

# Towards discrimination and improved modelling of dark-sector showers

*Sukanya Sinha*

University of Witwatersrand, Johannesburg

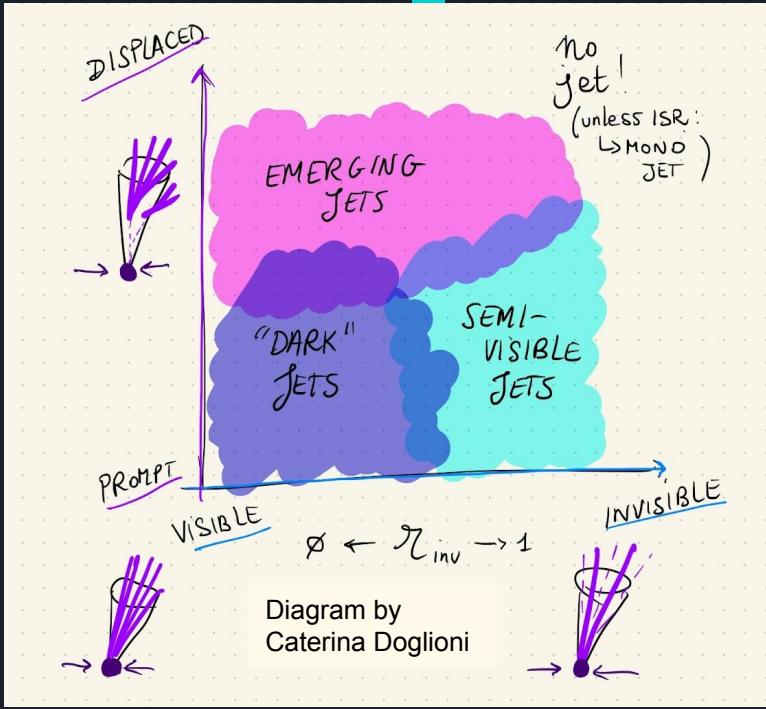
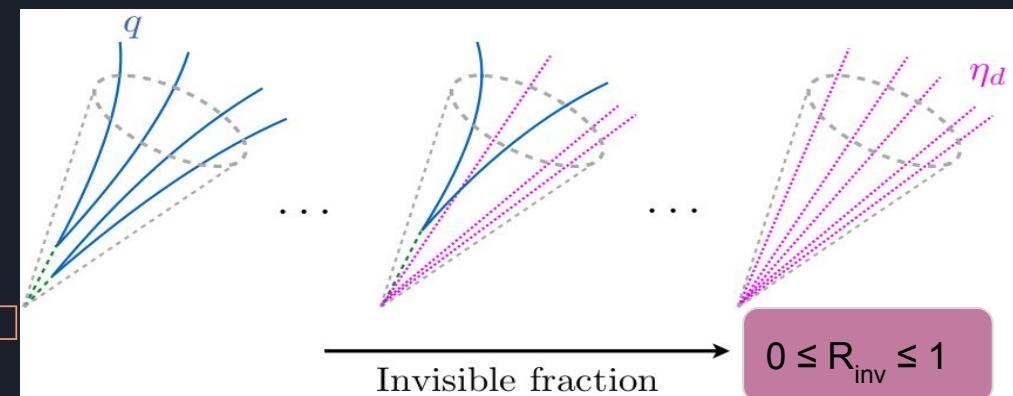
23rd MCnet Meeting  
December 2021

Alongwith Andy Buckley



# Introduction

- Look at unusual topologies & hidden phase space corners
- Dark hadrons decay promptly in a QCD-like fashion partially back to visible sector (**semi-visible jets “SVJ”**) Initial motivation
  - Showering using Pythia hidden valley module

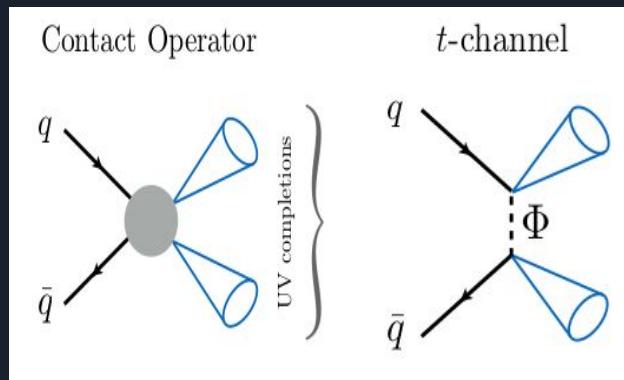
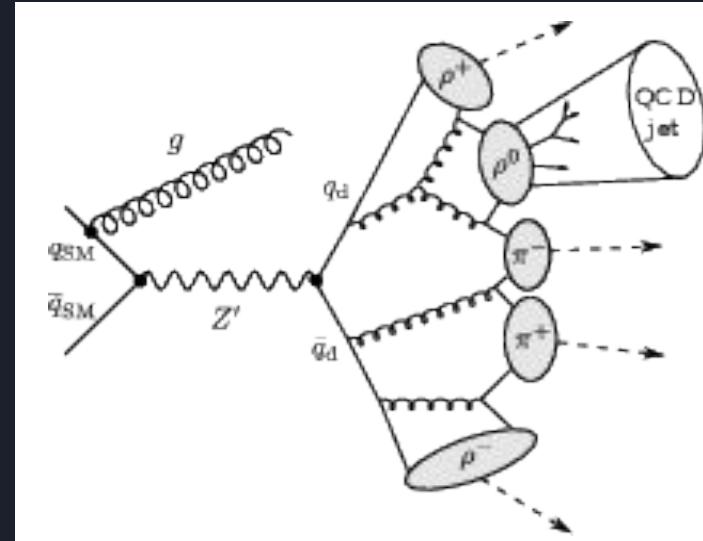


Based on the Paper:  
**LHC Searches for Dark Sector Showers** : Tim Cohen et al [arXiv: [1707.05326](https://arxiv.org/abs/1707.05326)]

# Hidden Valley: Semi-visible jets idea

Two different dark quark flavours

- ▶ Combine to form  $\pi^+$ ,  $\pi^-$ ,  $\pi^0$ , and  $\rho^+$ ,  $\rho^-$ ,  $\rho^0$  (assumed to be produced thrice as much as pions)
- ▶ Only  $\rho^0$  is unstable and (promptly) decays to SM quarks: more likely to decay to b pairs due to need for a mass insertion, to make the angular momentum conservation work out
- ▶ Other mesons are (collider-)stable → invisible



## Model Parameters:

1.  $M_\phi$  = Mass of Scalar Bi - fundamental (mediator)
2.  $M_d$  = Mass of dark hadrons
3.  $r_{inv}$  = no. of stable dark hadrons/ no. of hadrons

Signal xs very low compared to BG →  
More of a topology generator rather than full-blown theory model

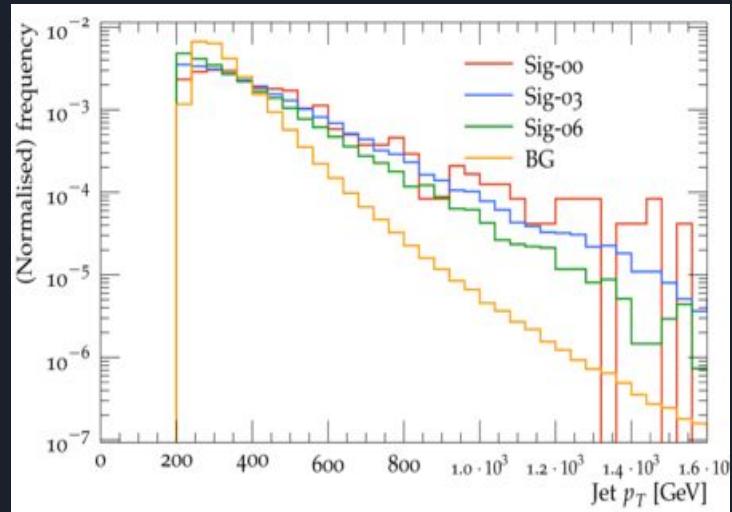
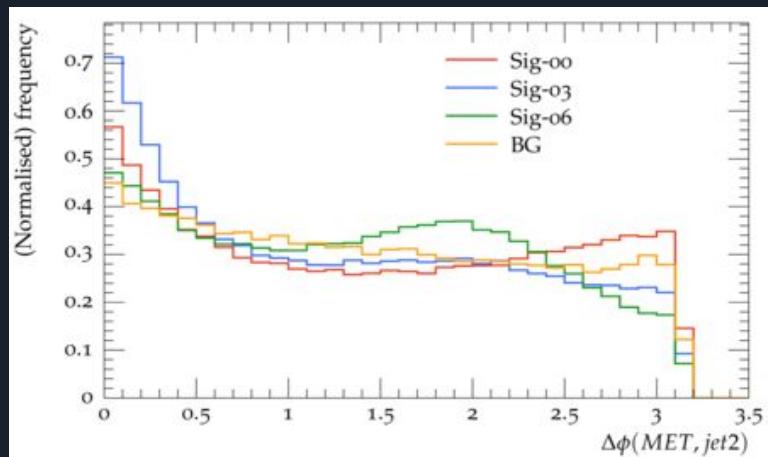
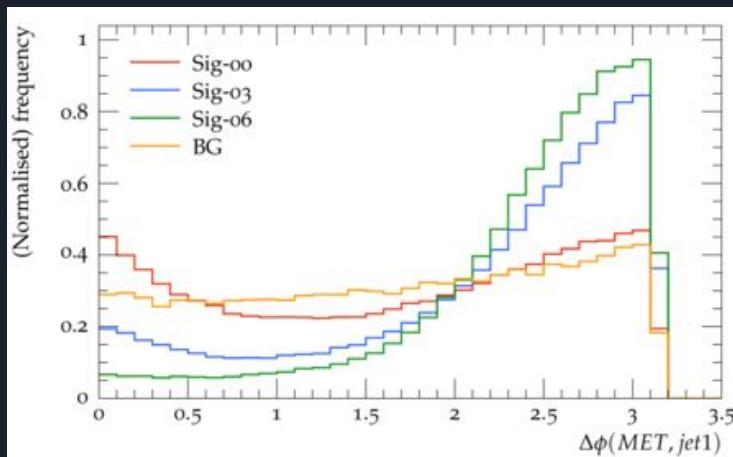
# Jet-substructure study

D.Kar & SS: [SciPostPhys.10.4.084](#)

- Comparing jet substructure variables to see if SVJ substructure is different from light quark/gluon jets (BG). Do they behave more multi-pronged as opposed to mostly single prong?
- Comparison can be done in  $p_T$  bins or in  $m/p_T$  bins, picked the former, as there is no resonance.

*t-channel makes it more challenging as no resonance peak*

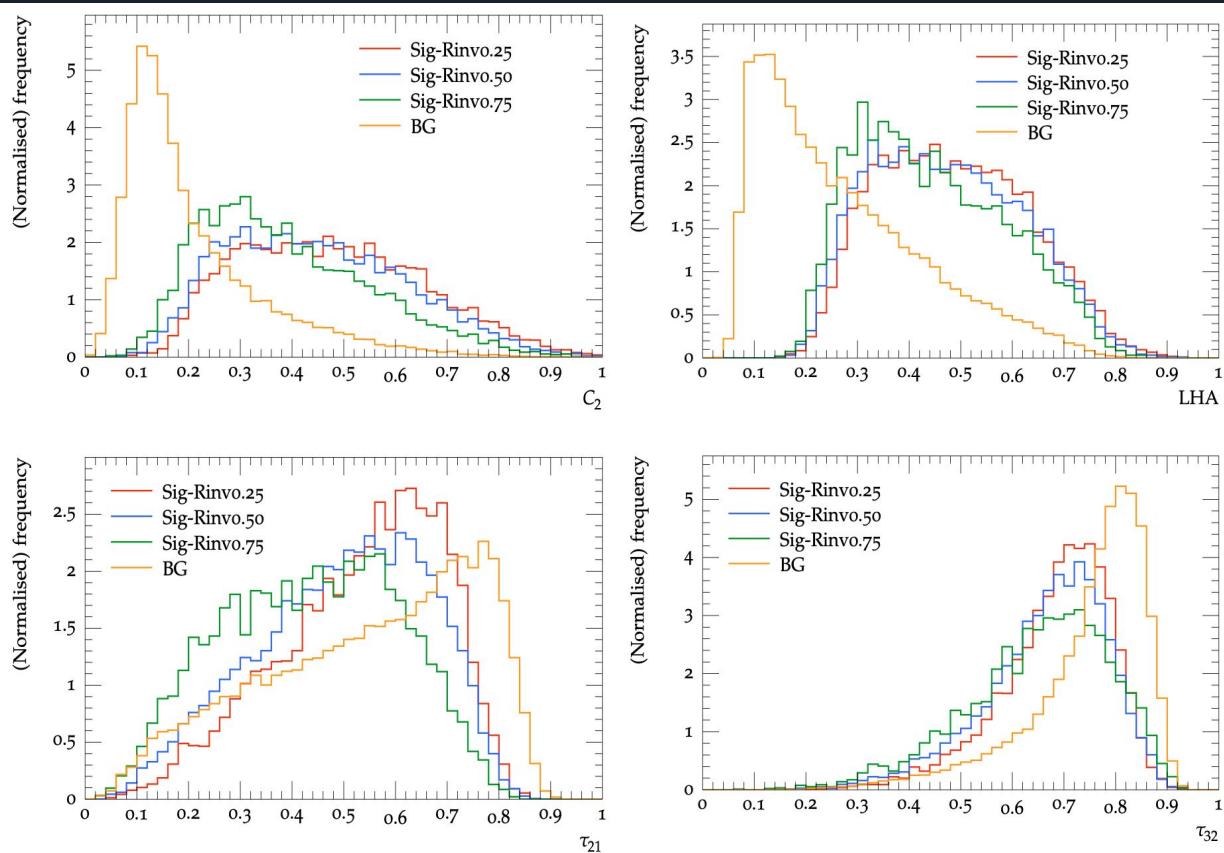
Subleading jets tend to align more with MET, which makes it harder to study



Signals  
( $r_{\text{inv}} = 0, 0.3,$   
 $0.6$ ) and multijet  
background  
generated using  
MG5 + Py8

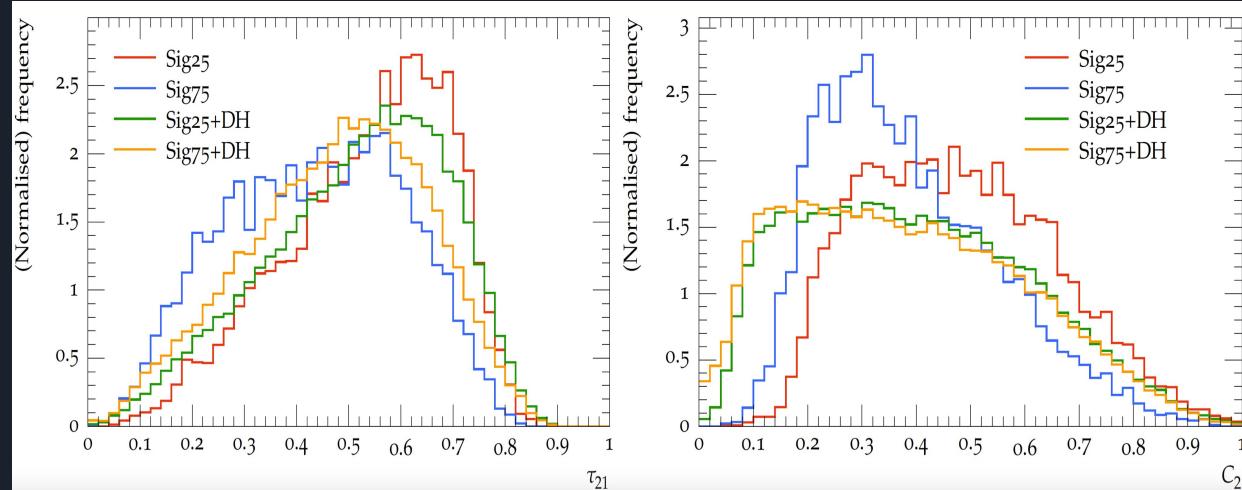
Normally signals are  
generated with upto two  
extra jets!

# What effects are responsible for specific jet-substructure of semi-visible jets?



For finite  $r_{\text{inv}}$  values, when only the visible hadrons are clustered in jets, subtle substructure difference observed for different Rinv values.

# What effects are responsible for specific jet-substructure of semi-visible jets?



If the final dark hadrons are also clustered in the jets ---> expect this difference to go away ----> the different amount of missing hadrons in each case presumably is responsible for the difference.

## Conclusions:

1. The substructure becomes less two-pronged with visible and dark hadrons in them, and the absence of the dark hadrons create the two-pronged structure → The substructure is created by the interspersing of visible hadrons with dark hadrons.
2. Specific hidden valley parameter configurations can reduce the dark shower model dependent features of the signal jets.

# However, the story does not end here....

- Several other observables may be out there, that can help discriminate these unconventional jets from the standard q/g jets
- Mostly due to presence of only one dark shower module, so far, all studies are somewhat model dependent

However, the story does not end here....

- **Several other observables may be out there, that can help discriminate these unconventional jets from the standard q/g jets**
- Mostly due to presence of only one dark shower module, so far, all studies are somewhat model dependent

# Energy - flow polynomials

- A complete linear basis for jet substructure

J. Thaler et al

<https://arxiv.org/pdf/1712.07124.pdf>

What are EFPs?

- Observables that are multiparticle energy correlators with specific angular structures which directly result from IRC safety

EFPs form a linear basis of all IRC-safe observables, making them suitable for a wide variety of jet substructure contexts where linear methods are applicable

For a multigraph  $G$  with  $N$  vertices and edges  $(k, l) \in G$ , the corresponding EFP takes the form:

$$\text{EFP}_G = \sum_{i_1=1}^M \cdots \sum_{i_N=1}^M z_{i_1} \cdots z_{i_N} \prod_{(k,\ell) \in G} \theta_{i_k i_\ell}$$

where the jet consists of  $M$  particles,  $z_i$  is the energy fraction carried by particle  $i$ , and  $\theta_{ij}$  is the angular distance between particles  $i$  and  $j$ .

Set of EFPs can be considered as the energy flow basis.

# How to compute EFPs?

$$\bullet_j \iff \sum_{i_j=1}^M z_{i_j} \quad k \xrightarrow{\hspace{1cm}} \ell \iff \theta_{i_k i_\ell}$$

Each edge  $(k, l)$  in a multigraph is in one-to-one correspondence with a term  $\theta$  in an angular monomial

Each vertex  $j$  in the multigraph corresponds to a factor of  $z$  and summation over  $i_j$  in the EFP

# How to compute EFPs?

$$\bullet_j \iff \sum_{i_j=1}^M z_{i_j} \quad k \xrightarrow{\hspace{1cm}} \ell \iff \theta_{i_k i_\ell}$$

Each edge  $(k, l)$  in a multigraph is in one-to-one correspondence with a term  $\theta$  in an angular monomial

Each vertex  $j$  in the multigraph corresponds to a factor of  $z$  and summation over  $i_j$  in the EFP

$$\bullet - \bullet = \sum_{i_1=1}^M \sum_{i_2=1}^M z_{i_1} z_{i_2} \theta_{i_1 i_2}$$

2 particles/constituents in a jet  $\rightarrow$  2 energy fractions, 1 angularity value  $\rightarrow$  degree 1 polynomial

# How to compute EFPs?

$$\bullet_j \iff \sum_{i_j=1}^M z_{i_j} \quad k \xrightarrow{\hspace{1cm}} \ell \iff \theta_{i_k i_\ell}$$

Each edge  $(k, l)$  in a multigraph is in one-to-one correspondence with a term  $\theta$  in an angular monomial

Each vertex  $j$  in the multigraph corresponds to a factor of  $z$  and summation over  $i_j$  in the EFP

$$= \sum_{i_1=1}^M \sum_{i_2=1}^M \sum_{i_3=1}^M \sum_{i_4=1}^M z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}$$

4 particles/constituents in a jet  $\rightarrow$  4 energy fractions, 5 angularity values  $\rightarrow$  degree 5 polynomial

Because the EFP basis is infinite, a suitable organization and truncation scheme is necessary to use the basis in practice.

# How to compute EFPs?

$d$	1	2	3	4	5	6	7	8	9	10
$N$	<b>2</b>	1	1	1	1	1	1	1	1	1
	<b>3</b>		1	2	3	4	6	7	9	11
	<b>4</b>			2	5	11	22	37	61	95
	<b>5</b>				3	11	34	85	193	396
	<b>6</b>					6	29	110	348	969
	<b>7</b>						11	70	339	1 318
	<b>8</b>							23	185	1 067
	<b>9</b>								47	479
	<b>10</b>									3 294
	<b>11</b>									106
										235

Several combinations for diagrams possible

# How to compute EFPs?

$d$	1	2	3	4	5	6	7	8	9	10
2	1	1	1	1	1	1	1	1	1	1
3		1	2	3	4	6	7	9	11	13
4			2	5	11	22	37	61	95	141
5				3	11	34	85	193	396	771
$N$					6	29	110	348	969	2445
						11	70	339	1318	4457
8							23	185	1067	4940
9								47	479	3294
10									106	1279
11										235

Several combinations for diagrams possible

Restricting ourselves to this regime...

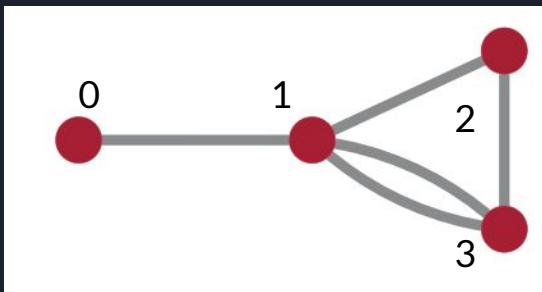
The linked paper discusses in great detail how different EFP combinations lead to well-known jss observables

# What we ideally want to achieve with EFPs?

- Implement EFPs in Rivet and see if any particular combination of EFPs helps to distinguish between standard q/g jets and more unconventional jets
  - Might lead to a new jet-substructure observable for dark shower discrimination
- Setup working ( computing EFP multigraphs till  $N = 7, d = N - 1, N, N + 1$  ) → python code taking into account the different possible orientations of the input “particles” and designing an array of possible EFP diagrams as a grid

# What we ideally want to achieve with EFPs?

- Implement EFPs in Rivet and see if any particular combination of EFPs helps to distinguish between standard q/g jets and more unconventional jets
  - Might lead to a new jet-substructure observable for dark shower discrimination
- Setup working ( computing EFP multigraphs till  $N = 7, d = N - 1, N, N + 1$ ) → python code taking into account the different possible orientations of the input “particles” and designing an array of possible EFP diagrams as a grid

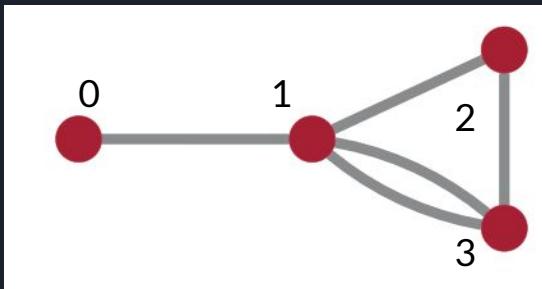


EFP diagram to “particle” pair translation

$\{\{0, 1\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 3\}\}$

# What we ideally want to achieve with EFPs?

- Implement EFPs in Rivet and see if any particular combination of EFPs helps to distinguish between standard q/g jets and more unconventional jets
  - Might lead to a new jet-substructure observable for dark shower discrimination
- Setup working ( computing EFP multigraphs till  $N = 7, d = N - 1, N, N + 1$ ) → python code taking into account the different possible orientations of the input “particles” and designing an array of possible EFP diagrams as a grid



EFP diagram to “particle” pair translation

$N=4, d = N - 2$	0	1	2	3
0	-	-	-	-
1	1	-	1	2
2	-	-	-	-
3	-	-	1	-

$\{\{0, 1\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 3\}\}$

# What we ideally want to achieve with EFPs?

- Individual EFP diagrams use  $R=0.2$  anti- $k_T$  subjets, from the leading jet in each event, as inputs → and the log-likelihood ratio value between pseudodata (Signal + SM background) and MC (SM background) is obtained after a very inclusive particle-level analysis ( no MET threshold, jet  $pT > 50$  GeV,  $|\eta| < 4.9$ , lepton vetoed)
- Gitlab repo: [link](#)

# Model parameters and Statistical analysis summary

**Signal:** semi-visible jets t-channel

1. PS - Py8 Hidden Valley module,  $R_{inv} = \#stable\ dark\ hadrons / \#hadrons = 0.4$ ,  $M_{med} = 1.5, 3\ TeV$

2.  $\alpha_D$  running,  $\lambda_D = 0.1$ ,  $pTminFSR = 1.1$  (can tweak different combinations for subsequent rounds)

3. Generated with upto two extra jets in ME level, MLM merging scheme applied

**BG** - Pythia dijet

Plotting for several EFP diagrams, of 20 bins each, and y axis is defined by the formula:

$\mu = 1$  in this case...

$$\ln(L(\mu)) = \sum_i^N n_i \ln(\mu s_i + b_i) - (\mu s_i + b_i),$$

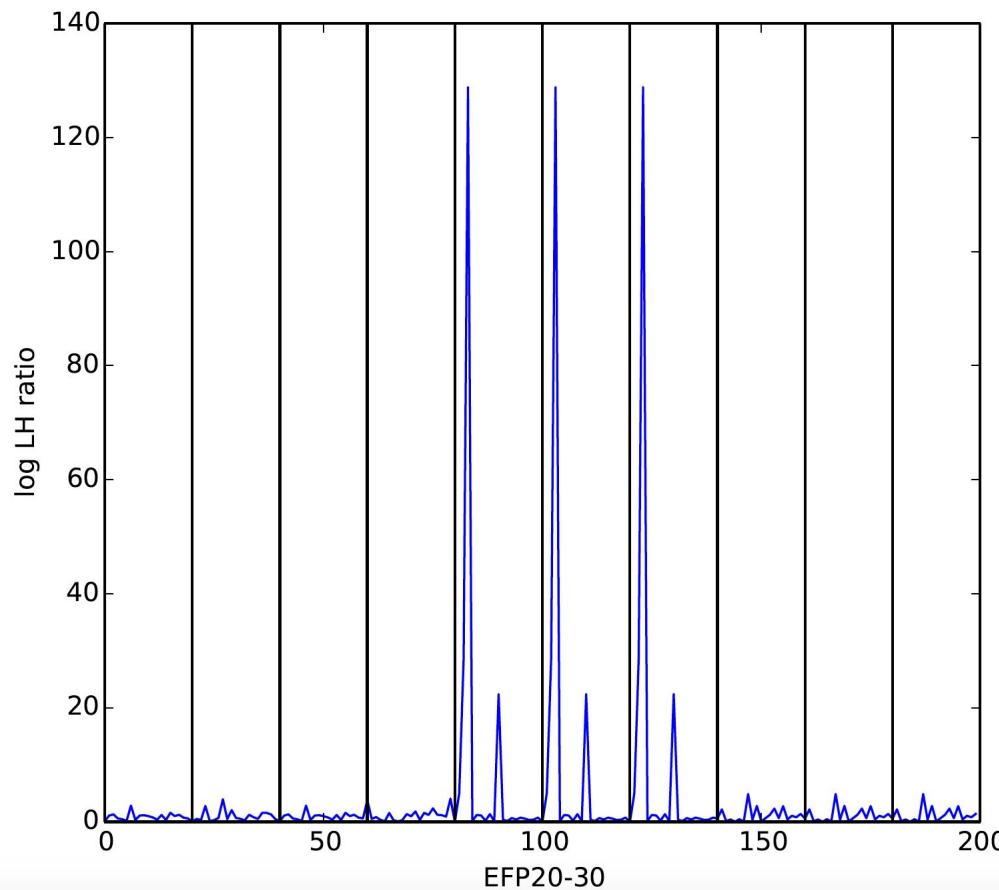
$$\ln L(0) = N_i * \ln(B) - B$$

$$\ln L(1) = N_i * \ln(S+B) - (S+B)$$

$$q_0 = -2 \ln \left( \frac{L(0)}{L(\hat{\mu})} \right)$$

where  $N_i$  = asimov data = expected  
signal+bkg events

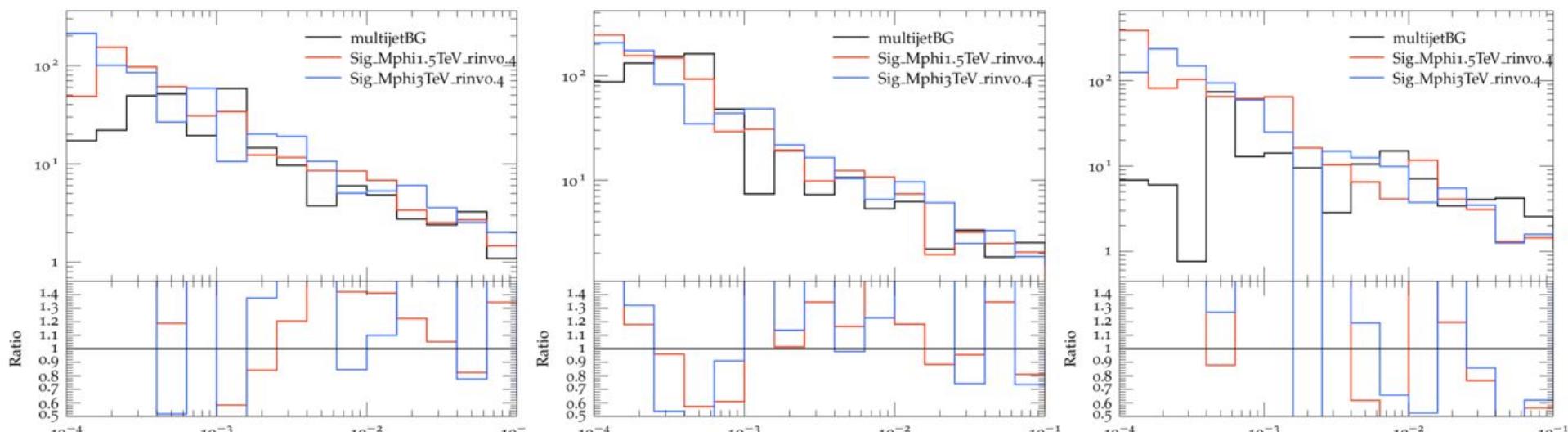
# Log-Likelihood ratio distributions



Certain EFP diagrams seem to have some bins of the jet-shape observables that QCD just doesn't populate at all, in which the DM signal dominates...

Compared several of these LLR distributions for identifying some distinct LLR that deviate from "SM" (multijet BG is treated as a null-hypothesis)

# Some key standalone EFP diagrams of interest



EFP 3

$\{\{0, 1\}, \{0, 2\}, \{1, 0\}, \{1, 2\}, \{2, 0\}, \{2, 1\}\}$

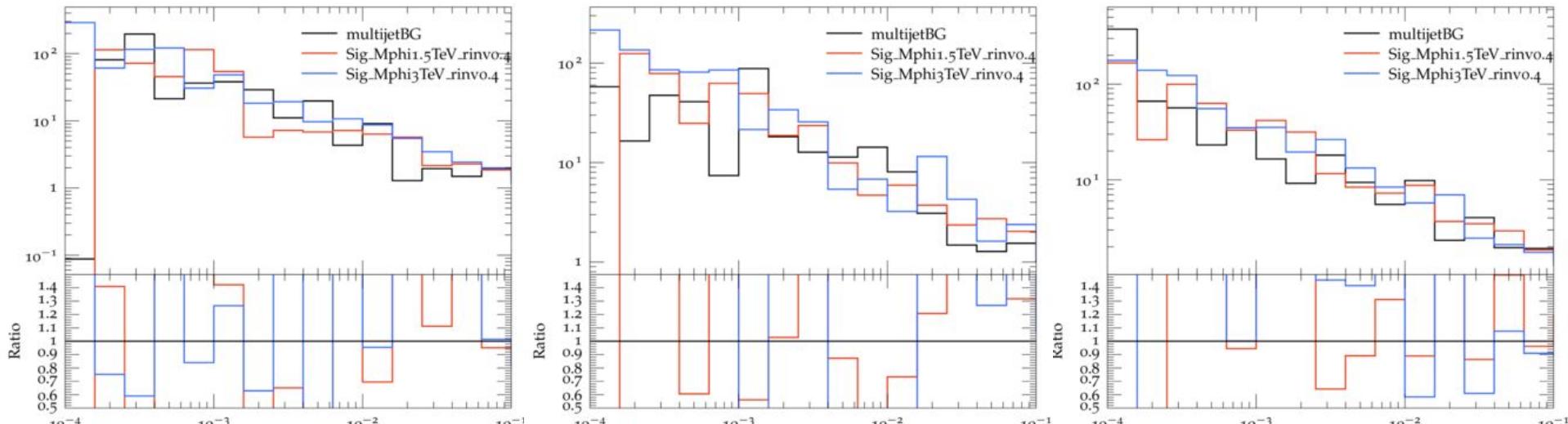
EFP 6

$\{\{0, 2\}, \{1, 2\}, \{1, 2\}, \{2, 0\}, \{2, 1\}, \{2, 1\}\}$

EFP 7

$\{\{0, 1\}, \{0, 1\}, \{0, 2\}, \{1, 0\}, \{1, 0\}, \{1, 2\}, \{2, 0\}, \{2, 1\}\}$

# Some key standalone EFP diagrams of interest



EFP 13

$\{\{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 0\}, \{1, 2\}, \{2, 0\}, \{2, 1\}, \{3, 0\}\}$

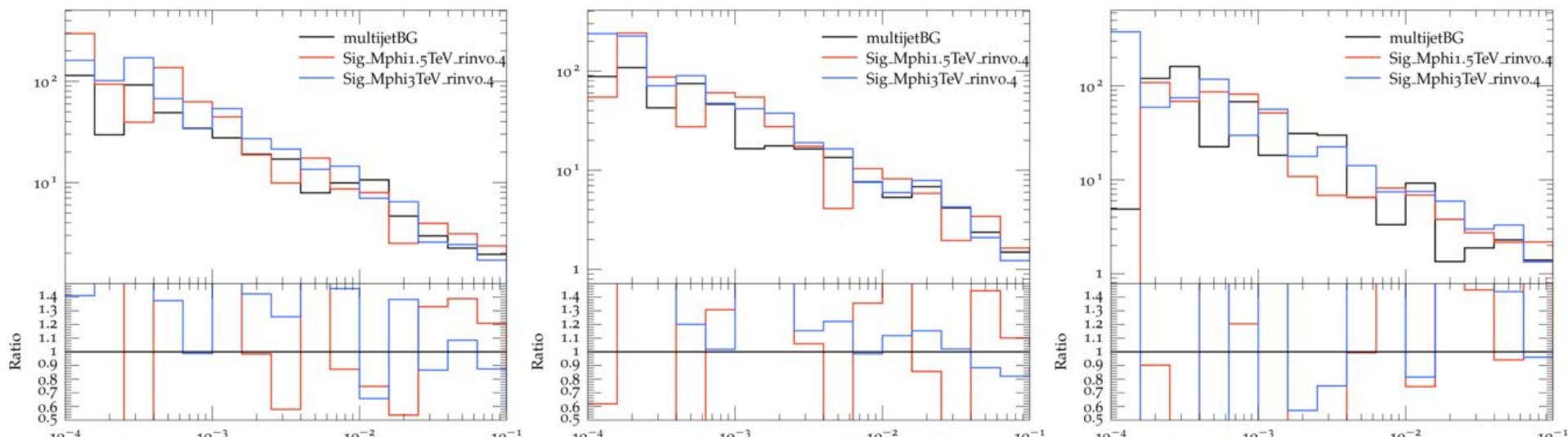
EFP 14

$\{\{0, 2\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \{2, 0\}, \{2, 0\}, \{2, 1\}, \{3, 0\}\}$

EFP 17

$\{\{0, 1\}, \{0, 2\}, \{0, 3\}, \{0, 3\}, \{1, 0\}, \{2, 0\}, \{3, 0\}, \{3, 0\}\}$

# Some key standalone EFP diagrams of interest



EFP 18

$\{\{0, 1\}, \{0, 3\}, \{0, 3\}, \{1, 0\}, \{1, 2\}, \{2, 1\}, \{3, 0\}, \{3, 0\}\}$

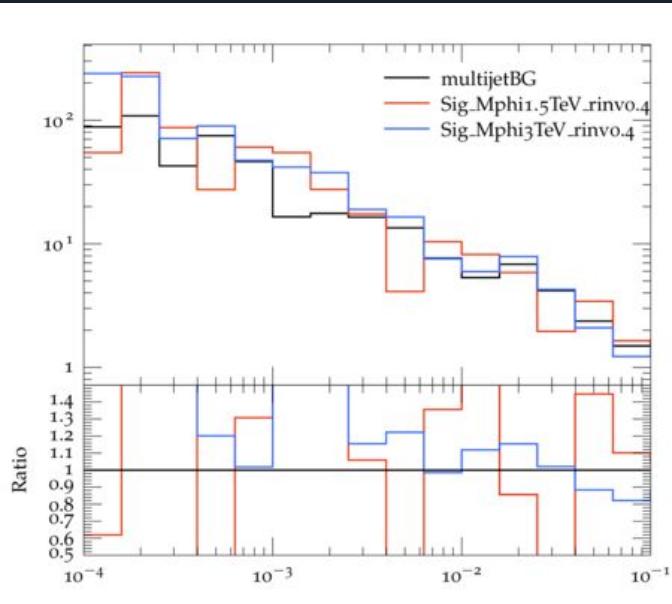
EFP 25

$\{\{0, 1\}, \{0, 2\}, \{0, 2\}, \{1, 0\}, \{1, 2\}, \{1, 3\}, \{2, 0\}, \{2, 0\}, \{2, 1\}, \{3, 1\}\}$

EFP 26

$\{\{0, 1\}, \{0, 2\}, \{1, 0\}, \{1, 2\}, \{1, 2\}, \{1, 3\}, \{2, 0\}, \{2, 1\}, \{2, 1\}, \{3, 1\}\}$

# Some key standalone EFP diagrams of interest



EFP 28

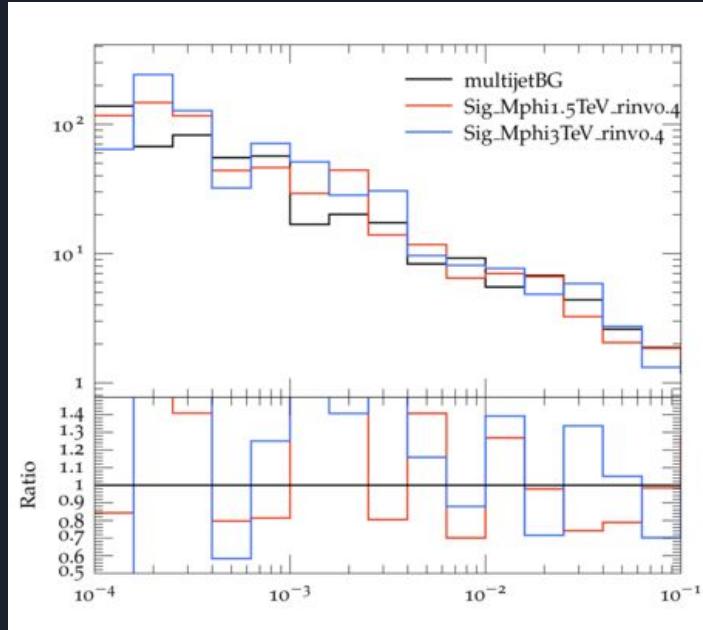
$\{\{0, 1\}, \{0, 2\}, \{0, 2\}, \{0, 3\}, \{1, 0\}, \{1, 3\}, \{2, 0\}, \{2, 0\}, \{3, 0\}, \{3, 1\}\}$

The default model predicts an invisibly small signal/background ratio

→ we treat it more as a signature generator, and shape comparisons are done with this particular setup

Treating multijet BG as a constant → mapping out how large an xsection would be needed, and which EFP would dominate → to exclude each model point in a scan...

There is a flexibility in the phenomenology as a function of model parameters, see [Danielle Wilson's presentation in Snowmass](#)



EFP 29

$\{\{0, 1\}, \{0, 1\}, \{0, 3\}, \{1, 0\}, \{1, 0\}, \{1, 2\}, \{1, 3\}, \{2, 1\}, \{3, 0\}, \{3, 1\}\}$

# However, the story does not end here....

- Several other observables may be out there, that can help discriminate these unconventional jets from the standard q/g jets
- **Mostly due to presence of only one dark shower module, so far, all studies are somewhat model dependent**

# Necessity of an additional dark-shower module

## Why Herwig?

- Herwig is using systematically different shower and hadronization models from Pythia
  - ◆ Herwig dark hadronization would naturally be cluster- rather than string-based, hence there really is a model difference.
- Helps to get a better estimate of theory associated uncertainties.
- Apart from the standard dark confinement scale, and the renormalisation scale multiplier in the shower, different hadronization procedure might also play a key role, and we suspect that the difference between cluster and string models could be similar to the individual physically feasible variations within either of them.

# Preliminary plan for implementation of Herwig HV hadronisation module

alongwith Aidin Masouminia, Andreas Papaefstathiou,  
Andrzej Siódmiak, Deepak Kar and Simon Plaetzar

- Start with dark mesons “only”
  - ◆ Masses comparable to QCD scale
- Have a separate dark hadronization prehandler, that mimics the SM hadronization
  - ◆ sans the fission, and other complicated machinery at play in the SM part
- This prehandler runs before the standard model hadronization handler

Thanks to all!

Because of completely different dark-sector radiation & hadronization relating to Pythia, it will help to better estimate model uncertainties.

Efforts ongoing to prepare a very minimal design of H7 HV module, however, still not at a phenomenological testing stage...