QCD VBFMET samples MadGraph: New proposal

João Pela

Imperial College London

2015-09-01

Introduction

In the last presentation

- Presented first results from MadGraph `pp>jj`
  - Pythia8 hadronization efficiency
  - Variable migration studies: parton → gen jets
- Questions were raised about Pythia8 vs MadGraph for parton level cuts

Pythia vs MadGraph

From conversations with Josh Bendavid:

- Yes, both madgraph and pythia also allow the implementation of custom cuts in the hard process.
- MadGraph is more flexible for cuts in general out of the box.
- Madgraph also does allow to include extra jets at the ME level (QCD samples produced for SUSY have 2,3,4 jets at ME level).
  - Better since these can also be included in the phase space cuts (ie you can require that any pair of jets satisfy your requirements).
- We could manage to generate 1 single inclusive sample, without having to break things down in pthat bins.

So it was decided to proceed with studies with pp to 2, 3 and 4 jets sample production

Implementation of new MadGraph cuts

After first multi-jet results came out and after checking MadGraph forums and direct code inspections it became clear mmjj cut was applied to all possible jet combinations.

Implementing MadGraph Cuts

- Had to learn Fortran77 and reverse engineer MadGraph cuts code.
- Implemented cuts to select events with at least one dijet with:
  - dijet_pt: min pT of at least one jet pair
  - dijet_eta: min delta era of at least one jet pair
  - dijet_mjj: min invariant mass of at least one jet pair
- According to Chayanit A. this can be integrated in the official MadGraph gridpacks for official production.

Problems found at IC

For a few days I was getting MadGraph crashes in specially long jobs (+10k diagrams)

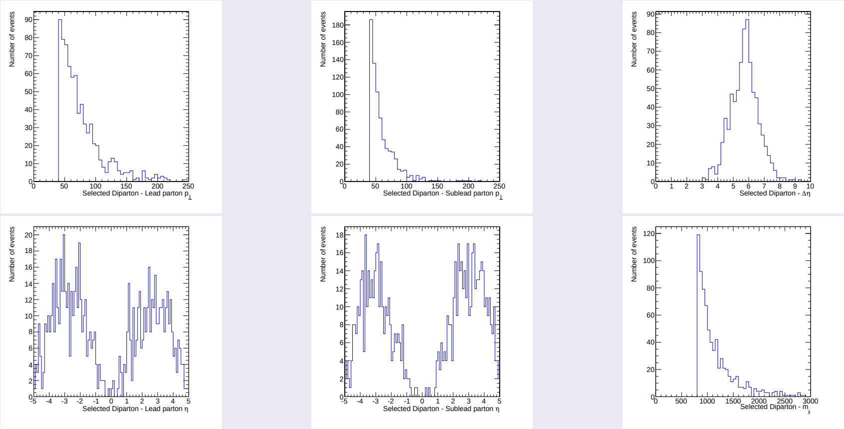
- Turns out MadGraph stores sub-process files at /tmp/ while running.
- At IC in most machines /tmp/ has its own partition and is limited in space to 1 GB.
- Big jobs would run out of space and crash.
- A work around was found with Simon to redirect temporary storage in the jobs.

Imperial College London

Checking Correct Implementation (Parton level)

This plots are for pp to 2, 3 and 4 jets.

Parton distributions - Cuts $p_\perp > 40$, $\eta < 5.0$, $\Delta\eta > 3.0$ and $m_{jj} > 800$



- Cuts look like they are correctly implemented
- It can be seen here and also in my studies $\Delta\eta$ cut provides almost no reduction power below 4. We should not use this cut at parton level.

Imperial College London

Obtained cross sections

$p_\perp > 30,\ m_{jj} > 800$

Process	Cross section [pb]
pp>jj	$1.736 \times 10^6 \pm 4.477 \times 10^3$
pp>jjj	$3.977 \times 10^6 \pm 1.173 \times 10^4$
pp>jjjj	$5.516 \times 10^6 \pm 1.248 \times 10^4$
pp>all	$1.11 \times 10^7 \pm 1.799 \times 10^4$

$p_\perp > 30,\ \Delta\eta > 2.5,\ m_{jj} > 800$

Process	Cross section [pb]
pp>jj	$1.797 \times 10^6 \pm 4.594 \times 10^3$
pp>jjj	$3.804 \times 10^6 \pm 9.742 \times 10^3$
pp>jjjj	$5.568 \times 10^6 \pm 1.083 \times 10^4$
pp>all	$1.078 \times 10^7 \pm 1.946 \times 10^4$

$p_\perp > 40,\ \Delta\eta > 3.0,\ m_{jj} > 800$

Process	Cross section [pb]
pp>jj	$8.621 \times 10^5 \pm 2.605 \times 10^3$
pp>jjj	$2.081 \times 10^6 \pm 4.714 \times 10^3$
pp>jjjj	$3.186 \times 10^6 \pm 5.790 \times 10^3$
pp>all	$5.7616 \times 10^6 \pm 1.2 \times 10^4$

Conclusions

- $\Delta\eta$ cut only provides an event reduction of $\sim 5\%$
- With the least restrictive working point we can produce an 10 fb^{-1} equivalent sample with $\sim 10^{11}$ parton level events.
- Compared with the Pythia8 study that implied full generation of 2×10^{12} events we have a factor 20 reduction.
- Hardest presented set cuts buys a further factor of 2 at the cost of less physics usability, but still under offline signal region cuts.
- Only caveat here is the size of such sample even at only parton level would be 20-30 TB (I think this is acceptable)
- I will proceed presenting results for the least restrictive working point, but I have all number/plots for all working points

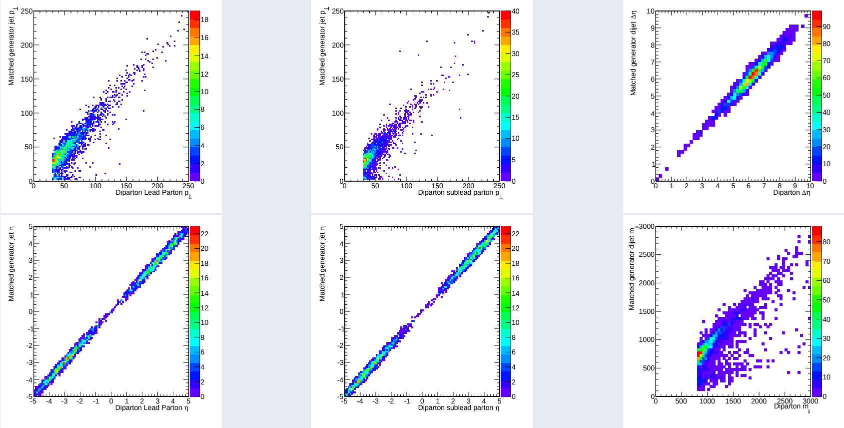
Pythia8 Hadronization

Next step is to proceed with Pythia8 Hadronization and matching.

		Events			Cross section [pb]	
Process	#	Tried	Passed	Accepted [%]	Pre-match	Post-match
pp>jj		10000	2197	22.0 ± 0.4	$1.736 \times 10^6 \pm 4.478 \times 10^3$	$3.815 \times 10^5 \pm 7.256 \times 10^3$
pp>jjj		10000	696	7.0 ± 0.3	$3.977 \times 10^6 \pm 1.173 \times 10^4$	$2.768 \times 10^5 \pm 1.015 \times 10^4$
pp>jjjj		10000	581	5.8 ± 0.2	$5.517 \times 10^6 \pm 1.248 \times 10^4$	$3.205 \times 10^5 \pm 1.293 \times 10^4$
pp>all	jj	1454	312	21.5 ± 1.1	$1.617 \times 10^6 \pm 2.620 \times 10^3$	$3.469 \times 10^5 \pm 1.741 \times 10^4$
pp>all	jjj	3603	251	7.0 ± 0.4	$3.946 \times 10^6 \pm 6.394 \times 10^3$	$2.749 \times 10^5 \pm 1.674 \times 10^4$
pp>all	jjjj	4943	337	6.8 ± 0.4	$5.538 \times 10^6 \pm 8.973 \times 10^3$	$3.775 \times 10^5 \pm 1.986 \times 10^4$
pp>all	all	10000	900	9.0 ± 0.3	$1.110 \times 10^7 \pm 1.132 \times 10^4$	$9.993 \times 10^5 \pm 3.127 \times 10^4$

Matching Dipartons to Generator Dijets

Parton distributions



- At the parton $p_\perp$ cut line the upper tail of finishes around 50 GeV
- At the diparton m_{jj} cut line the upper tail of finishes around 1000 GeV
- Good correlation in η variables, but no migration problems since there is no cut at parton level

Filter efficiency

Testing several generator filter configurations.

Efficiencies and Expected events

Generator Cuts					Results	
$p_\perp$	η	$\Delta\eta$	$\Delta\phi$	m_{jj}	Filter Efficiency	Events for 10 fb^-1
50	4.8	3.5	-	1000	$8.248 \times 10^{-2} \pm 4.534 \times 10^{-3}$	$8.242 \times 10^{8} \pm 4.530 \times 10^{7}$
40	4.8	3.5	1.5	1000	$1.470 \times 10^{-2} \pm 1.914 \times 10^{-3}$	$1.469 \times 10^{8} \pm 1.913 \times 10^{7}$
45	4.8	3.5	1.5	1000	$1.121 \times 10^{-2} \pm 1.672 \times 10^{-3}$	$1.121 \times 10^{8} \pm 1.670 \times 10^{7}$
50	4.8	3.5	1.5	1000	$7.725 \times 10^{-3} \pm 1.387 \times 10^{-3}$	$7.719 \times 10^{7} \pm 1.386 \times 10^{7}$
50	4.8	3.5	1.5	800	$1.196 \times 10^{-2} \pm 1.726 \times 10^{-3}$	$1.195 \times 10^{8} \pm 1.725 \times 10^{7}$
50	4.8	3.5	1.5	900	$9.718 \times 10^{-3} \pm 1.556 \times 10^{-3}$	$9.712 \times 10^{7} \pm 1.555 \times 10^{7}$
50	4.8	3.5	1.75	1000	$9.469 \times 10^{-3} \pm 1.536 \times 10^{-3}$	$9.463 \times 10^{7} \pm 1.535 \times 10^{7}$
50	4.8	3.5	2.0	1000	$1.196 \times 10^{-2} \pm 1.726 \times 10^{-3}$	$1.195 \times 10^{8} \pm 1.725 \times 10^{7}$
50	4.8	3.5	2.25	1000	$1.495 \times 10^{-2} \pm 1.930 \times 10^{-3}$	$1.494 \times 10^{8} \pm 1.929 \times 10^{7}$
50	4.8	3.5	2.5	1000	$1.844 \times 10^{-2} \pm 2.144 \times 10^{-3}$	$1.843 \times 10^{8} \pm 2.142 \times 10^{7}$
50	4.8	3.5	2.75	1000	$2.018 \times 10^{-2} \pm 2.243 \times 10^{-3}$	$2.017 \times 10^{8} \pm 2.241 \times 10^{7}$
50	4.8	3.0	1.5	1000	$7.725 \times 10^{-3} \pm 1.387 \times 10^{-3}$	$7.719 \times 10^{7} \pm 1.386 \times 10^{7}$
50	4.8	3.0	2.0	1000	$1.196 \times 10^{-2} \pm 1.726 \times 10^{-3}$	$1.195 \times 10^{8} \pm 1.725 \times 10^{7}$

- Events passing this filters would going through RECO
- Compared with Pythia8 approach we would be sending about a factor of 2-3 less event through RECO.
- Depending on cuts a sample could be done with $\sim 80 - 120$ with no MET cut
- All cuts can be set below analysis threshold but $\Delta\phi$ cut cannot be removed from generator level.

Conclusions

Summary

- Full MadGraph production of pp to 2, 3 and 4 jets has been setup
- New MadGraph custom cuts have been implemented and tested
- Pythia8 hadronization/matching preformed and variable migrations studied
- A generator filter study has been also performed.

Proposal

- MadGraph parton cuts: $p_\perp > 30$, $m_{jj} > 800$
- Generator cuts: Dijet $p_\perp > 50$ GeV, $\eta < 4.8$, $\Delta\eta > 3.0$, $\Delta\phi < 2.0$ and $m_{jj} > 1000$
- Offline cuts: none

Proposed course of action

- Propose this to PPD (this thick all the boxes, less Pythia running, better filter efficiency, less RECO)
- Pass custom MadGraph code to Chayanit for gridpark production
- Pass generator filter code to Chayanit for CMSSW integration
- Wait for samples to be produced :)