymmErrors* new_eff3 = new TGraphAsymmErrors(hMatched_n, hTotal,"cl=0.683 b(1,1) mode");
    cout << "With error bars on hMatched, the eff errors are:" << endl;
    cout << "Lower errors: " << new_eff2->GetErrorYlow(0) << ", " << new_eff2->GetErrorYlow(1) << ", " << new_eff2->GetErrorYlow(2) << "
        , " <<new_eff2->GetErrorYlow(3) << ", " <<new_eff2->GetErrorYlow(4) << ", " <<new_eff2->GetErrorYlow(5) <<  endl;
    cout << "Upper errors: " << new_eff2->GetErrorYhigh(0) << ", " << new_eff2->GetErrorYhigh(1) << ", " << new_eff2->GetErrorYhigh(2) <<
         ", " <<new_eff2->GetErrorYhigh(3) << ", " <<new_eff2->GetErrorYhigh(4) << ", " <<new_eff2->GetErrorYhigh(5) << endl;

    cout << "Without error bars on hMatched, the eff errors are:" << endl;
    cout << "Lower errors: " << new_eff3->GetErrorYlow(0) << ", " << new_eff3->GetErrorYlow(1) << ", " << new_eff3->GetErrorYlow(2) << "
        , " <<new_eff3->GetErrorYlow(3) << ", " <<new_eff3->GetErrorYlow(4) << ", " <<new_eff3->GetErrorYlow(5) << endl;
    cout << "Upper errors: " << new_eff3->GetErrorYhigh(0) << ", " << new_eff3->GetErrorYhigh(1) << ", " << new_eff3->GetErrorYhigh(2) <<
         ", " <<new_eff3->GetErrorYhigh(3) << ", " <<new_eff3->GetErrorYhigh(4) << ", " <<new_eff3->GetErrorYhigh(5) << endl;

}
```

```
With error bars on hMatched, the eff errors are:
Lower errors: 0.000859964, 0.00106934, 0.000827385, 0.000374413, 0.000840199, 0.000661177
Upper errors: 1.14341e-06, 0.000766622, 0.000653977, 2.91142e-08, 0.000714811, 0.000555899
Without error bars on hMatched, the eff errors are:
Lower errors: 0.000859964, 0.00106934, 0.000827385, 0.000374413, 0.000840199, 0.000661177
Upper errors: 1.14341e-06, 0.000766622, 0.000653977, 2.91142e-08, 0.000714811, 0.000555899
```

Summary

- The option "cl=0.683 b(1,1) mode”, which is a replacement for the old method BayesDivide in dividing two histograms for efficiency uses Bayesian method, the code was written by Marc Paterno, and the description of method is available here, https://home.fnal.gov/~paterno/images/effic.pdf .
- This method utilizes only uncertainty in fitting total events but not that of the matched events.
  - The fitting is a simultaneous fitting, and number of total events and its uncertainty is obtained by simultaneously fitting into matched and unmatched data.

Back up (reasons for using shortest confidence interval by the author)

I recommend using the *shortest 68.3% confidence interval* as the measure of the uncertainty in the efficiency measurement. It has two attractive features. First, it has a known probability content, one chosen to be the same as a "1 σ" Gaussian error. Therefore such error intervals will behave as we most often expect. Second, it is the most constrained region which has this probability content, so that we present our measurement in the fashion that most constrains the range in which we believe the true value exists.

References

- https://home.fnal.gov/~paterno/images/effic.pdf Marc Paterno
- https://root.cern/root/html524/TGraphAsymmErrors.html#TGraphAsymmErrors:BayesDivide
- https://root.cern.ch/doc/master/classTEfficiency.html, Section IV: Statistic Options