Search for semi-visible jets in ATLAS and CMS

Sukanya Sinha
University of Manchester
(on behalf of ATLAS & CMS collaborations)

LHCP 2023, Belgrade

26/05/2023

The big picture!

We have not found any concrete sign of new physics … yet!

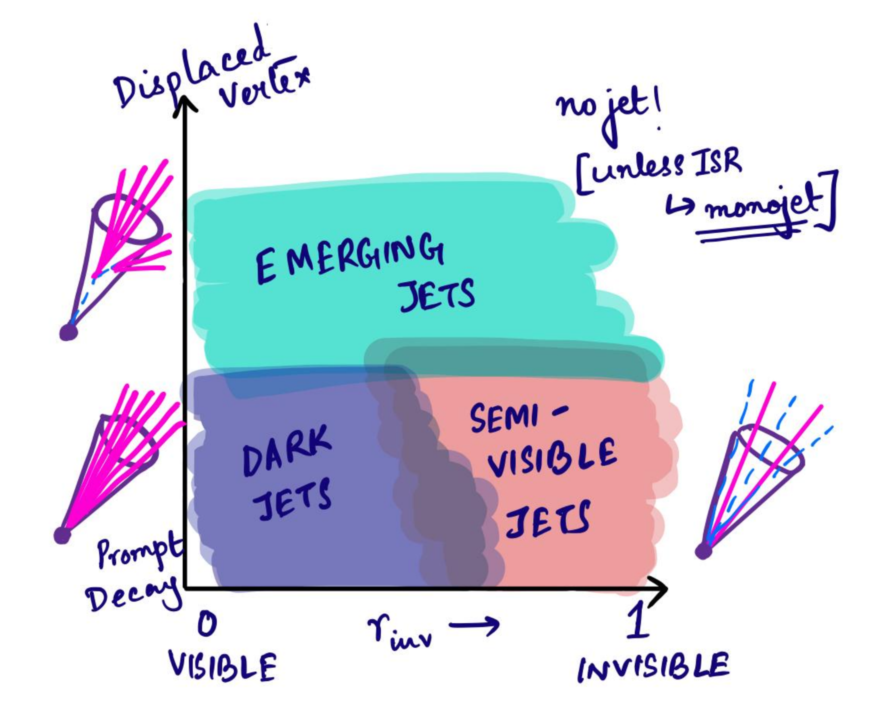
Looking at unusual topologies and hidden corners of the phase space

→ **signature based searches, using benchmark models.**

Dark hadrons decaying PROMPTLY in a QCD-like fashion, fully (dark jets)

or partially back to visible sector (semi-visible jets)

Dark hadrons undergoing DISPLACED decays in a QCD-like fashion (emerging jets)



source

Showering using Pythia hidden valley module: at best a guesstimate!

Pythia 8 Hidden Valley Module

Different dark quark flavours

- Combine to form π_d^+, π_d^-, π_d^0, and ρ_d^+, ρ_d^-, ρ_d^0 (assumed to be produced thrice as much as pions)

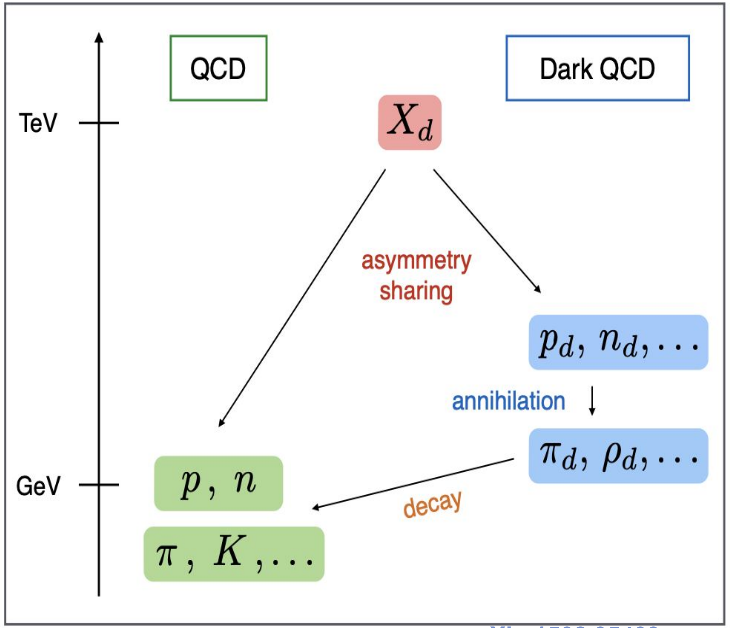
- Only ρ_d^0 is unstable and (promptly) decays to SM quarks: more likely to decay to b pairs due to need for a mass insertion, to make the angular momentum conservation work out

- Other mesons are (collider-)stable → invisible

Pythia 8 Hidden Valley Module

Different dark quark flavours

- Combine to form π_d^+, π_d^-, π_d^0, and ρ_d^+, ρ_d^-, ρ_d^0 (assumed to be produced thrice as much as pions)
- Only ρ_d^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



arXiv:1502.05409

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

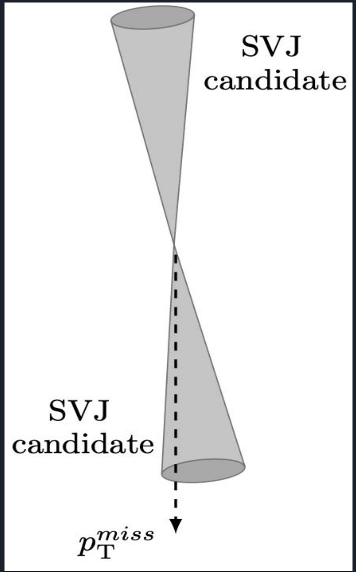
Decay chains are rather complex and the showering model is still being developed by the theory community

Baryon and DM asymmetries shared via a mediator X_d → asymmetry in stable dark baryons.

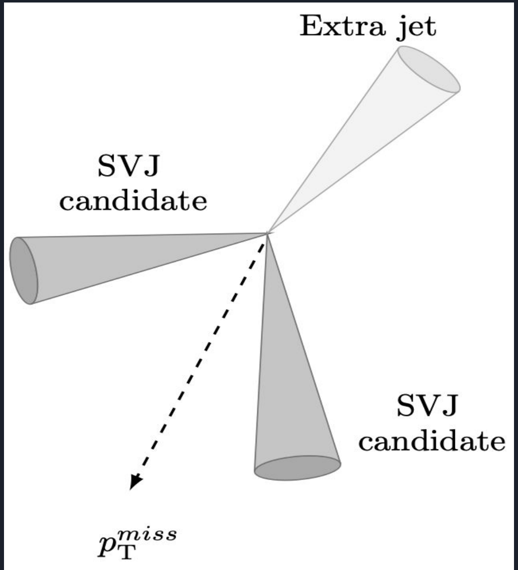
The symmetric relic density annihilated into dark pions → decay into SM particles.

Correct DM relic density obtained when dark baryon masses are in the 10 GeV range.

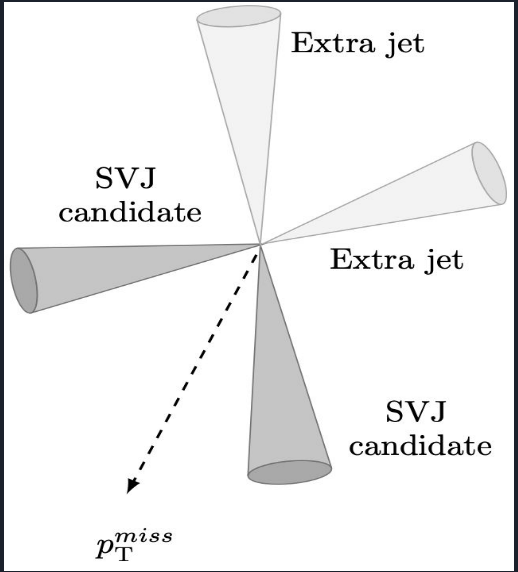
The semi-visible jet topology



At Leading Order



With one extra jet

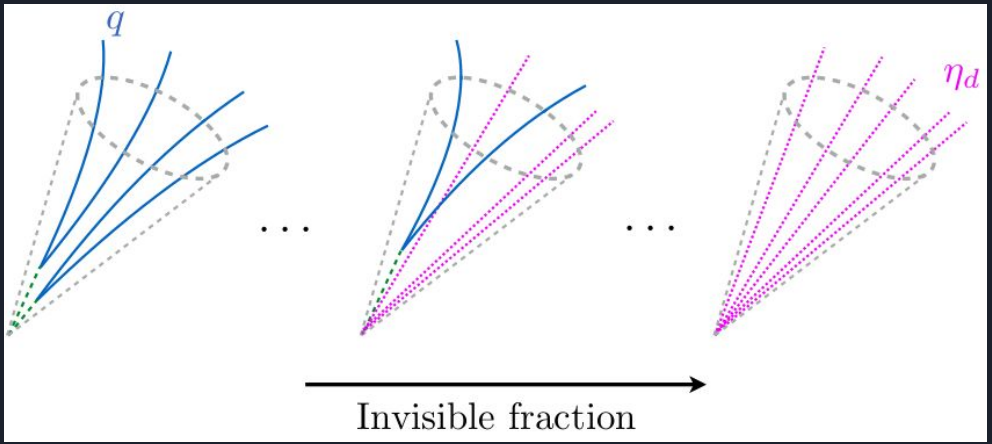


With two extra jets

Length of the cones do not represent the visible energy of the particles, and invisible energies are expected in the directions of the two SVJ candidates.

Semi-visible jets in CMS

Semi-visible jet production



Model Parameters:

1. $M_{Z'}$ = Mass of Z' mediator
2. R_{inv} = #stable/(#stable + #unstable)
3. M_d = Mass of dark hadrons
4. α_{dark} = running coupling of dark QCD

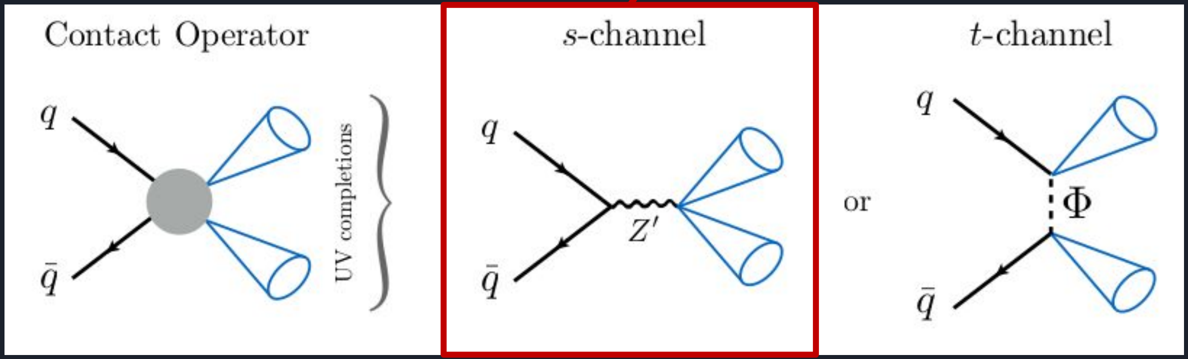
Link to the paper: https://arxiv.org/abs/1503.00009

Signal samples: Pythia8 with

R_{inv} = 0 - 0.8 and M_d = 1 - 100 GeV,

$M_X = M_d/2$, g_q = 0.25, g_X = 0.5,

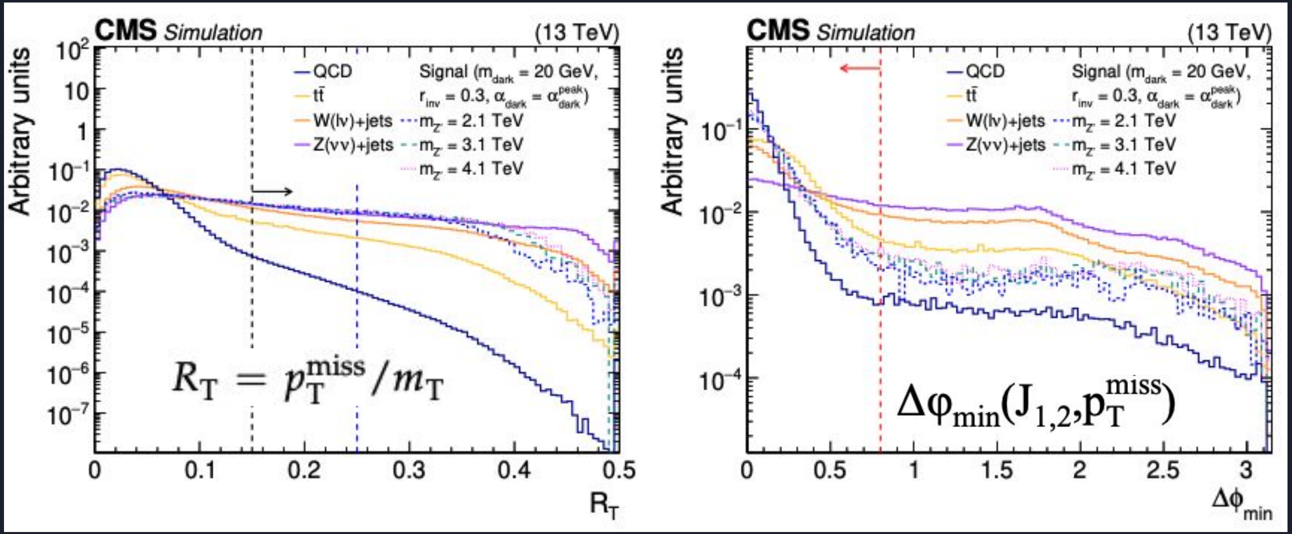
$M_{Z'}$ = 1.5 - 5.1 TeV



Analysis Strategy

Two-fold analysis strategy:

- Inclusive search: using only event-level kinematic observables (applicable for models with similar kinematic behaviour)
- BDT based: optimised for semi-visible jet tagger (assumes chosen signal models are accurate)

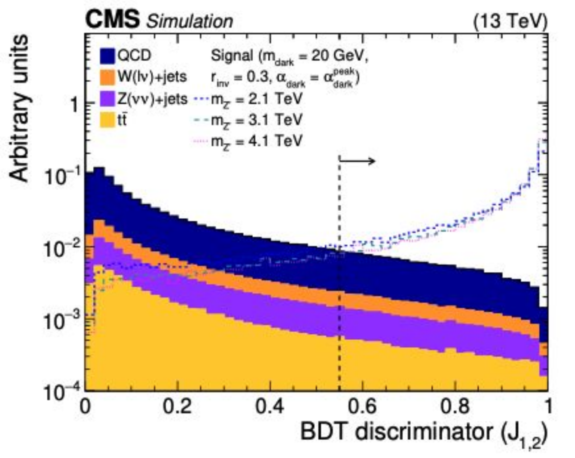


Looking at two central R=0.8 jets, with large radius jet or H_T triggers.

m_T is the search variable, with selections applied based on high R_T (used to uncorrelate p_T^{miss} and m_T) and low $\Delta\varphi_{min}$ thresholds.

$$m_T^2 = m_{JJ}^2 + 2p_T^{miss}\left[\sqrt{m_{JJ}^2 + p_{T,JJ}^2} - p_{T,JJ}\cos(\phi_{JJ,miss})\right]$$

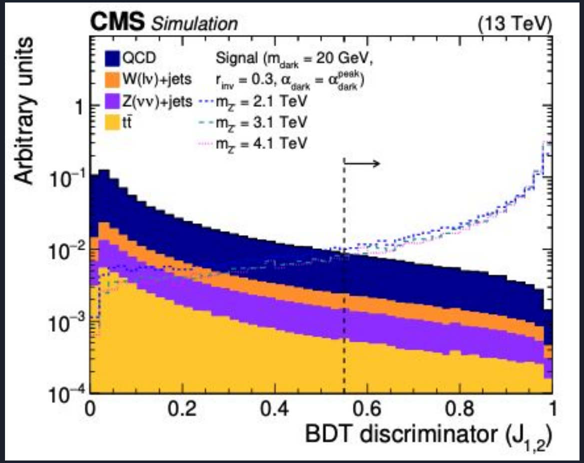
Analysis Strategy



Semi-visible jet identification:

- Used to discriminate between svj and light q/g jets
- 15 input variables (associated to jet substructure, flavour, q/g discrimination) combined in a BDT
  - Two highest *p*T jets from simulated signal and background samples are inputs, with the 15 variables computed for each jet.
  - final WP = 0.55; the BDT rejects 84–88% of simulated bkg jets, while correctly classifying 87% of jets from the benchmark signal model

Analysis Strategy



Semi-visible jet identification:

- Used to discriminate between svj and light q/g jets
- 15 input variables (associated to jet substructure, flavour, q/g discrimination) combined in a BDT
  - Two highest p_T jets from simulated signal and background samples are inputs, with the 15 variables computed for each jet.
  - final WP = 0.55; the BDT rejects 84–88% of simulated bkg jets, while correctly classifying 87% of jets from the benchmark signal model

Define two signal regions (SRs) for the inclusive search: low-R_T ($0.15 < R_T \leq 0.25$) and high- R_T ($R_T > 0.25$).

For BDT: subsets of the high-R_T and low-R_T inclusive SRs are selected by requiring both jets in each event are tagged as semi-visible – termed as high-SVJ2 and low-SVJ2.

Fit the observed m_T distribution in each signal region with an analytic smoothly falling functional form:

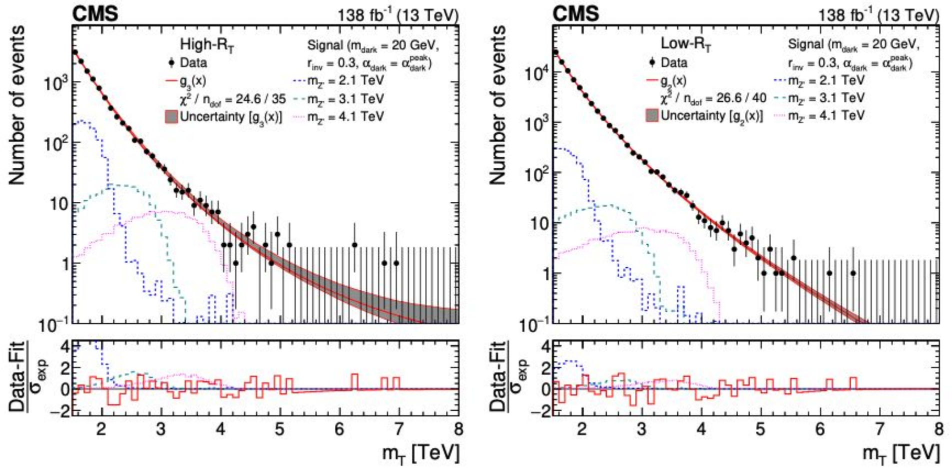
$$g(x) = \exp(p_1 x) x^{p_2 [1 + p_3 \ln(x)]}$$

With p_i being free parameters in the fit, and normalisation is allowed to freely float.

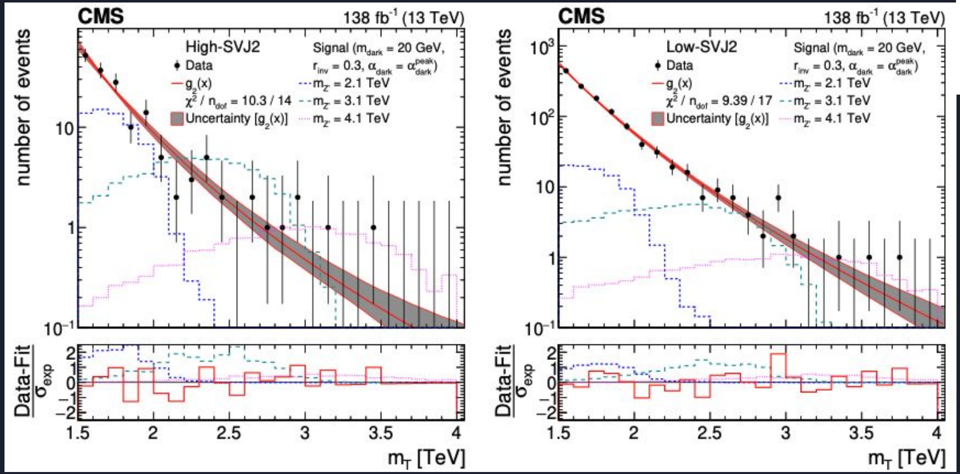
Results JHEP 06 (2022) 156, CDS

Background-only fits are compared to the observed data for:

the inclusive signal regions



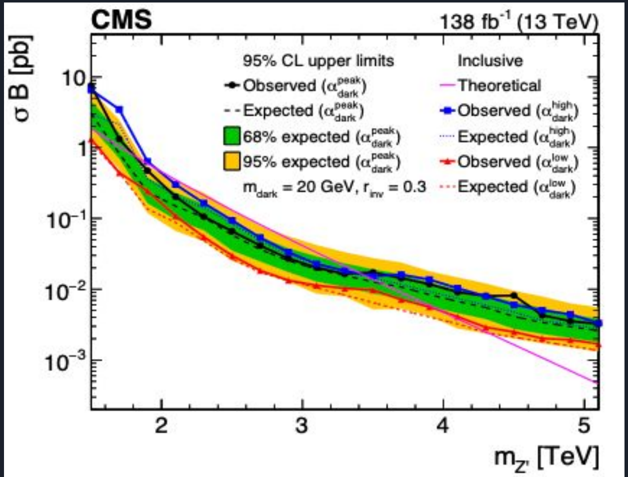
BDT - based regions

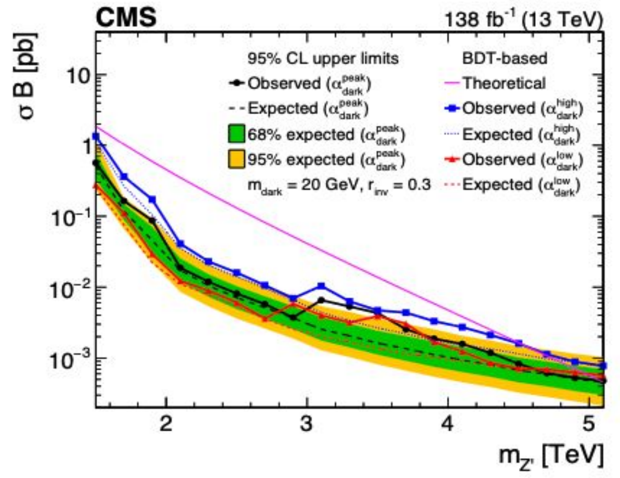


Lower panel shows the difference between the observation and the prediction divided by the statistical uncertainty in the observation

Large improvement compared to analysis without BDT identification of semi-visible jets

Results [JHEP 06 (2022) 156](), [CDS]()





Inclusive signal regions

Excludes models with

1.5 < $M_{Z'}$ < 4.0 TeV and

0.07 < R_{inv} < 0.53

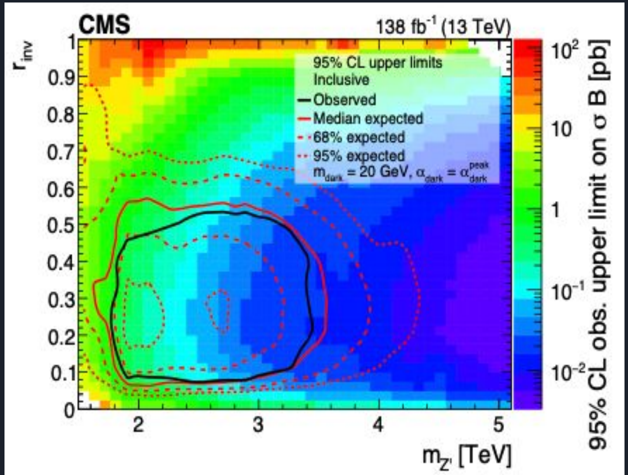
BDT - based regions

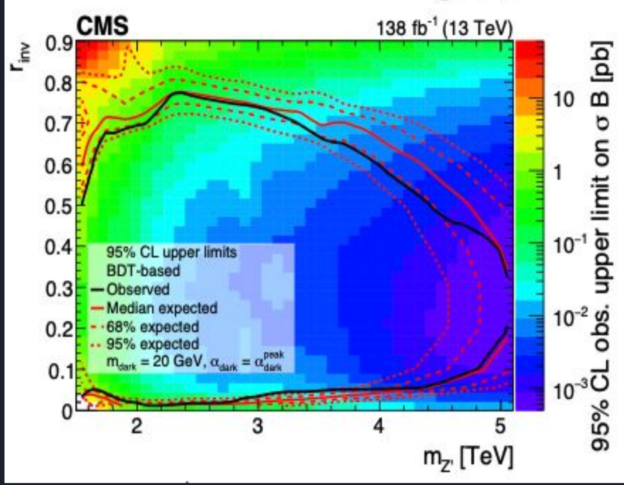
Excludes models with

1.5 < $M_{Z'}$ < 5.1TeV and

0.01 < R_{inv} < 0.77

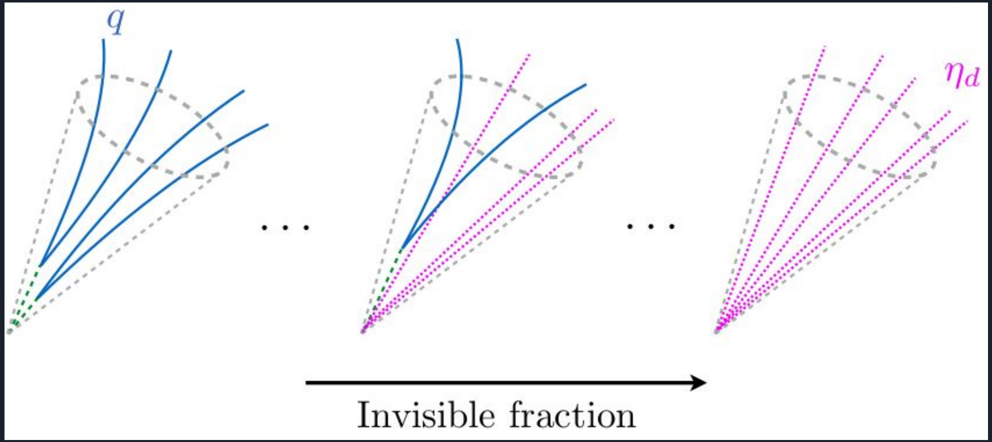
For M_{dark} = 20 geV





Semi-visible jets in ATLAS

Semi-visible jet production



Model Parameters:

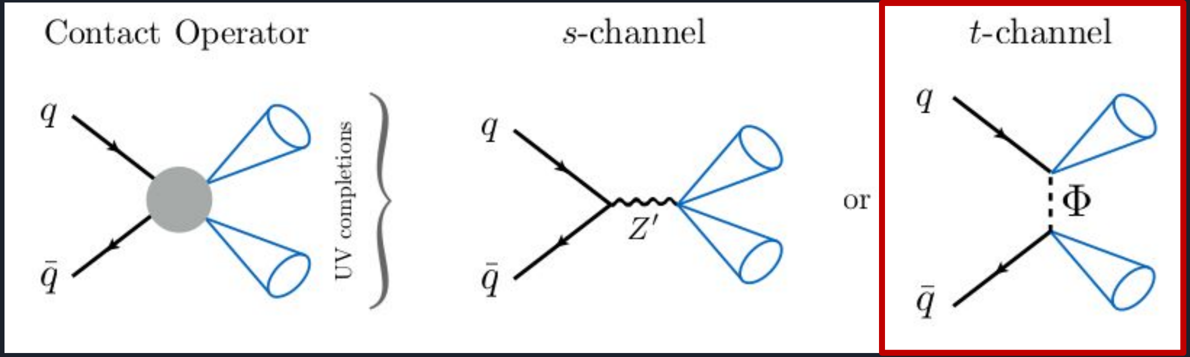
1. M_ϕ = Mass of scalar bi-fundamental mediator
2. r_{inv} = #stable dark hadrons/#all dark hadrons
3. M_d = Mass of dark hadrons
4. $\boldsymbol{\lambda}$ = q - ϕ - q_d coupling strength

Link to the paper: https://arxiv.org/abs/1707.05326

Signal samples: Madgraph + Pythia8

with R_{inv} = 0.1 - 0.9 and

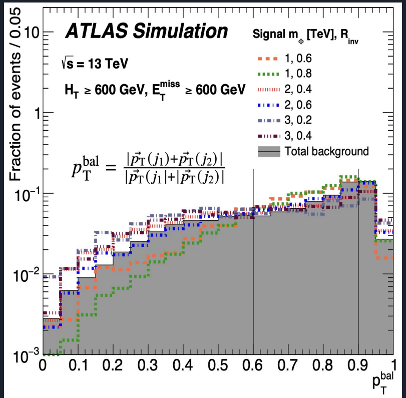
M_d = 10 GeV, M_ϕ = 1 - 5 TeV

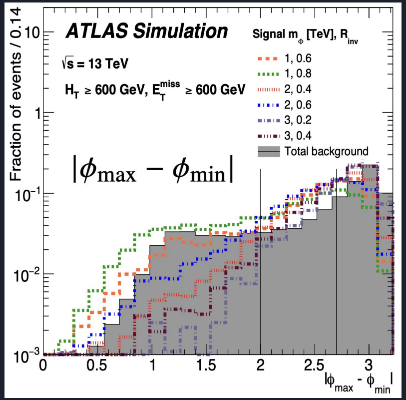


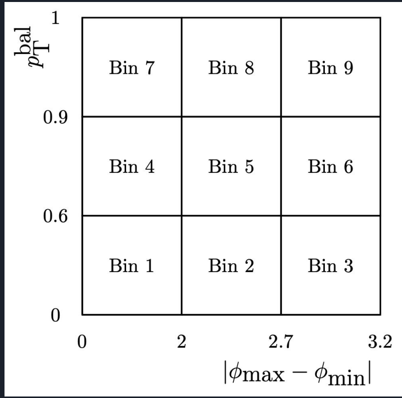
Analysis Strategy

Looking at two central R=0.4 jets, MET trigger, Δφ (closest jet, MET) < 2.0, leading jet pT > 250 GeV, HT > 600 GeV, MET > 600 GeV

Two key observables used to design a 9 bin grid: Yield in each bin treated as an observable







The signal events typically have high MET —- better sensitivity for signals with higher mediator masses and Rinv fraction if search is performed at a high MET range.

Background samples:

W/Z+jets, ttbar,

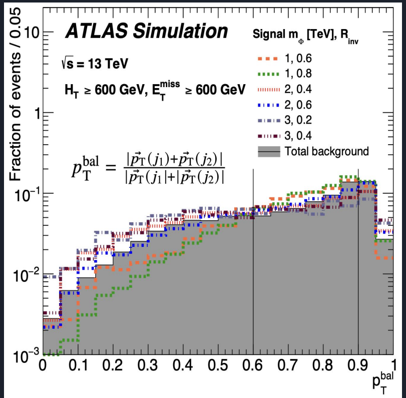
singletop, multi-jet,

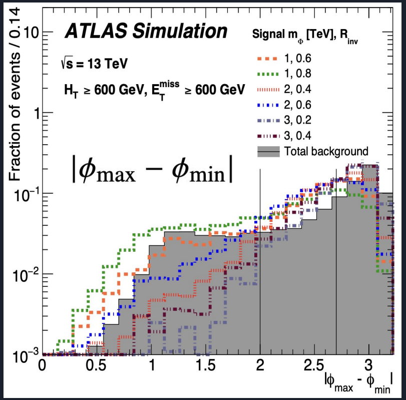
diboson

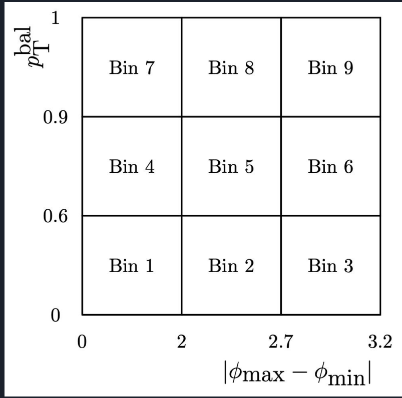
Analysis Strategy

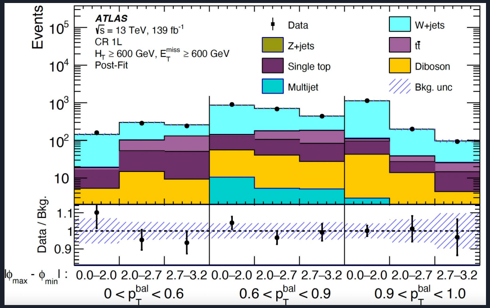
Looking at two central R=0.4 jets, MET trigger, Δφ (closest jet, MET) < 2.0, leading jet pT > 250 GeV, HT > 600 GeV, MET > 600 GeV

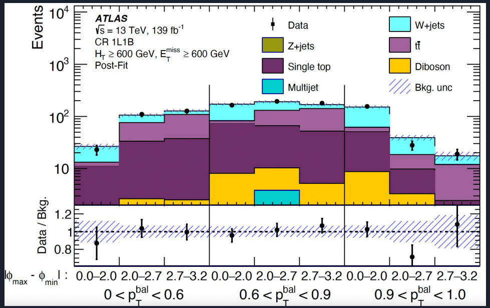
Two key observables used to design a 9 bin grid: Yield in each bin treated as an observable











The signal events typically have high MET —- better sensitivity for signals with higher mediator masses and Rinv fraction if search is performed at a high MET range.

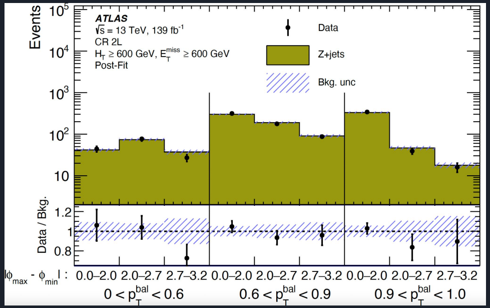
MET > 600 GeV and HT > 600 GeV – define SR & CR (1L, 1L1B & 2L)

Background samples:

W/Z+jets, ttbar,

singletop, multi-jet,

diboson

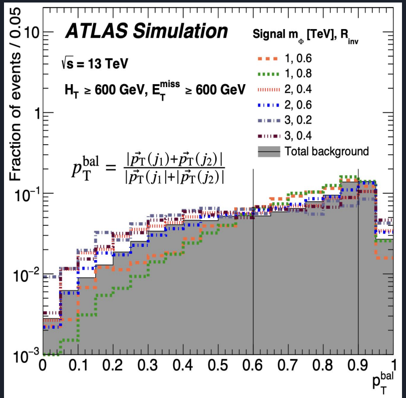


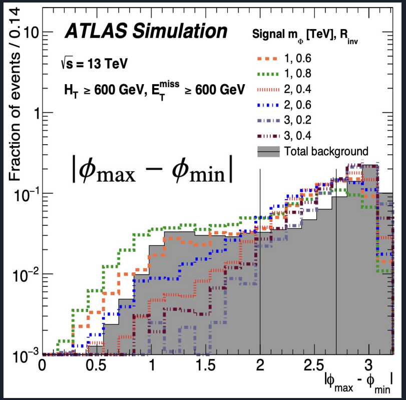
Partially data-driven method, simultaneously fit SR and three CRs to obtain scale factors for each bg process

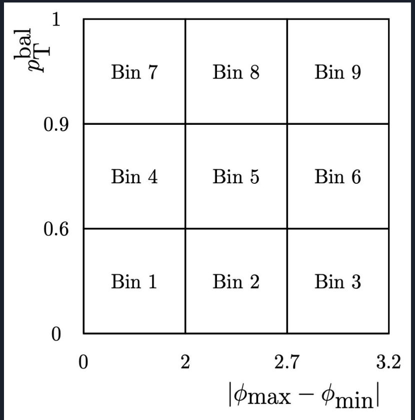
Analysis Strategy & Results

ATLAS-EXOT-2022-37 CDS

Two key observables used to design a 9 bin grid: Yield in each bin treated as an observable



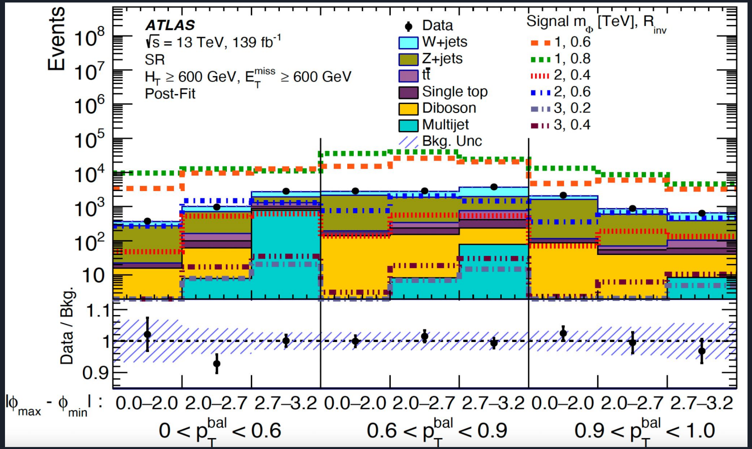




The signal events typically have high MET —- better sensitivity for signals with higher mediator masses and Rinv fraction if search is performed at a high MET range.

MET > 600 GeV and HT > 600 GeV – define SR & CR (1L, 1L1B & 2L)

Partially data-driven method, simultaneously fit SR and three CRs to obtain scale factors for each bg process — absence of signal, good postfit agreement!

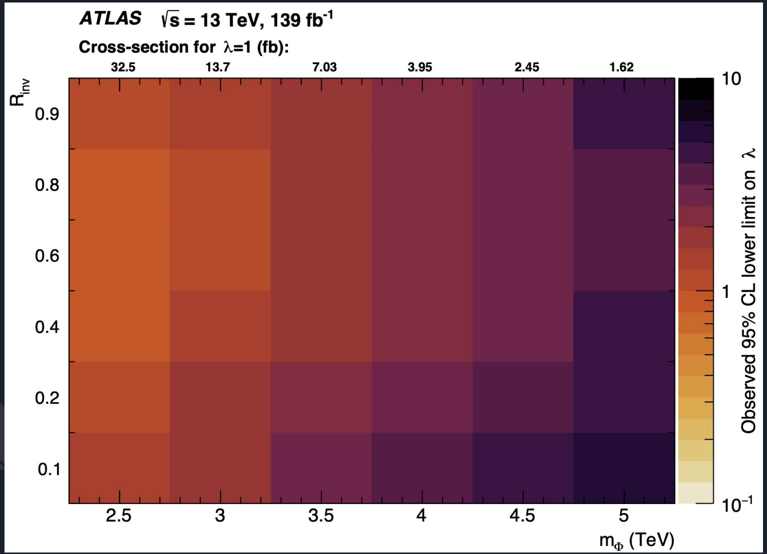


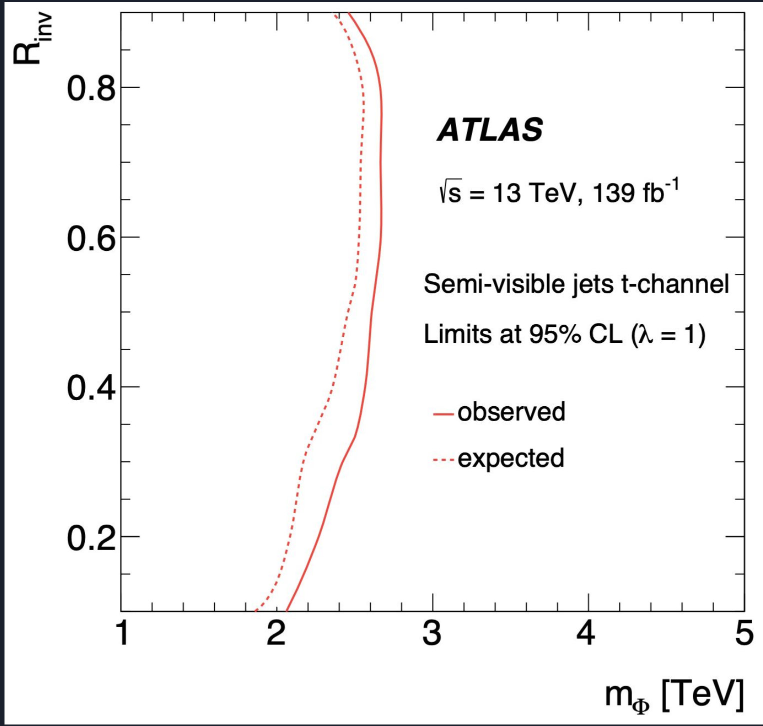
Results

First semi-visible jets result from ATLAS… but not the last!

The largest post-fit effects: signal modelling uncertainties ~8%, Z+jets modelling uncertainties ~7%, top process modelling uncertainties ~4%. The rest of the contributions are less than 2%.

- Excellent agreement between data and background prediction
- Assuming a coupling strength of unity between the mediator, a Standard Model quark and a dark quark, mediator masses up to 2.7 TeV can be excluded.
- For mediator mass of 2.5 TeV or higher can also express the limits in terms of the q-qd-ϕ vertex coupling strength λ, as XS ~ λ^4





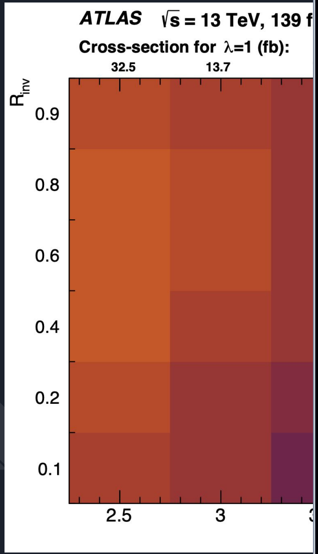
ATLAS-EXOT-2022-37 CDS

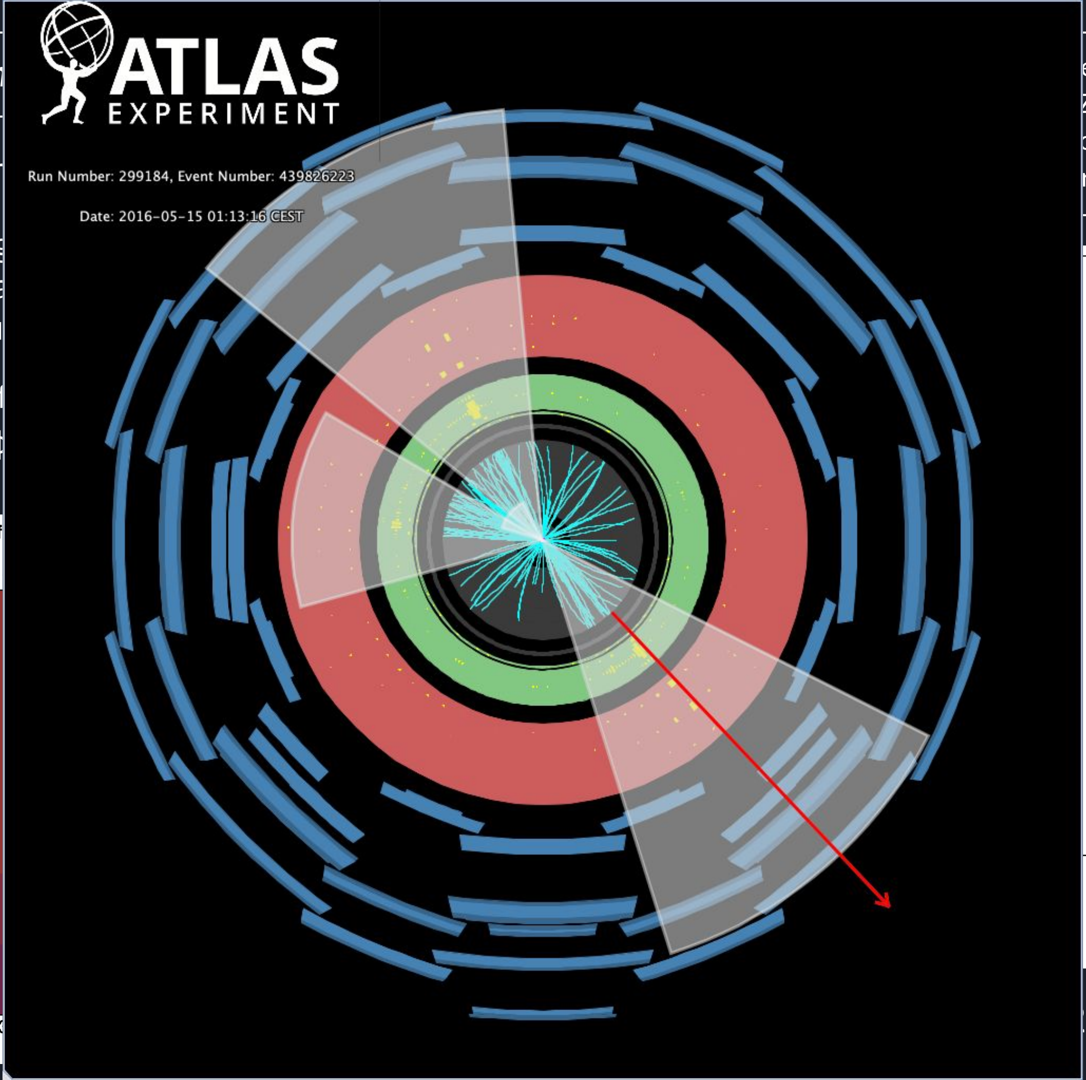
Results

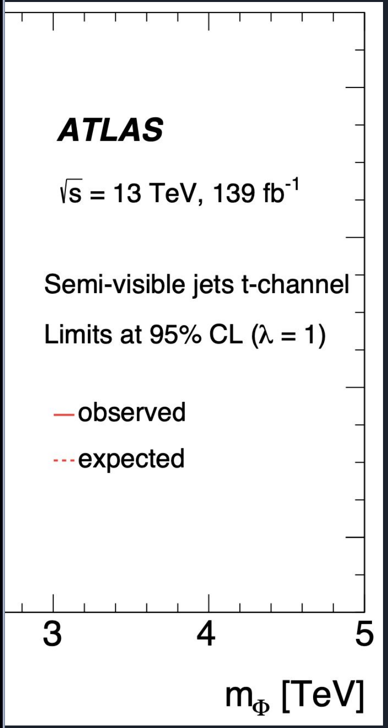
First sen

effects: signal modelling
Z+jets modelling uncertainties
odelling uncertainties ~4%.
ributions are less than 2%.

- Excellent agreemen
- Assuming a coupling Standard Model qua 2.7 TeV can be exclu
- For mediator mass of in terms of the q-qd-ϕ

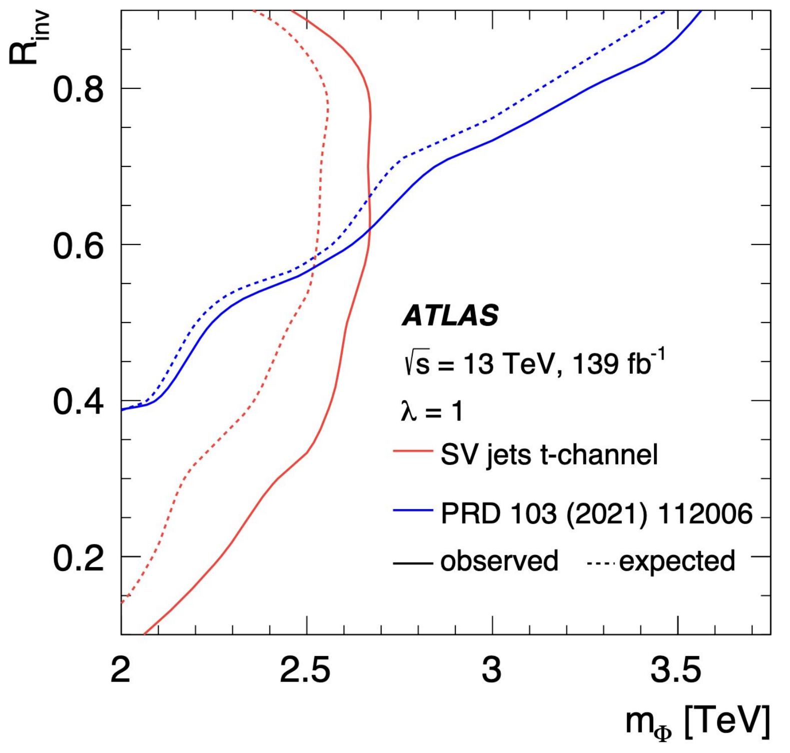






2-37 CDS

Monojet reinterpretation of semi-visible jet t-channel signals



Expected and observed exclusion contours at 95% CL for semi-visible jets signal, using the monojet analysis selection.

The analysis selections:
Number of anti-kt R=0.4 jets <= 4,
leading jet pT > 150 GeV
$\Delta\phi$ (jets, E^{T}_{miss}) > 0.4,
MET trigger, E^{T}_{miss} > 200 GeV.

The samples are then run over the inclusive E^{T}_{miss} threshold of 600 GeV in order to be consistent with semi-visible jet E^{T}_{miss} threshold.

Good complementarity of the dedicated semi-visible jet analysis with the monojet analysis

Summary

- Several avenues of strongly interacting dark sector open for exploration
- General idea evolving around the need of more signature based searches
- **First bounds set on these kind of signatures in the s/t-channel production modes from CMS/ATLAS (many more to come)**

BACKUP

Semi-visible jets in ATLAS - HV Parameters (why and what)

Parameter	value
HiddenValley:Ngauge	2
HiddenValley:FSR	on
HiddenValley:spinFv	0
HiddenValley:fragment	on
HiddenValley:pTminFSR	1.1
HiddenValley:probVector	0.75
HiddenValley:alphaOrder	1
HiddenValley:Lambda	0.1
HiddenValley:alphaFSR	1.0

All parameters set as per theory paper

Running HV alpha selected, after discussions with theorists in different platforms (Snowmass, LHC DMWG). Advised to be the safest choice for first analysis.

Semi-visible jets in ATLAS - Analysis Samples

Signal: Madgraph + Pythia8 with R_{inv} = 0.2, 0.4, 0.6, 0.8 and M_d = 10 GeV, M_ϕ = 1 - 5 TeV (in 500 GeV intervals)

Background samples:

Process	Generator	ME order	PDF	Parton shower	Tune
W/Z+jets	Sherpa 2.2.11 [25,26]	NLO (up to 2 jets)	NNPDF3.0nnlo [12]	Sherpa MEPS@NLO	Sherpa
$t\bar{t}$	Powheg Box v2 [27,28,29]	NLO	NNPDF3.0nlo [12]	Pythia 8.230 with NNPDF2.3lo	A14 [15]
Single top	Powheg Box v2	NLO	NNPDF3.0nnlo	Pythia 8.230 with NNPDF2.3lo	A14
Multijet	Pythia 8.230 [14]	LO	NNPDF2.3lo [12]	Pythia 8.230	A14
Diboson	Sherpa 2.2.1	NLO (up to 2 jets)	NNPDF3.0nnlo	Sherpa MEPS@NLO	Sherpa

Data samples:

2015: 3.20 \pm 0.07 fb^{-1}
2016: 32.9 \pm 0.72 fb^{-1}
2017: 44.3 \pm 1.06 fb^{-1}
2018: 59.9 \pm 1.19 fb^{-1}

Semi-visible jets in ATLAS - Systematic Uncertainties

- Largest contribution from theoretical components (~25% on signal cross-sections mostly from scale variations).
  - Apart from usual scale and PDF variations, also included ttbar and single top I/FSR variation, ME and PS variation by using alternate generators, DR/DS subtraction scheme difference for tW.
  - W+jets split into heavy and light flavour, and an extra 30% normalisation uncertainty was used for heavy flavour, since Sherpa 2.2 has been found to underestimate V+heavy-flavour by about a factor of 1.3
  - There is known mismodelling in multijet processes, so a data-otherMC vs multijet reweighting is done in 250 < MET < 300 GeV in 9bin distribution → the reweighting factors are obtained in bin 3,6,9, and applied to 1-3, 4-6, 7-9 respectively.
- Standard experimental uncertainties: JES/JER, MET soft term, luminosity, PU reweighting, flavour tagging, reconstruction/identification/isolation/trigger efficiencies on muon and tau leptons.

Semi-visible jets in ATLAS

Process	k_i^{SF}
Z+jets	1.18 ± 0.05
W+jets	1.09 ± 0.04
Top processes	0.64 ± 0.04
Multijet	1.10 ± 0.04

Process	SR	CR 1L	CR 1L1B	CR 2L
Z+jets	8 490 ± 260	11.6 ± 1.4	2.2 ± 0.6	1 120 ± 40
W+jets	5 820 ± 300	3 190 ± 170	351 ± 41	-
$t\bar{t}$	920 ± 70	350 ± 29	304 ± 24	-
Single top	533 ± 47	358 ± 29	290 ± 25	-
Multijet	850 ± 100	28 ± 11	7.7 ± 3.1	-
Diboson	757 ± 10	187 ± 9	34.5 ± 2.8	-
Total bkg.	17 370 ± 280	4 120 ± 100	990 ± 35	1 120 ± 40
Data	17 388	4 136	999	1 124
Signal:				
$m_\Phi = 1$ TeV, $R_{inv} = 0.6$	101 000 ± 23 000	-	-	-
$m_\Phi = 1$ TeV, $R_{inv} = 0.8$	160 000 ± 40 000	-	-	-
$m_\Phi = 2$ TeV, $R_{inv} = 0.4$	2 800 ± 600	-	-	-
$m_\Phi = 2$ TeV, $R_{inv} = 0.6$	8 900 ± 2 000	-	-	-
$m_\Phi = 3$ TeV, $R_{inv} = 0.2$	59 ± 13	-	-	-
$m_\Phi = 3$ TeV, $R_{inv} = 0.4$	126 ± 29	-	-	-

Semi-visible jets in ATLAS - Statistical analysis

- To determine individual N_i → simultaneous binned maximum likelihood function fit is performed using product of all PDF_i and nine bin yields, using the MC templates
- The fit maximises the likelihood function constructed from the product of all relevant Poisson and Gaussian pdfs. The scale factors for the individual backgrounds, k^{SF} are determined from the fit:

$$\mathcal{L}(\mu, \theta) = \prod_{j \in \text{ 36 bins}} \text{Poisson}(N_j^{\text{obs}} | \mu N_j^{\text{sig}}(\theta) + \sum_{i \in \text{bg}} k_i^{\text{SF}} \times N_{i,j}^{\text{bg}}(\theta)) \times f^{\text{constr}}(\theta)$$

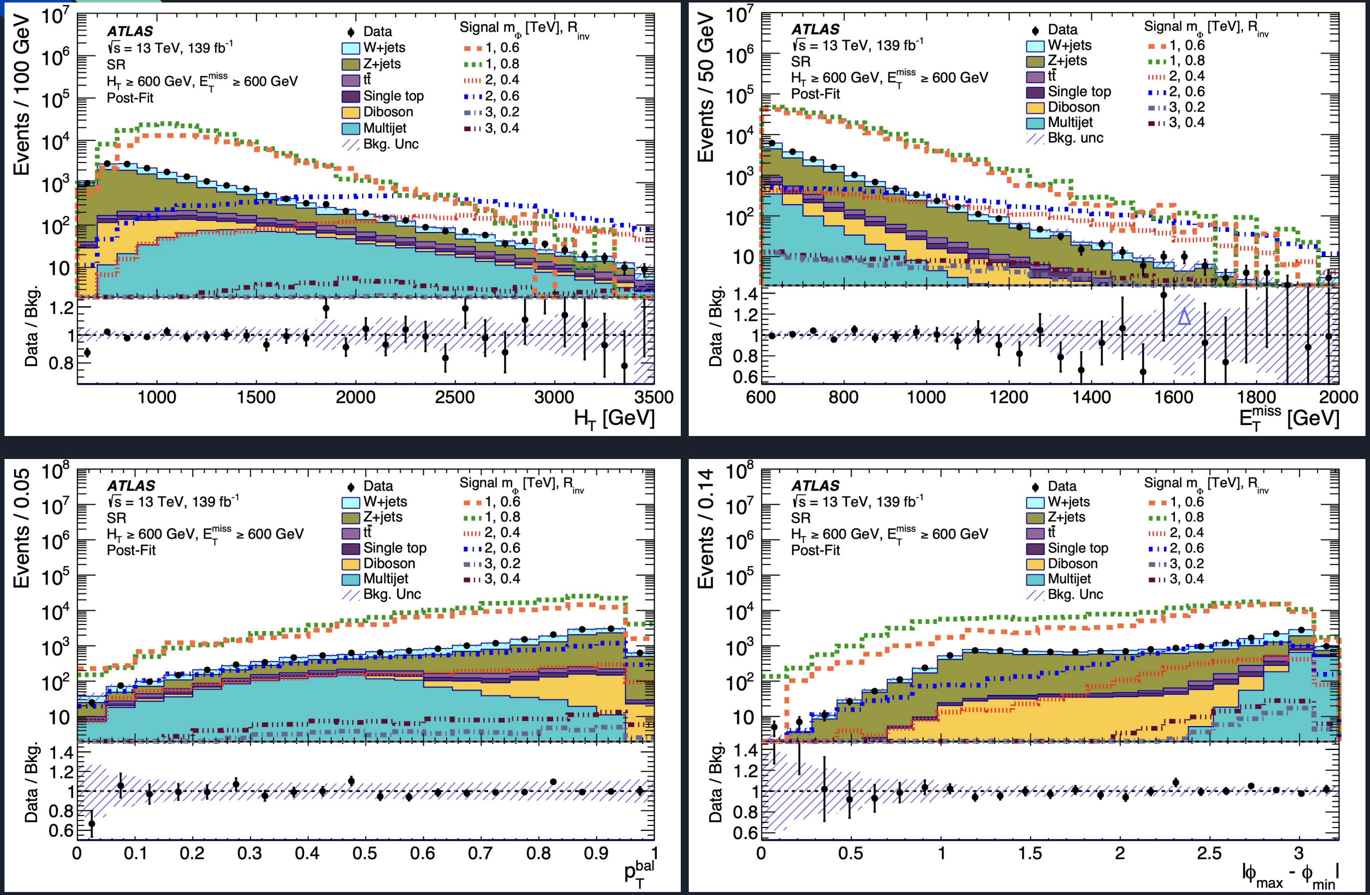
Here, $N_j^{expected}$ is the observed total yield in the bin j, signal strength is \mu, systematic uncertainties in the fit are denoted by

nuisance parameters \theta, N_j^{bg}(\theta) is the combined background yield in bin j

The term f_{constr}(\theta) of represents the product of the gaussian constraints applied to each of the nuisance parameters,

$$f_{constr}(\theta) = \prod_{k=1}^{M} G(\theta_k^0 - \theta_k)$$

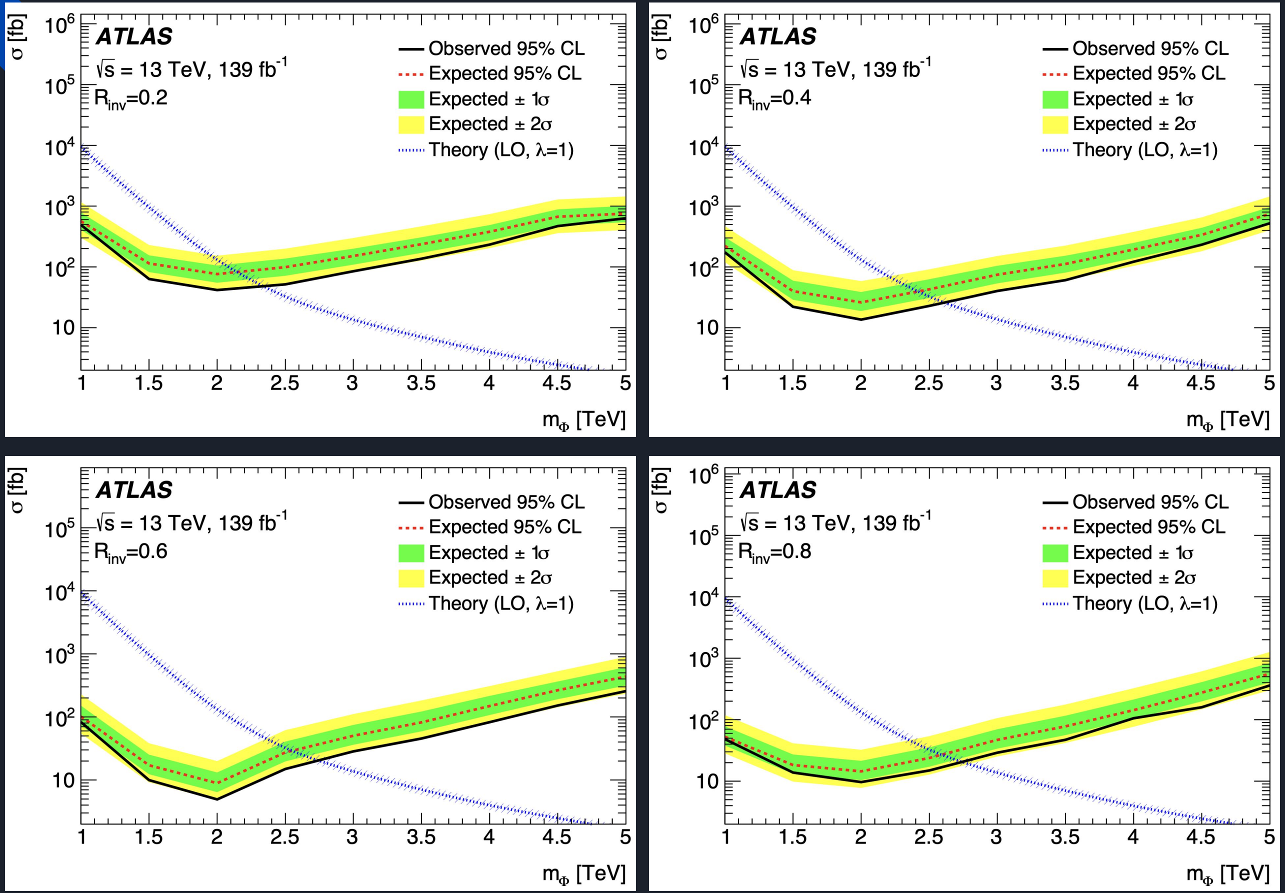
Semi-visible jets in ATLAS - Kinematic distributions - SR



We haven't found new physics :-(

Excellent agreement between data and estimated background...

Semi-visible jets in ATLAS - 1D exclusion limit plots



The observed exclusions tend to be slightly stronger than the expected ones due to a slight deficit in data in individual SR bins and the fit's preference of a negative signal yield to improve the data agreement.

Semi-visible jets in CMS

Scan	$m_{Z'}$ [TeV]	m_{dark} [GeV]	r_{inv}	α_{dark}
1	1.5–5.1	1–100	0.3	$\alpha_{\text{dark}}^{\text{peak}}$
2	1.5–5.1	20	0–1	$\alpha_{\text{dark}}^{\text{peak}}$
3	1.5–5.1	20	0.3	$\alpha_{\text{dark}}^{\text{low}}$–$\alpha_{\text{dark}}^{\text{high}}$

Preselection requirements

$p_T(J_{1,2}) > 200\,\text{GeV},\ \eta(J_{1,2}) < 2.4$

$R_T > 0.15$

$\Delta\eta(J_1, J_2) < 1.5$

$m_T > 1.5\,\text{TeV}$

$N_\mu = 0$

$N_e = 0$

p_T^{miss} filters

$\Delta R(j_{1,2}, c_{\text{nonfunctional}}) > 0.1$

Final selection requirements

veto $f_\gamma(j_1) > 0.7$ & $p_T(j_1) > 1.0\,\text{TeV}$

veto $-3.05 < \eta_j < -1.35$ & $-1.62 < \phi_j < -0.82$ *

$\Delta\phi_{\min} < 0.8$

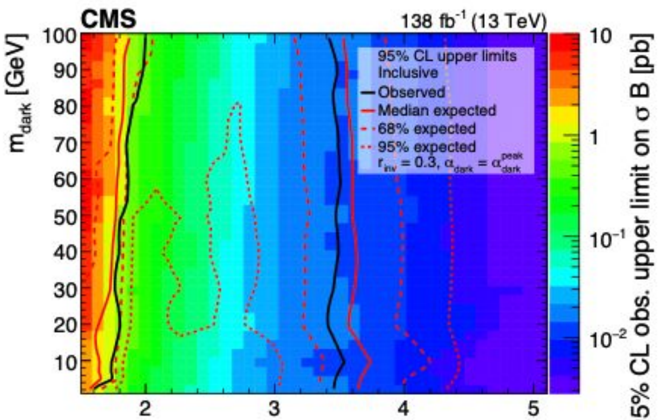
The 15 BDT input variables, computed for each jet, originate from several categories. From heavy object identification, the N-subjettiness ratios τ_{21} and τ_{32} [81], the energy correlation functions $N_2^{(1)}$ and $N_3^{(1)}$ [82], and the soft-drop mass m_{SD} [83] are used. From quark-gluon discrimination, the jet girth g_{jet} [84], the major and minor jet axes σ_{major} and σ_{minor} [85], and the p_T dispersion D_{p_T} [85] are used. The flavor-related variables used are the jet energy fractions for each type of constituent identified by the PF algorithm: $f_{h^\pm}$, f_e, f_μ, f_{h^0}, and f_γ. The angle between the jet and the missing transverse momentum, $\Delta\phi(\vec{J}, \vec{p}_T^{\text{miss}})$, is also included.

Semi-visible jets in CMS

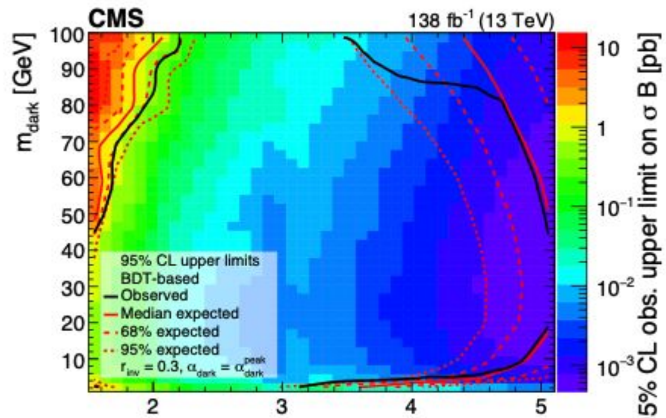
Table 3: Metrics representing the performance of the BDT for the benchmark signal model ($m_{Z'} = 3.1\,\text{TeV}$, $m_{\text{dark}} = 20\,\text{GeV}$, $r_{\text{inv}} = 0.3$, $\alpha_{\text{dark}} = \alpha_{\text{dark}}^{\text{peak}}$), compared to each of the major SM background processes.

Background	Acc.	AUC	$1/\varepsilon_{\text{bkg}}$ ($\varepsilon_{\text{sig}} = 0.3$)
QCD	0.883	0.946	636.8
$t\bar{t}$	0.883	0.932	290.0
W+jets	0.883	0.936	477.1
Z+jets	0.883	0.930	455.4

Uncertainty	Yield effect [%]
Integrated luminosity	1.6
Jet energy corrections	0.05–12
Jet energy resolution	0.02–2.3
Jet energy scale	0.29–21
PDF	0.0–5.3
Parton shower FSR	0.07–9.4
Parton shower ISR	0.0–2.9
Pileup reweighting	0.0–1.3
Renormalization and factorization scales	0.0–0.12
Statistical	1.2–4.9
Trigger control region	0.24–0.40
Trigger efficiency	2.0
Total	3.3–22



Inclusive signal regions



BDT - based regions