Paper link: arXiv:2403.01556

Search for emerging jets using CMS Run 2 data

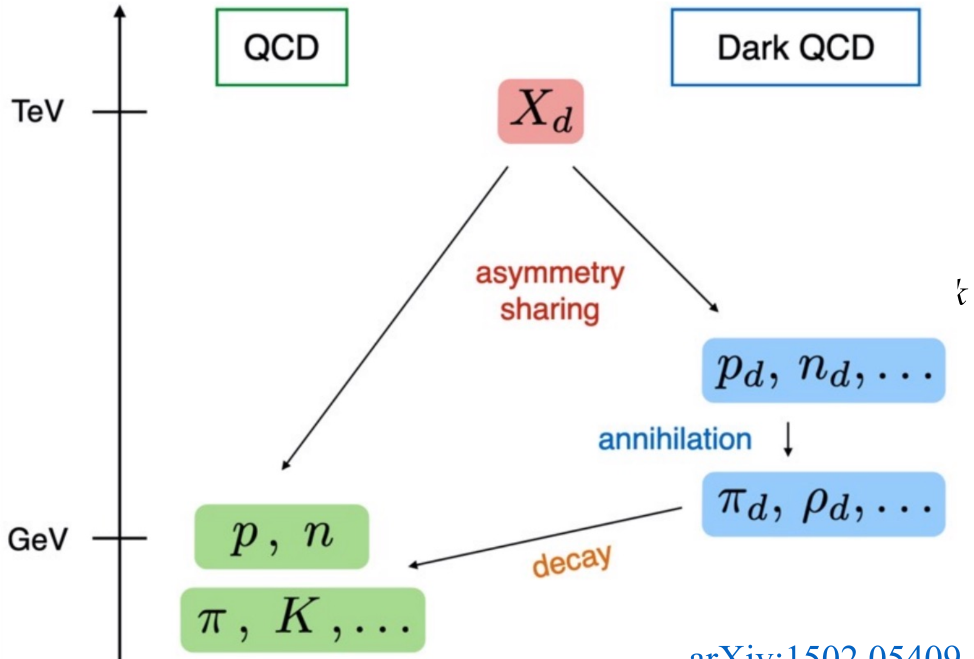
Claire Savard on behalf of the CMS collaboration
University of Colorado, Boulder

13 May 2024

Emerging jets (EMJ) theory

- Dark matter model = QCD-like hidden sector
- Dark hadrons with ~ Λ_{dark} (GeV), dark pions unstable $m_{\pi_{dark}} < \Lambda_{dark}$
- Heavy mediator particle ~ TeV couples to dark and visible sectors

- Energy scales reachable at LHC



arXiv:1502.05409

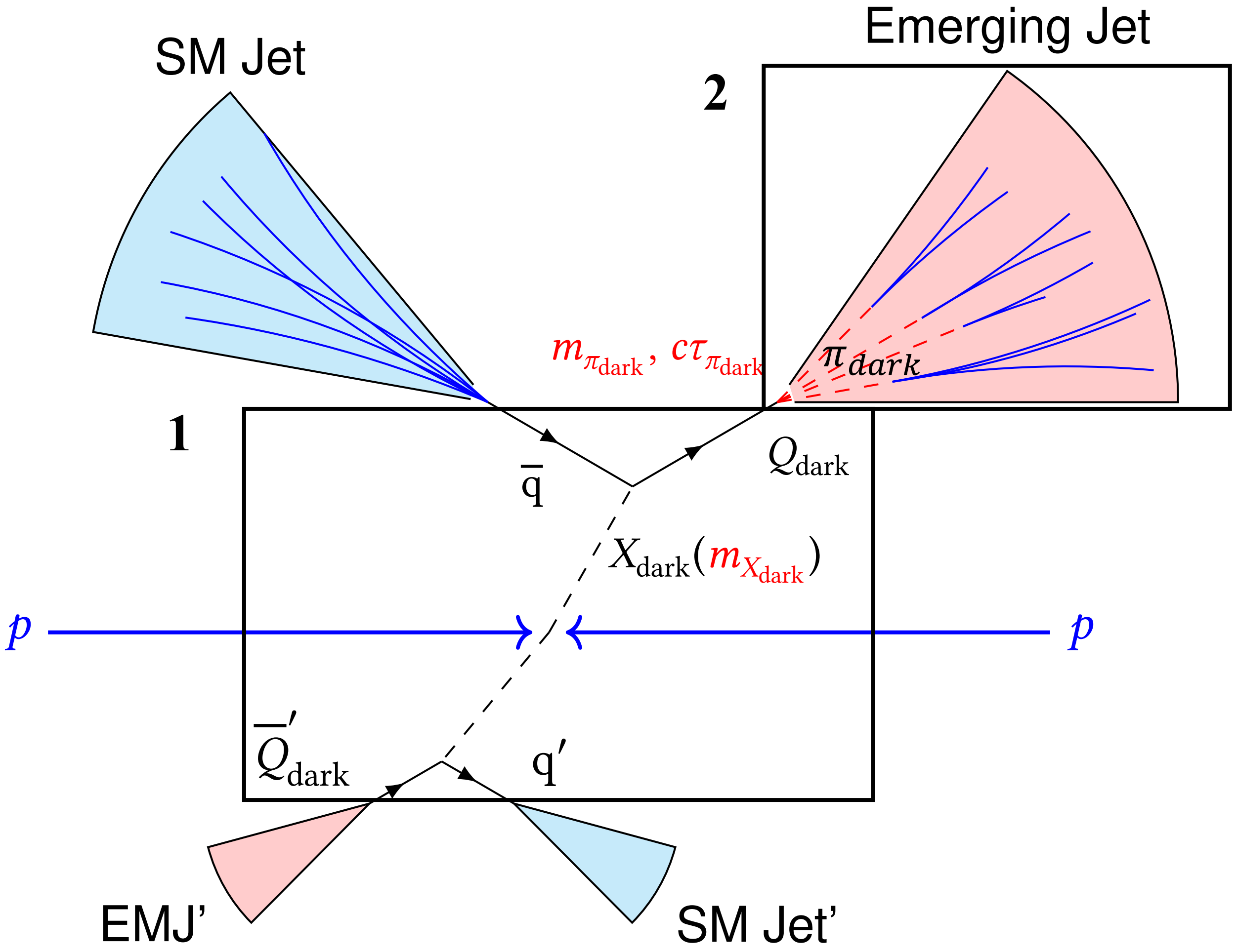
EMJ production at the LHC

1. $pp \rightarrow 2X_{dark} \rightarrow 2(qQ_{dark})$
2. $Q_{dark} \xrightarrow{\text{hadronizes}} N\,\pi_{dark}$ &

   $\pi_{dark} \xrightarrow{\text{travel } c\tau}$ SM particles

Free parameters:

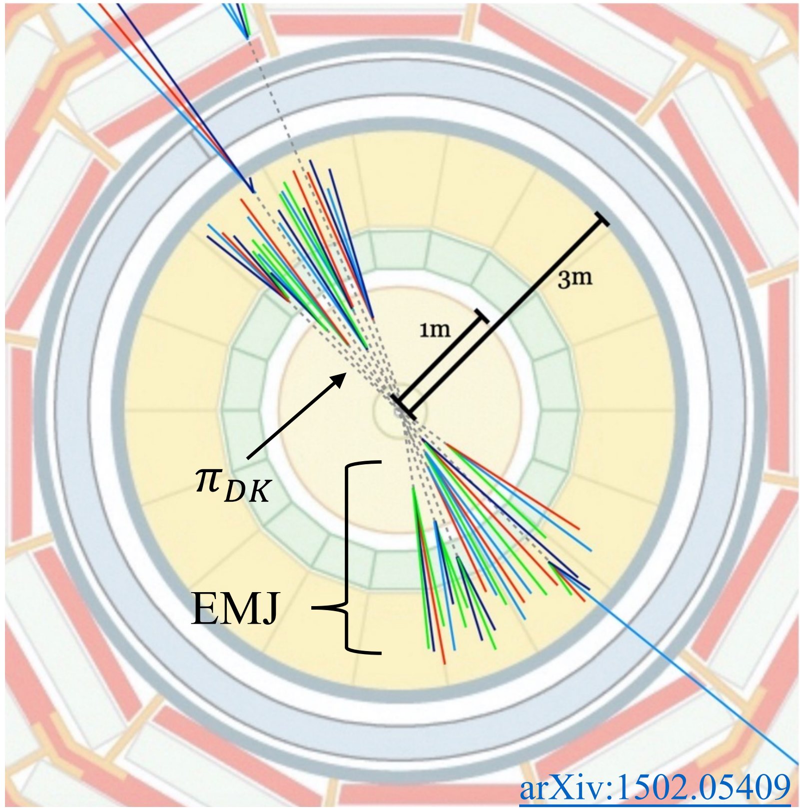
- m_X: [1, 2.5] TeV
- $m_{\pi_{dark}}$: [6, 20] GeV
- $c\tau_{\pi_{dark}}$: [1, 1000] mm



EMJs in CMS detector

- General-purpose particle detector
- Silicon tracker:
  - Charged particle reconstruction
  - Within 3.8 T solenoid for momentum resolution
  - Vertex z_0 resolution of ~15 μm
  - Extends from collision point to 1 m
- $c\tau_{\pi_{dark}}$ 1 – 1000 mm, contained in tracker

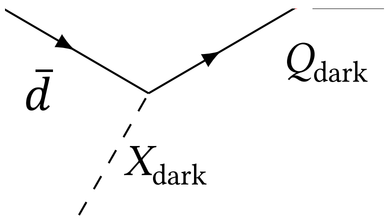
CMS cross section



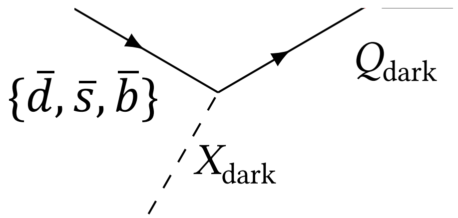
arXiv:1502.05409

EMJ coupling scenarios

1. Unflavored down scenario
    - Dark quarks couple to down quarks ONLY
    - All π_{dark} have same $c\tau$
    - Previous CMS search ([arXiv:1810.10069](https://arxiv.org/abs/1810.10069))



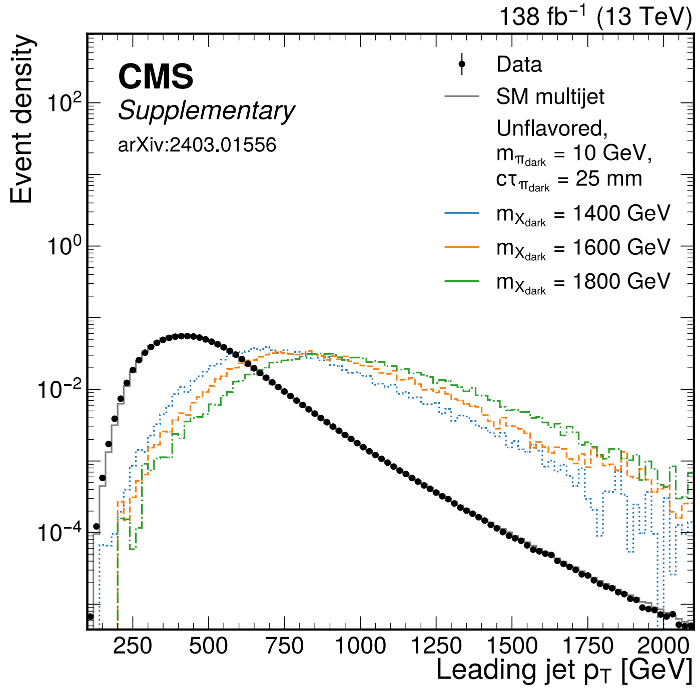
2. Flavor-aligned down scenario
    - Dark quarks couple to down-type quarks ONLY (d, s, b)
    - π_{dark} lifetime depends on dark pion composition

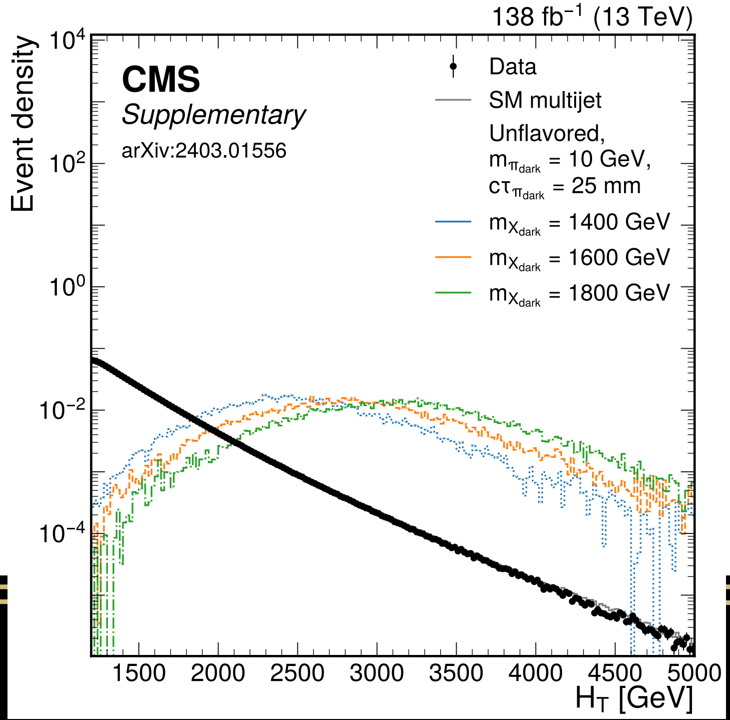


- Scenarios phenomenologically different, search methods tuned separately

Event selection

- ≥ 4 high p_T jets
- High event H_T $(= \sum_{jets} p_T)$
- ≥ 2 EMJ-tagged jets

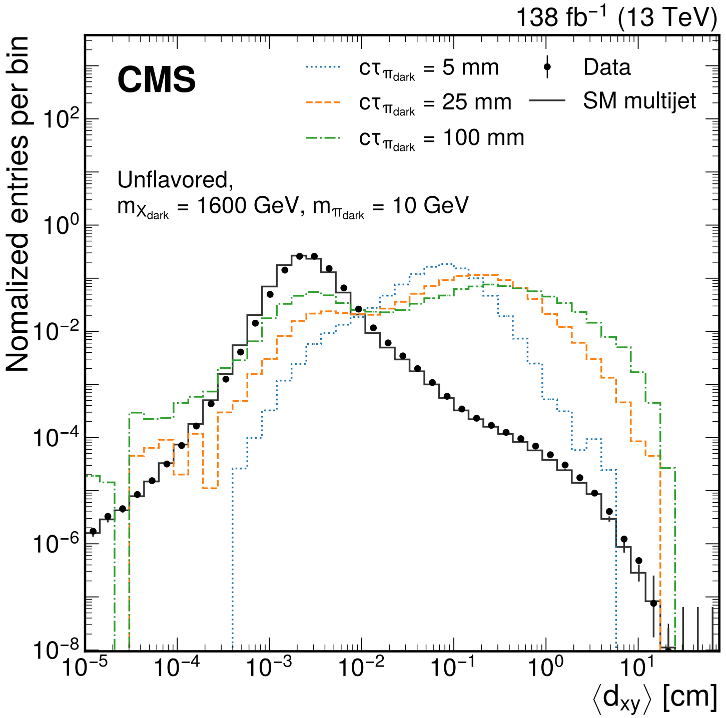


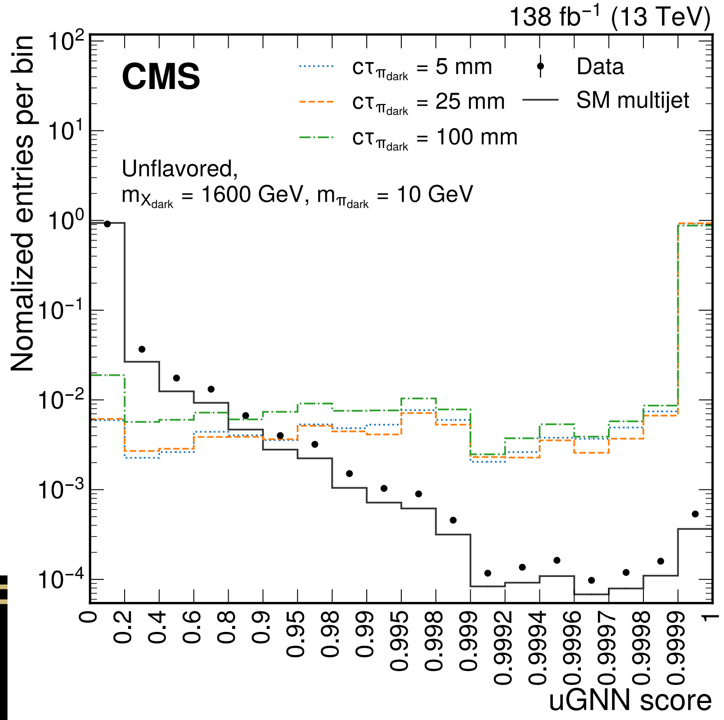


Jet tagging

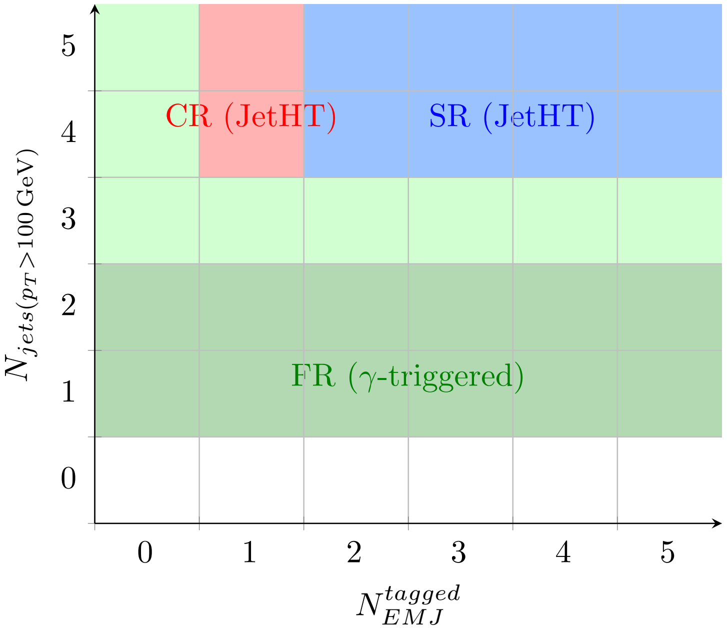
1. Jet variable selections (cut-based)
   - Unflavored: leverage track displacement
   - Flavor-aligned: leverage track multiplicity

2. Graph neural network classifier
   - 2 models trained separately on unflavored and flavor-aligned scenarios





Background estimation



Estimate # of bkg. events pass into SR using CR events and mistag rates from FR

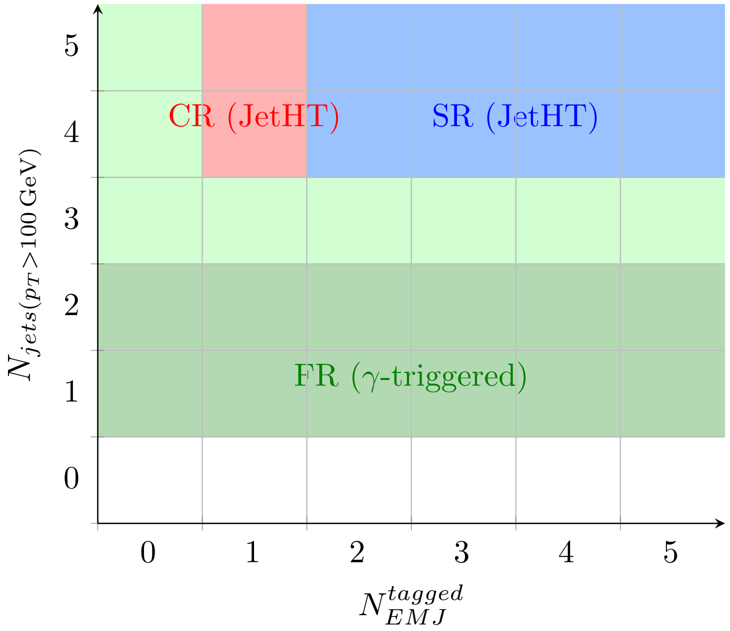
FR

$$N_{\mathrm{SR}} = \sum_{evt\in\mathrm{CR}} SF \sim \sum_{evt\in\mathrm{CR}} \frac{1}{2} \sum_{j\notin tagged} \epsilon(p_{T,j})$$

CR

- Fully data-driven estimation
- Mistag rate (ϵ) binned along jet p_T

Background estimation



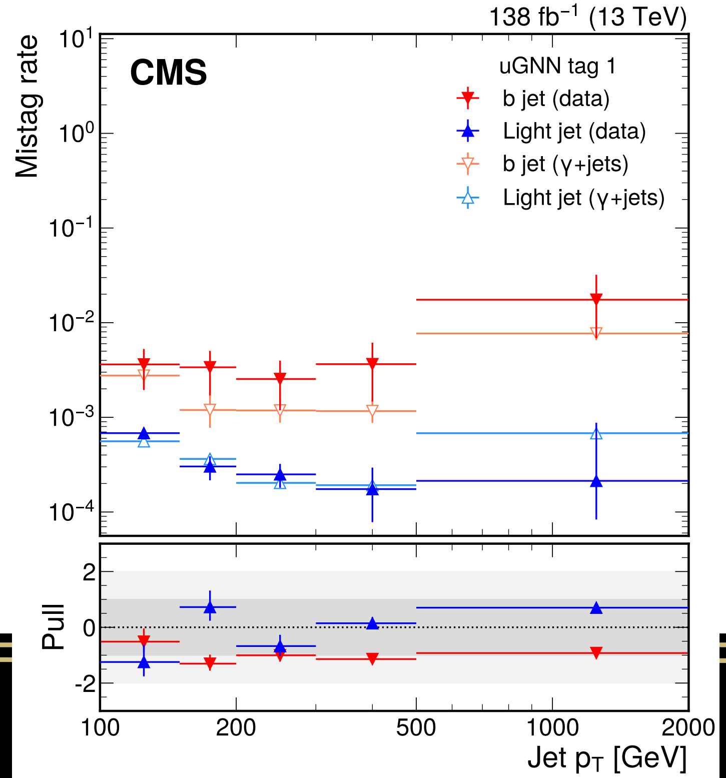
Estimate # of bkg. events pass into SR using CR events and mistag rates from FR

FR

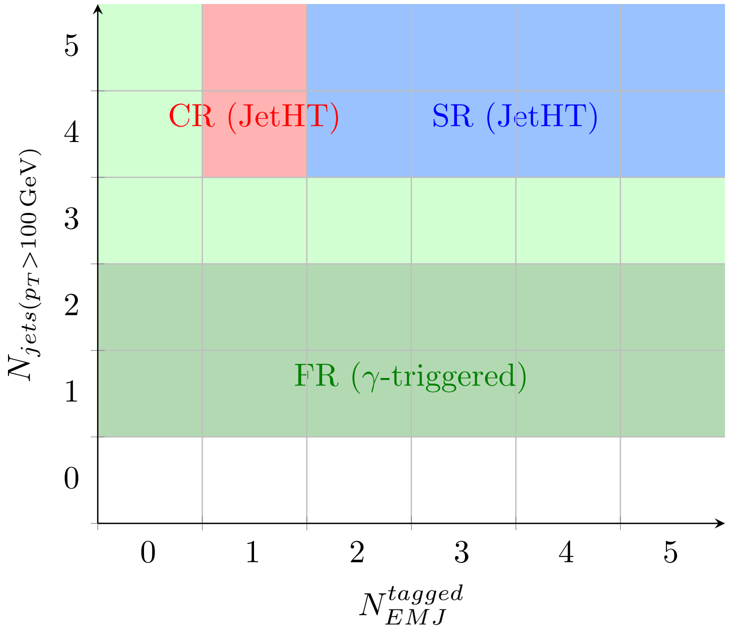
$$N_{\mathrm{SR}} = \sum_{evt\in\mathrm{CR}} SF \sim \sum_{evt\in\mathrm{CR}} \frac{1}{2} \sum_{j\notin tagged} \epsilon(p_{T,j})$$

CR

- Fully data-driven estimation
- Mistag rate (ϵ) binned along jet p_T



Background estimation



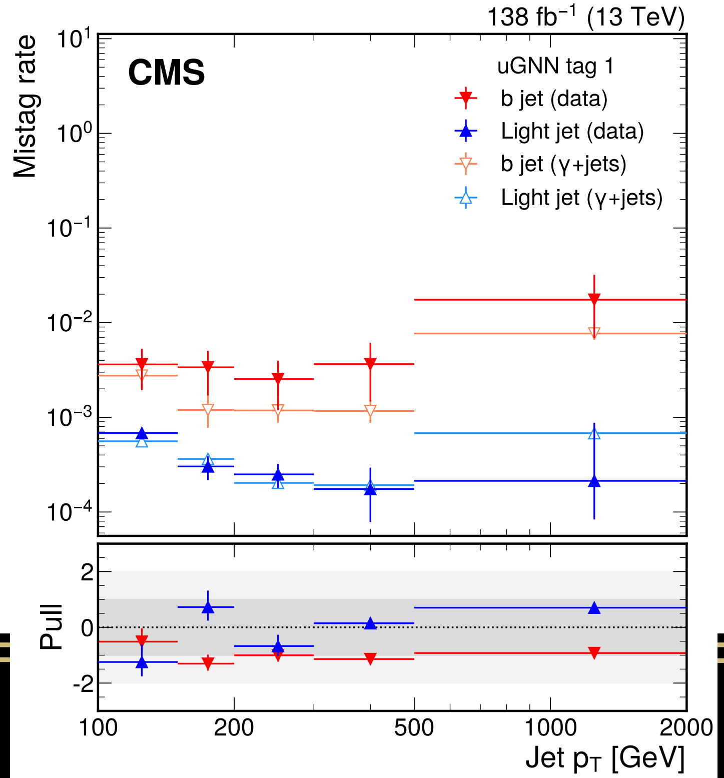
Estimate # of bkg. events pass into SR using CR events and mistag rates from FR

FR

$$N_{\mathrm{SR}} \sim \sum_{evt\in\mathrm{CR}} \frac{1}{2} \sum_{j\notin\mathrm{tagged}} B^{\mathrm{CR}}\epsilon(b,p_{T,j}) + (1-B^{\mathrm{CR}})\epsilon(l,p_{T,j})$$

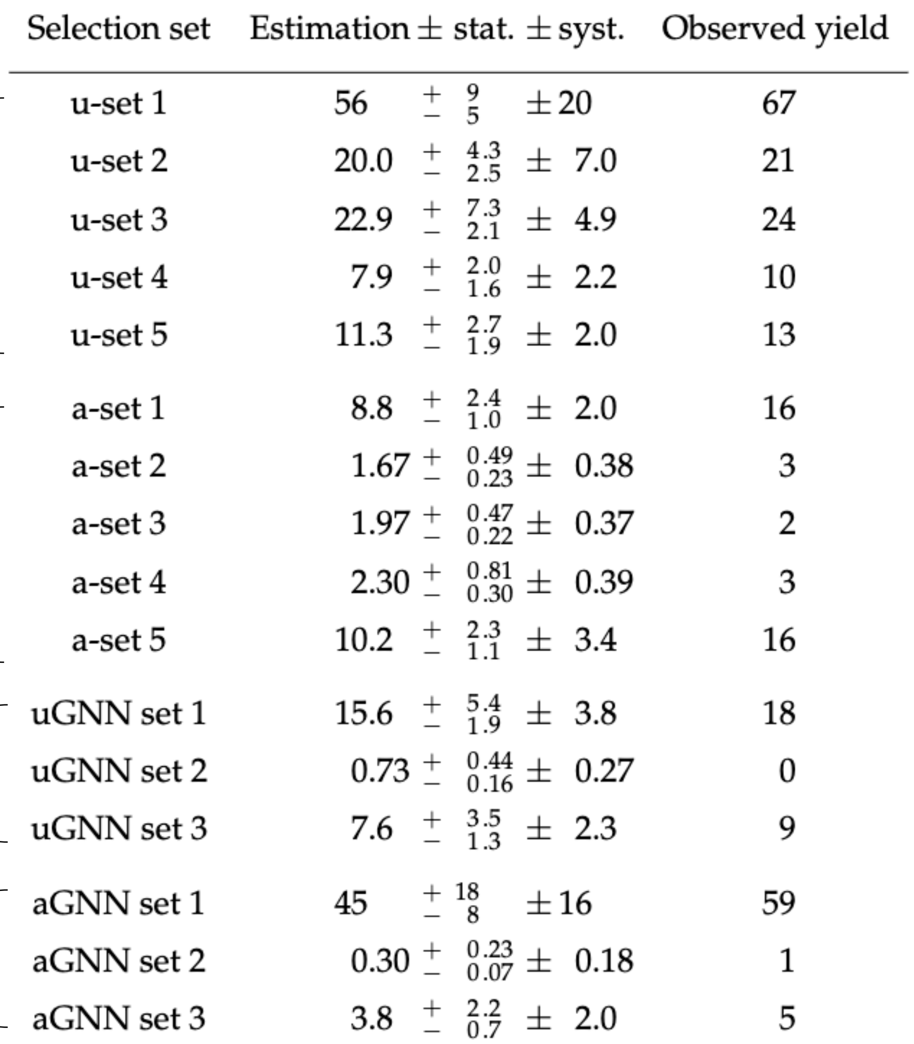
CR

- Fully data-driven estimation
- Mistag rate (ϵ) binned along jet p_T
- B-jet discriminator to calculate mistag on bs separately and b-jet fraction (B^{CR})



Results

- Different event selection criteria for various EMJ free parameters $(m_{X_{dark}}, m_{\pi_{dark}}, c\tau_{\pi_{dark}})$
- Systematic uncertainties on ϵ parameterization, jet flavor estimation, CR/FR selection



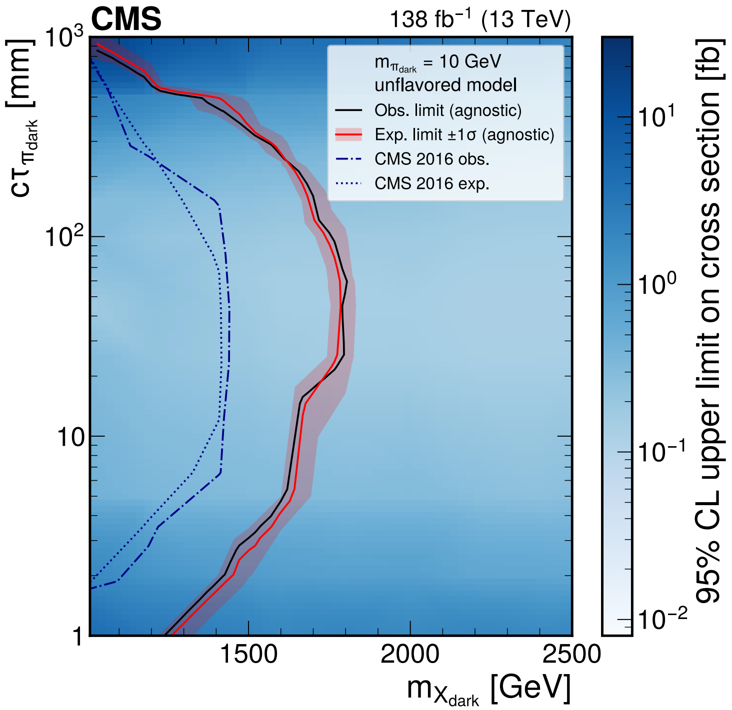
	Selection set	Estimation ± stat. ± syst.	Observed yield
Cut-based unflavored	u-set 1	$56 \, ^{+9}_{-5} \pm 20$	67
	u-set 2	$20.0 \, ^{+4.3}_{-2.5} \pm 7.0$	21
	u-set 3	$22.9 \, ^{+7.3}_{-2.1} \pm 4.9$	24
	u-set 4	$7.9 \, ^{+2.0}_{-1.6} \pm 2.2$	10
	u-set 5	$11.3 \, ^{+2.7}_{-1.9} \pm 2.0$	13
Cut-based flavored	a-set 1	$8.8 \, ^{+2.4}_{-1.0} \pm 2.0$	16
	a-set 2	$1.67 \, ^{+0.49}_{-0.23} \pm 0.38$	3
	a-set 3	$1.97 \, ^{+0.47}_{-0.22} \pm 0.37$	2
	a-set 4	$2.30 \, ^{+0.81}_{-0.30} \pm 0.39$	3
	a-set 5	$10.2 \, ^{+2.3}_{-1.1} \pm 3.4$	16
GNN unflavored	uGNN set 1	$15.6 \, ^{+5.4}_{-1.9} \pm 3.8$	18
	uGNN set 2	$0.73 \, ^{+0.44}_{-0.16} \pm 0.27$	0
	uGNN set 3	$7.6 \, ^{+3.5}_{-1.3} \pm 2.3$	9
GNN flavored	aGNN set 1	$45 \, ^{+18}_{-8} \pm 16$	59
	aGNN set 2	$0.30 \, ^{+0.23}_{-0.07} \pm 0.18$	1
	aGNN set 3	$3.8 \, ^{+2.2}_{-0.7} \pm 2.0$	5

No statistically significant excess between estimated and observed # of events

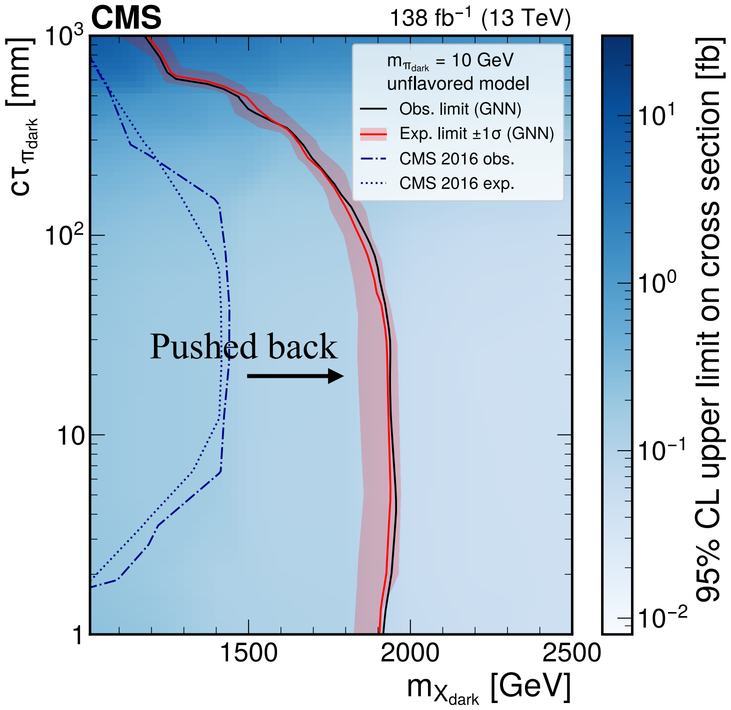
Exclusion limit results – unflavored

- Large gain in sensitivity using GNN at lower $c\tau_{\pi_{dark}}$
- Limits pushed back by ~ 400 GeV compared to previous publication

Cut-based jet tagging



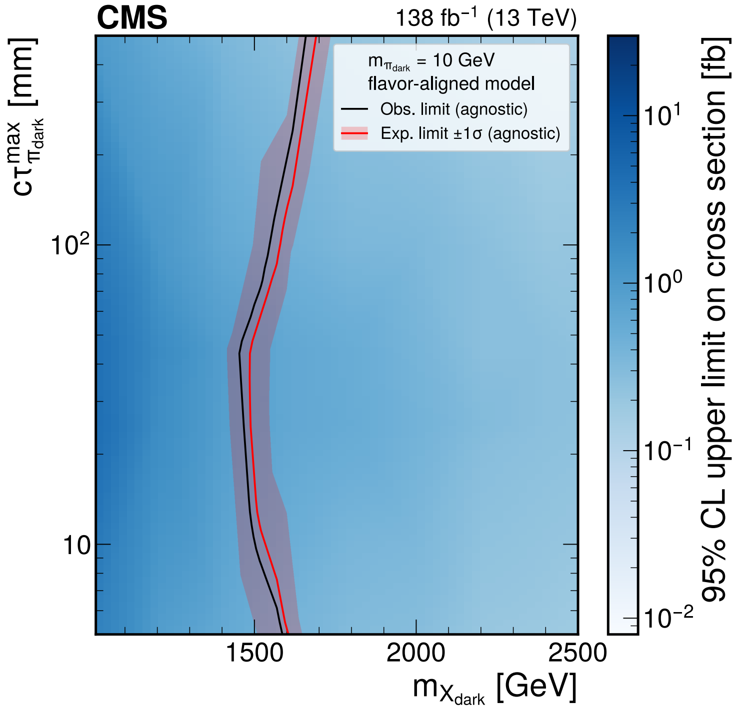
GNN jet tagging



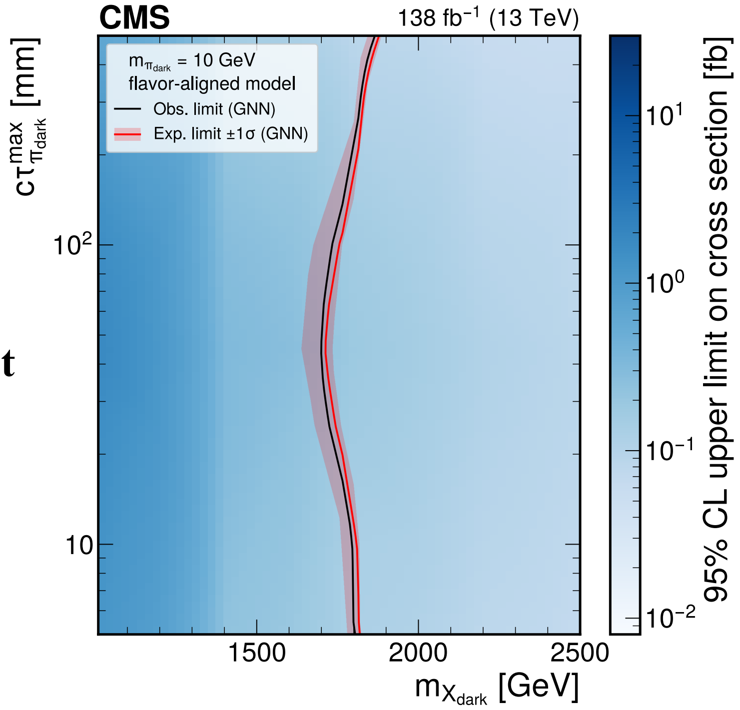
Exclusion limit results – flavor-aligned

- Completely new limits, GNN has best sensitivity
- Exclusion of up to ~1950 GeV in $m_{X_{dark}}$

Cut-based jet tagging



GNN jet tagging



Conclusions

- Pushed back previous limits on unflavored $m_{\pi_{dark}} = 10$ GeV models
- Completely new limits for:
  - Unflavored $m_{\pi_{dark}} = 20$ GeV
  - All flavor-aligned models

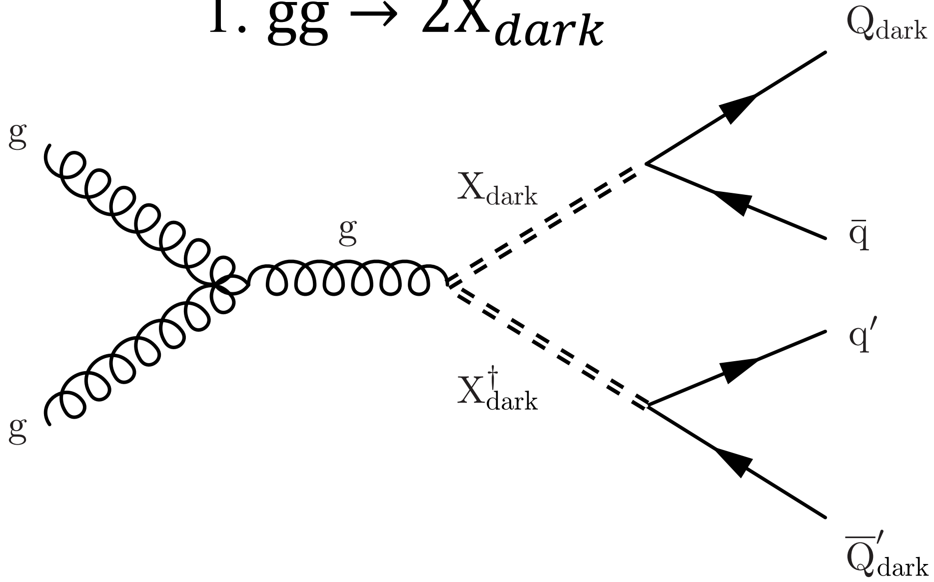
- One of first analyses to use GNN tagger
- GNN has better limits than cut-based algorithm by 150 – 600 GeV in $m_{X_{dark}}$ everywhere

Backup

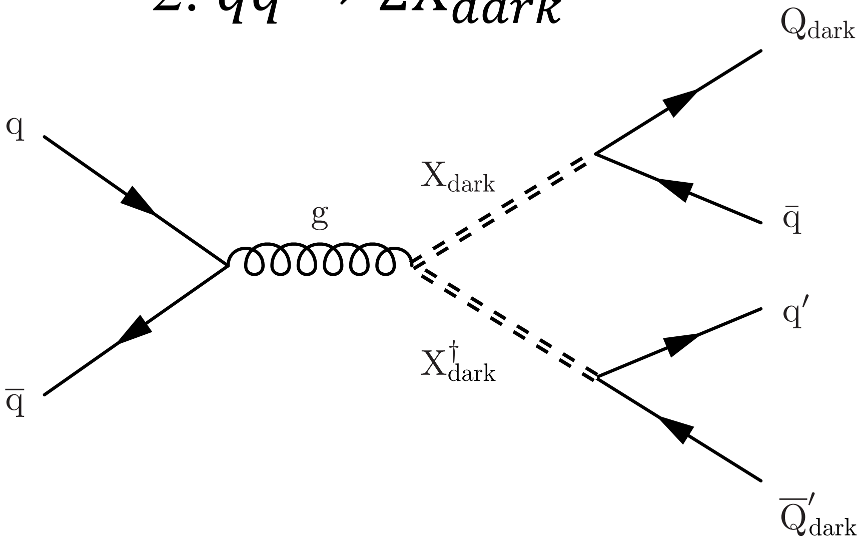
EMJ production at LHC

- 2 methods of X_{dark} production:

1. $gg \rightarrow 2X_{dark}$



2. $qq \rightarrow 2X_{dark}$

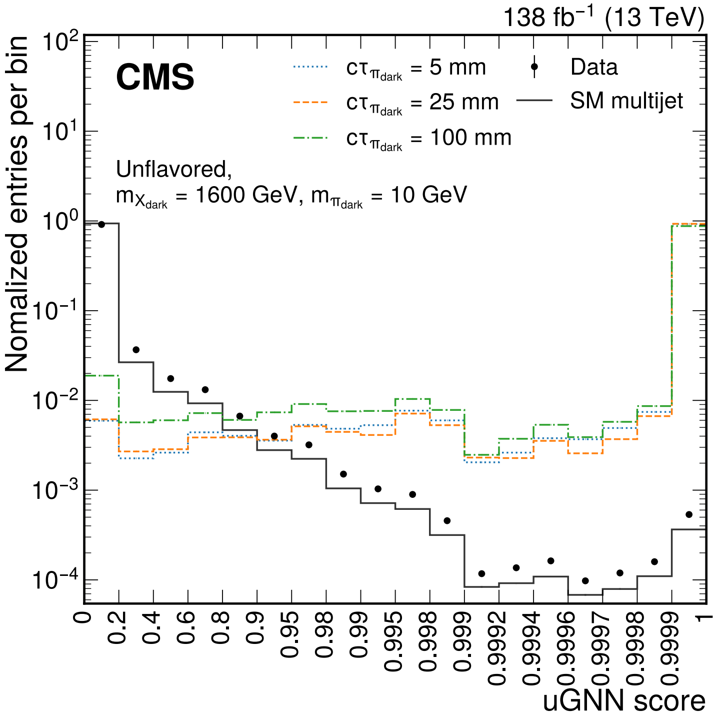


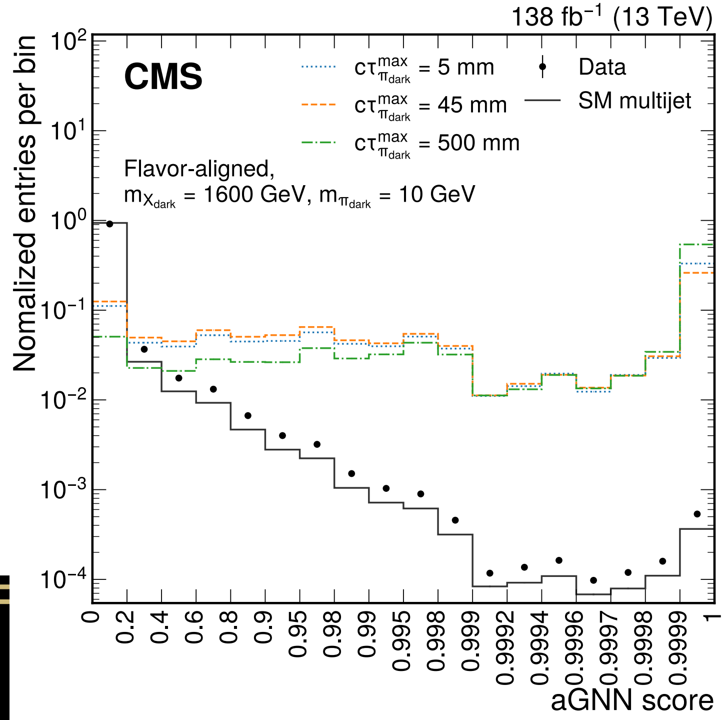
- Pairs of X_{dark} from gluon (g), decay to visible (q) and dark quarks ($\bar{Q}_{dark}$)

GNN EMJ tagger

- Associate tracks to jets within 0.8 angular separation from jet axis
- Track coordinates with respect to jet axis ($\Delta\phi$, $\Delta\eta$)
- Track features within jet:
    - Angular separation ΔR
    - p_T and p_T fraction
    - Impact parameters d_{xy} and d_z

- GNN score output used to classify
- 2 GNNs: unflavored and aligned





Mistag rate calculations

2 equations, 2 unknowns: solve for $\epsilon(b/l, p_T)$
(inverse method)

Mistag rate equations:

$$\begin{aligned} FR_E: &\quad \epsilon^E(p_T) = B^E(p_T)\,\epsilon(b,p_T) + (1-B^E)\,\epsilon(l,p_T) \\ FR_S: &\quad \epsilon^S(p_T) = B^S(p_T)\,\epsilon(b,p_T) + (1-B^S)\,\epsilon(l,p_T) \end{aligned}$$

b-enhanced and b-suppressed region defined using b-jet discriminator on a jet in FR, can measure mistag rate directly

Calculate b-jet fraction using DeepJet discriminator template

Mistag rate scale factor (SF)

Use flavor averaging* to get final background estimation:

$$N_{SR} \sim \sum_{evt \in CR} \frac{1}{2} \sum_{j \notin tagged} \epsilon\left(f_j, p_{T,j}\right)$$

$$N_{SR} \sim \sum_{evt \in CR} \frac{1}{2} \sum_{j \notin tagged} B^{CR} \epsilon\left(b, p_{T,j}\right) + (1 - B^{CR}) \epsilon(l, p_{T,j})$$

*Same method implemented in arXiv:1810.10069

Background uncertainties

Most accurate background estimation

$$Est_{true}^{JetHT}\left(\epsilon_{true}^{JetHT}(\vec{\theta}_{\infty})\right)$$

Cannot evaluate ϵ in infinitely fine jet kinematics bins

$$\downarrow$$

$$Est_{true}^{JetHT}\left(\epsilon_{true}^{JetHT}(p_T)\right)$$

Cannot evaluate ϵ in SR, potential signal contamination

$$\downarrow$$

$$Est_{true}^{JetHT}\left(\epsilon_{true}^{\gamma+jets}(p_T)\right)$$

Cannot determine flavor directly in data

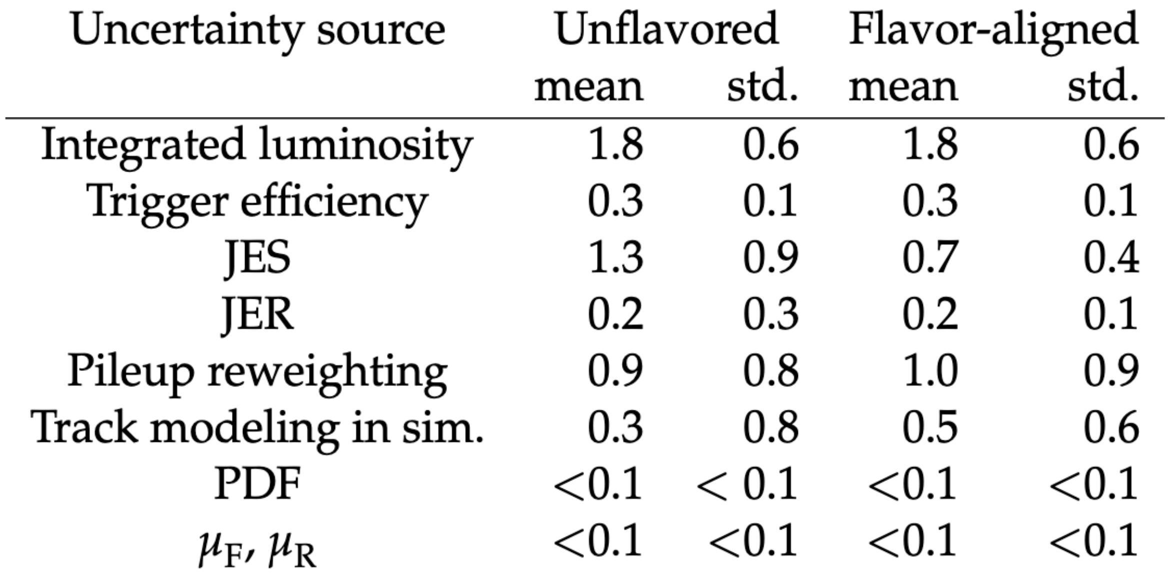
$$\downarrow$$

$$Est_{avg}^{JetHT}\left(\epsilon_{inv}^{\gamma+jets}(p_T)\right)$$

What we CAN evaluate

Each change leads to uncertainty in final estimation

Signal uncertainties



Uncertainty source	Unflavored		Flavor-aligned	
	mean	std.	mean	std.
Integrated luminosity	1.8	0.6	1.8	0.6
Trigger efficiency	0.3	0.1	0.3	0.1
JES	1.3	0.9	0.7	0.4
JER	0.2	0.3	0.2	0.1
Pileup reweighting	0.9	0.8	1.0	0.9
Track modeling in sim.	0.3	0.8	0.5	0.6
PDF	<0.1	< 0.1	<0.1	<0.1
μ_F, μ_R	<0.1	<0.1	<0.1	<0.1

Evaluated per EMJ signal sample