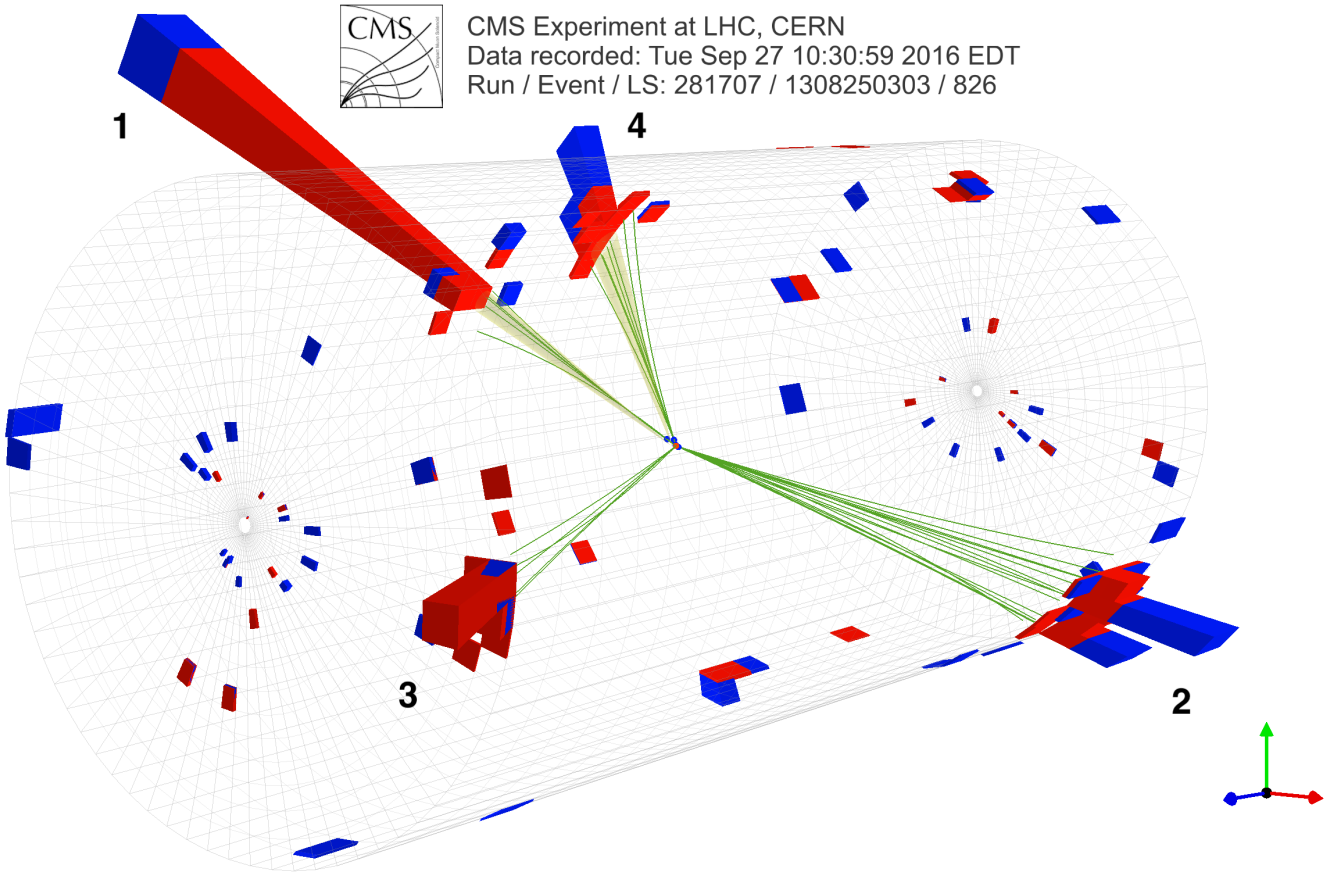


Emerging Jets with CMS in Run 2

Jannicke Pearkes

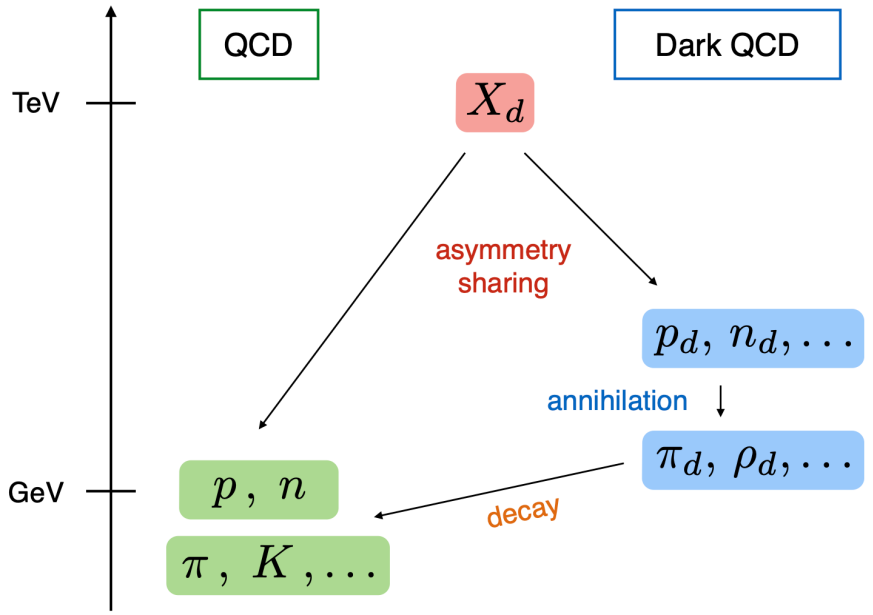
Youngst@rs – Colours in Darkness

18.10.2023

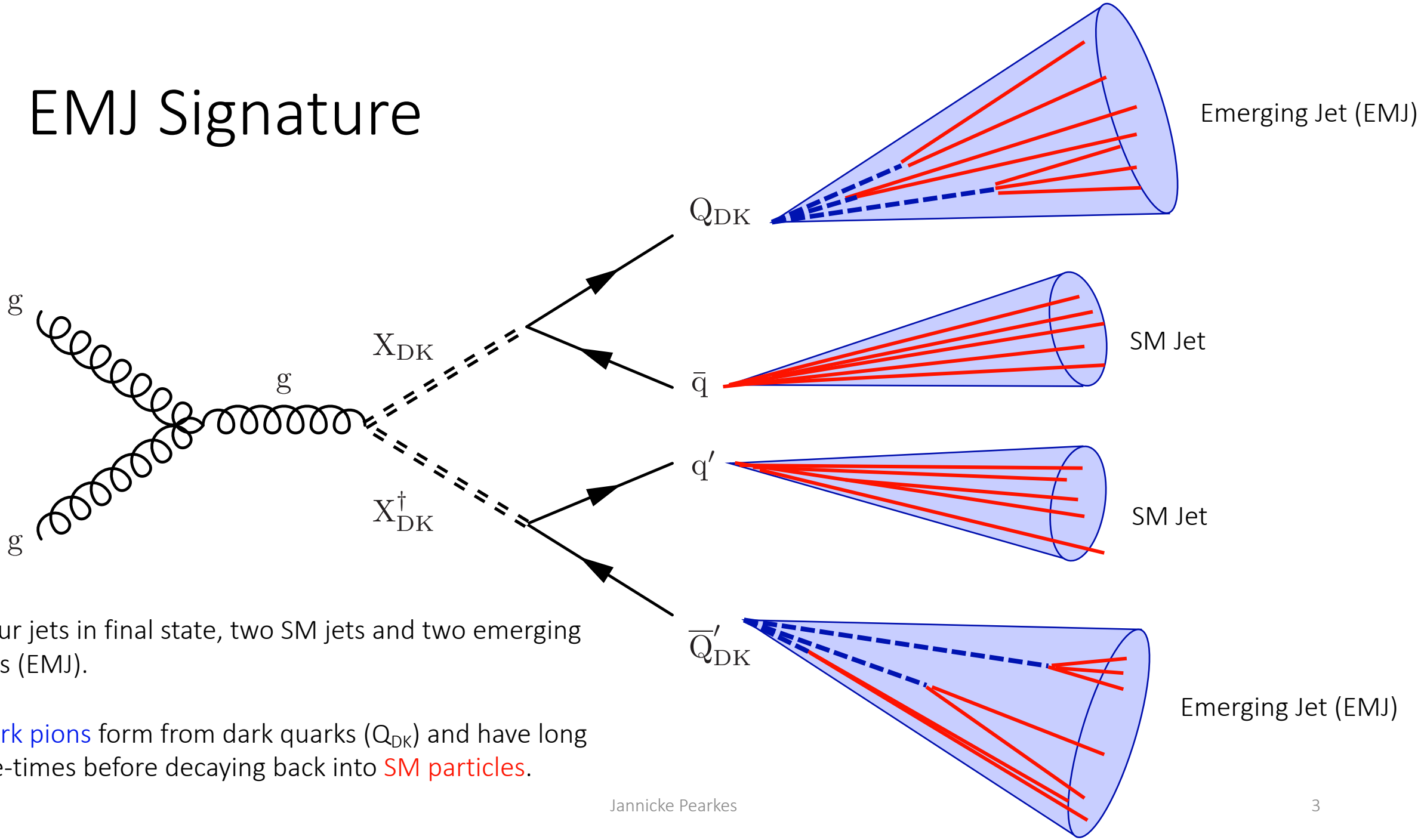
University of Colorado **Boulder**

Outline

- Focusing on Early Run 2 Result (16.1 fb^{-1} at 13 TeV) with 2016 data
- Search for G_{SM} ×$SU(N_{DK})$, dark QCD model with emerging jet signature as described in Schwaller, Stolarski, and Weiler, JHEP 05 (2015) 059 (https://arxiv.org/abs/1502.05409)
- Our result published in JHEP 02 (2019) 179 (https://arxiv.org/abs/1810.10069)
- New results with the Full Run 2 dataset coming soon!



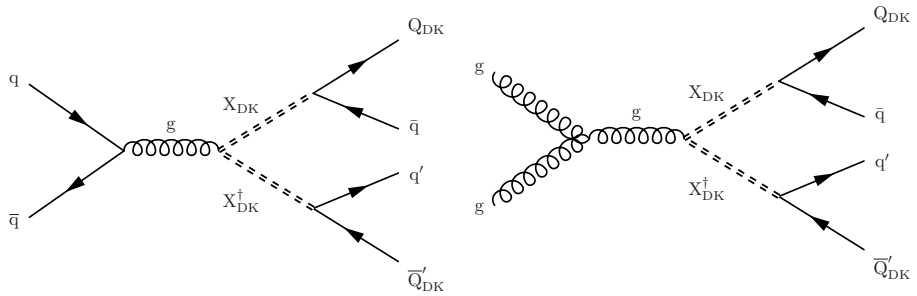
EMJ Signature



Four jets in final state, two SM jets and two emerging jets (EMJ).

Dark pions form from dark quarks (Q_{DK}) and have long life-times before decaying back into SM particles.

Simulation



EMJ signal simulated at leading order with $SU_{DK}(3)$ hidden valley model implemented with modified PYTHIA 8.212

Number of dark colours = 3, number of dark flavours = 7
$\Lambda = m_{Q,DK}$, $\Gamma_{X,DK} = 10$ GeV
Q_{DK} all mass degenerate, couple to SM down quark
$m_{\pi,DK} = 0.5\ m_{Q,DK}$
$m_{\rho,DK} = 2\ m_{Q,DK}$

By making these assumptions, we reduce down to three free parameters: $m_{X,DK}$, $m_{\pi,DK}$, $c\tau_{\pi,DK}$

Signal model parameters	List of values
$m_{X_{DK}}$ [GeV]	400, 600, 800, 1000, 1250, 1500, 2000
$m_{\pi_{DK}}$ [GeV]	1, 2, 5, 10
$c\tau_{\pi_{DK}}$ [mm]	1, 2, 5, 25, 45, 60, 100, 150, 225, 300, 500, 1000

SM QCD background simulated at leading order MADGRAPH5 aMC@NLO 2.2.2 or PYTHIA 8.2 with NNPDF3.0.
Parton shower simulated with PYTHIA using the underlying-event tune CUETP8M1.

Event Selection

Trigger

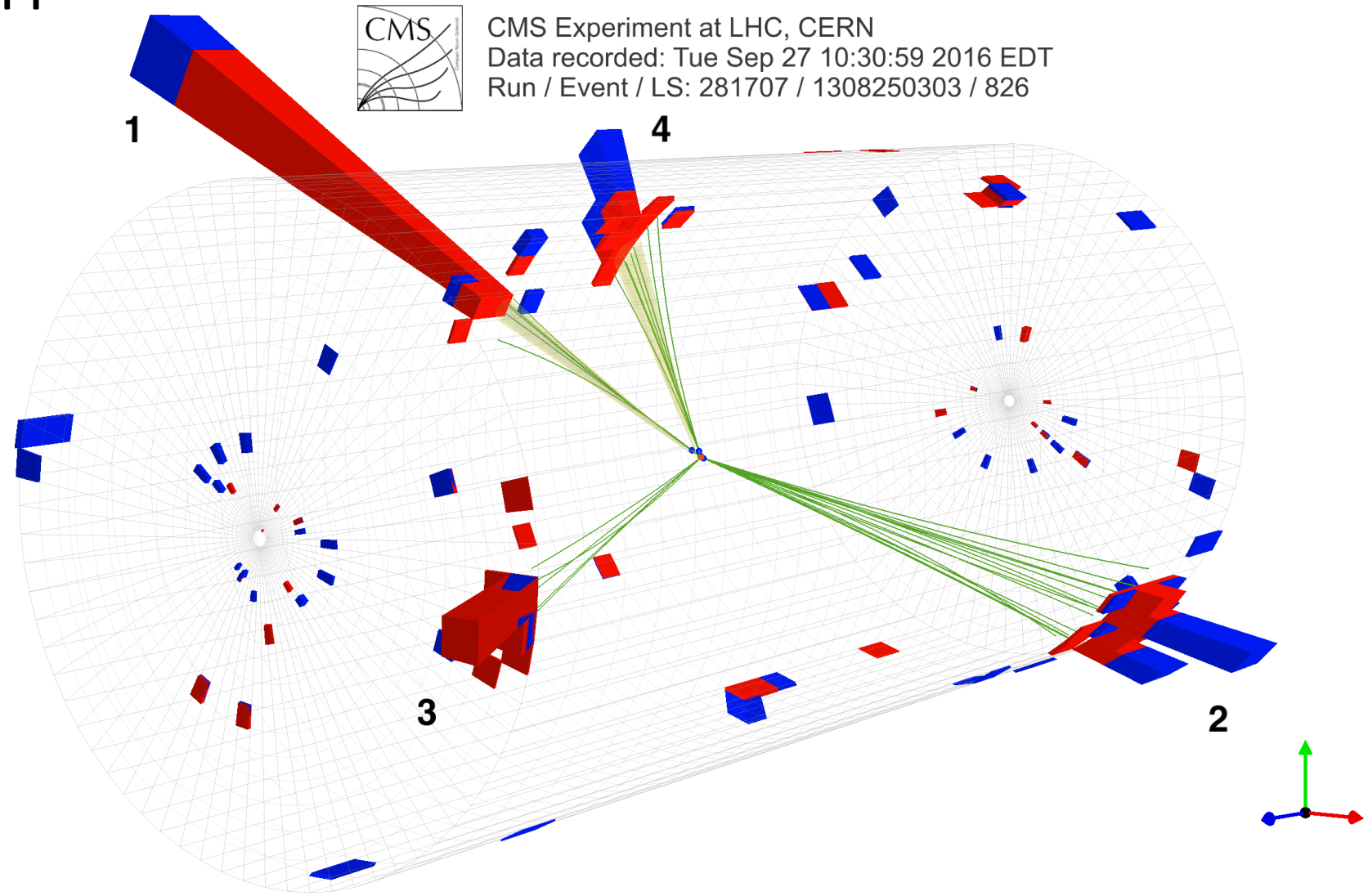
- H_T trigger, $\sum$ jet p_T > 900 GeV

Event selection

- Four jets (R = 0.4) with $|\eta| < 2.0$ and $p_T > 100$ GeV
- At least one track in each jet
- < 90% of jet energy from electrons or photons to reduce SM electron background

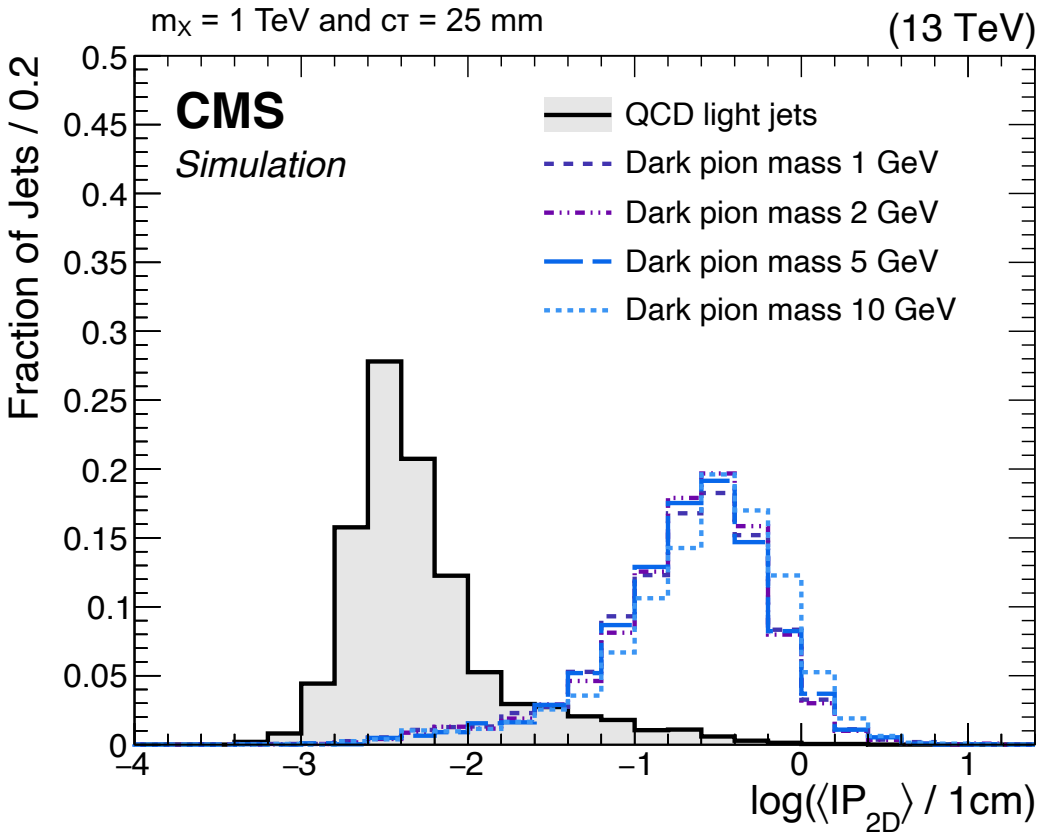
Track selection

- High purity tracks selected with p_T>1 GeV within R = 0.4 of jet for EMJ tagging

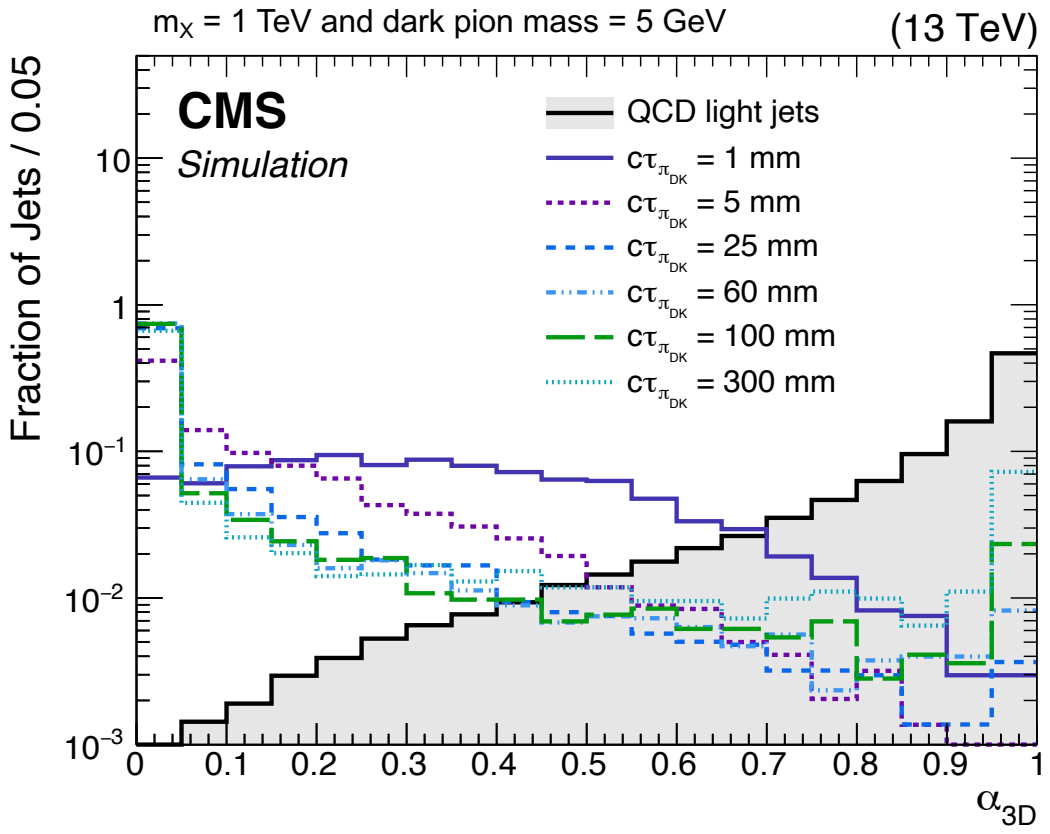


Emerging Jet Selection Variables

Variables designed to reject QCD background



Median unsigned transverse impact parameter



Prompt track momentum fraction

Cut based selections

distance of track from primary vertex in z
(used for pile-up rejection)

prompt track selection

Emerging jet "tagging" groups
Require 1 or more jets to fall into EMJ sets 1-6

Groups 7-8 used to produce SM QCD-enhanced regions for background estimate

Criteria group	PU_{dz} (<) [cm]	D_N (<)	$\langle IP_{2D} \rangle$ (>) [cm]	α_{3D} (<)
EMJ-1	2.5	4	0.05	0.25
EMJ-2	4.0	4	0.10	0.25
EMJ-3	4.0	20	0.25	0.25
EMJ-4	2.5	4	0.10	0.25
EMJ-5	2.5	20	0.05	0.25
EMJ-6	2.5	10	0.05	0.25
EMJ-7	2.5	4	0.05	0.40
EMJ-8	4.0	20	0.10	0.50

Cut based selections

distance of track from primary vertex in z (used for pile-up rejection)

prompt track selection

Emerging jet "tagging" groups
Require 1 or more jets to fall into EMJ sets 1-6

Groups 7-8 used to produce SM QCD-enhanced regions for background estimate

Criteria group	PU_{dz} (<) [cm]	D_N (<)	$\langle IP_{2D} \rangle$ (>) [cm]	α_{3D} (<)
EMJ-1	2.5	4	0.05	0.25
EMJ-2	4.0	4	0.10	0.25
EMJ-3	4.0	20	0.25	0.25
EMJ-4	2.5	4	0.10	0.25
EMJ-5	2.5	20	0.05	0.25
EMJ-6	2.5	10	0.05	0.25
EMJ-7	2.5	4	0.05	0.40
EMJ-8	4.0	20	0.10	0.50

Cut sets use different cuts on H_T, jet p_T, p_T^{miss}, number of EMJs, and EMJ "tagging" groups to target different regions of parameter space.

Set 8 & 9 used for validation of background estimate.

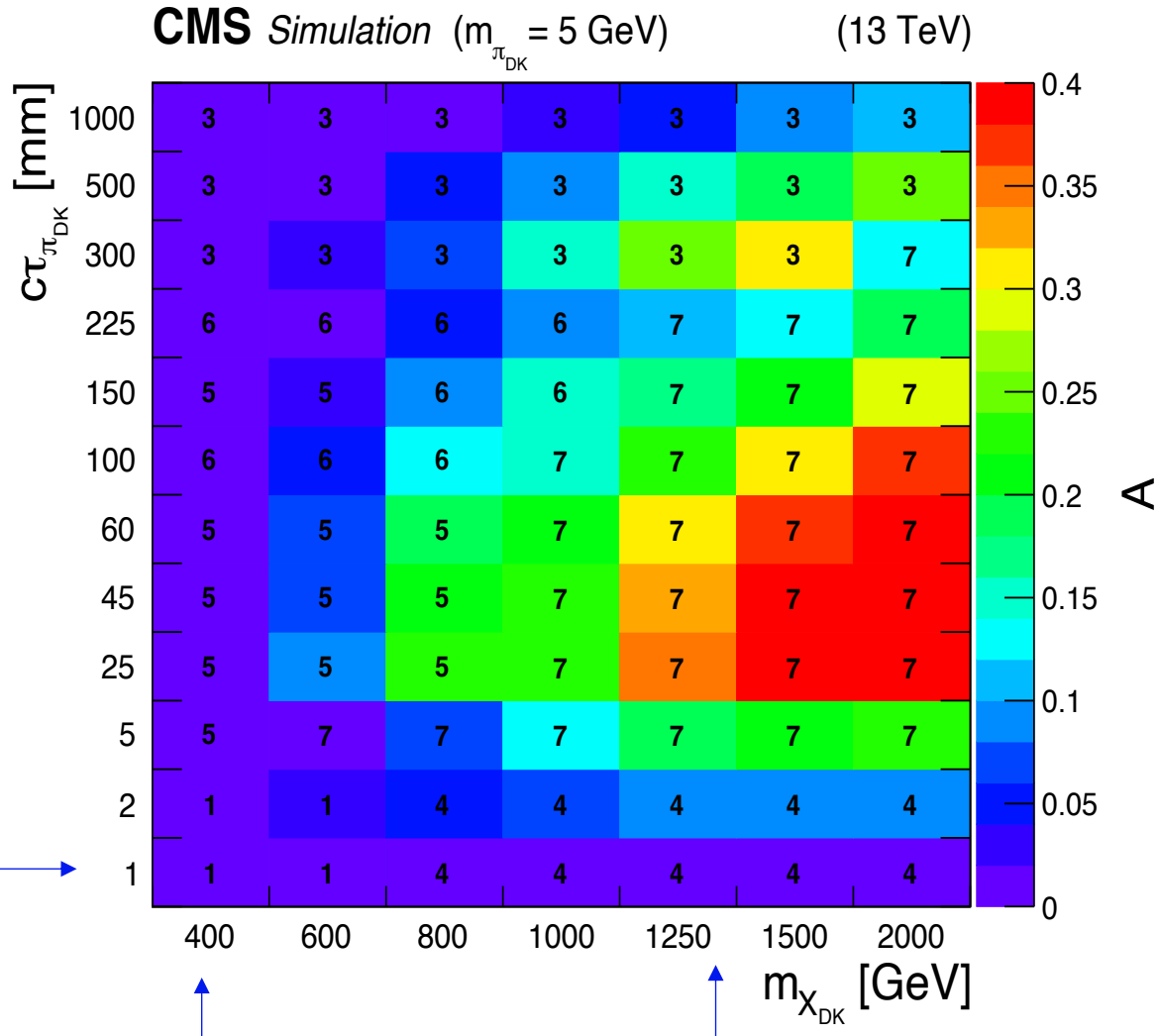
Set number	H_T	$p_{T,1}$	$p_{T,2}$	$p_{T,3}$	$p_{T,4}$	p_T^{miss}	$n_{EMJ}(\geq)$	EMJ group	no. models
1	900	225	100	100	100	0	2	1	12
2	900	225	100	100	100	0	2	2	2
3	900	225	100	100	100	200	1	3	96
4	1100	275	250	150	150	0	2	1	49
5	1000	250	150	100	100	0	2	4	41
6	1000	250	150	100	100	0	2	5	33
7	1200	300	250	200	150	0	2	6	103
8	900	225	100	100	100	0	2	7	SM QCD-enhanced
9	900	225	100	100	100	200	1	8	

Signal Acceptance

Large cτ - more displaced tracks outside of tracking volume

Small cτ – dark pions decay very fast, difficult to distinguish from prompt QCD background

Number in each bin is cut set with highest significance

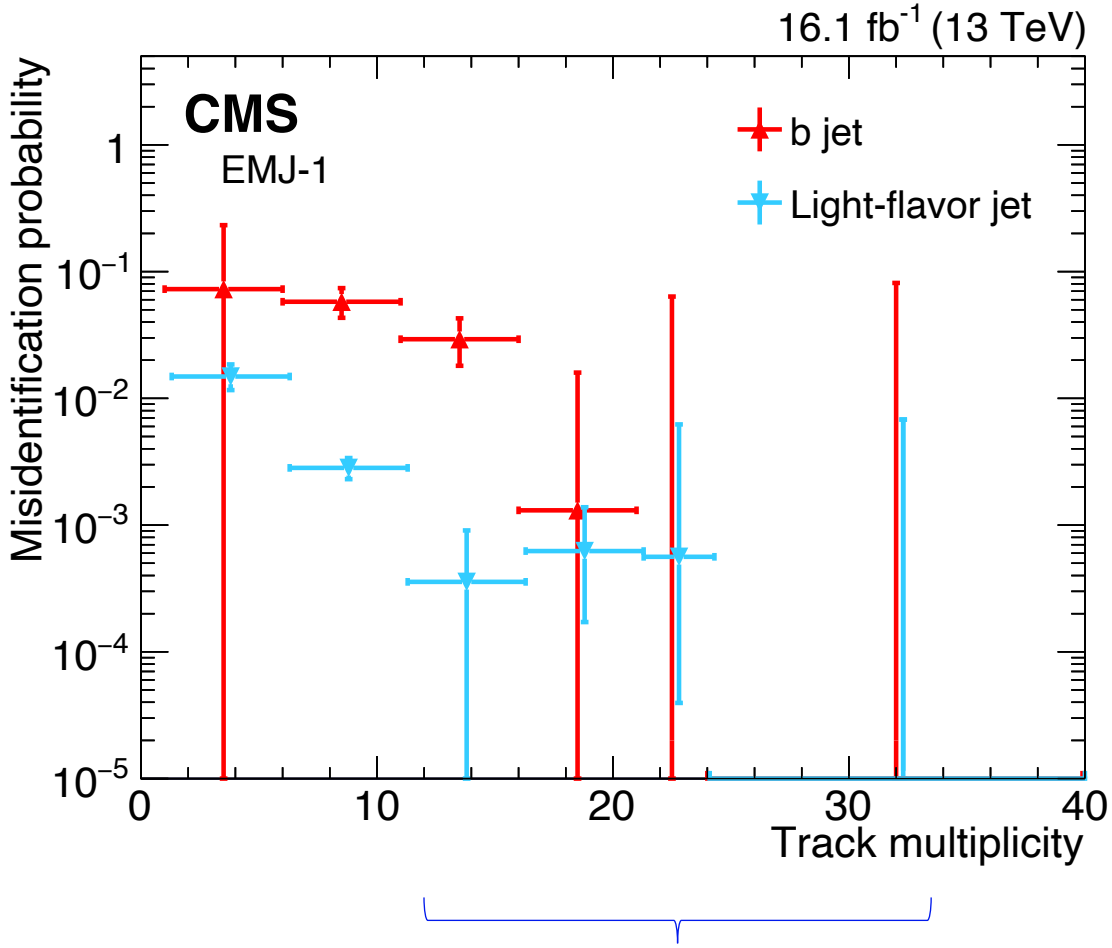


Lower efficiency at low mediator mass (m_X) due to hadronic jet momentum trigger of H_T > 900 GeV

Better signal efficiency at high m_X, easier to trigger, and larger boost results in longer distances travelled by dark pion.

QCD Background

- Background comes from SM QCD jets misidentified as emerging jets
- Misidentification probability depends on jet flavour and track multiplicity
- Bottom quark jets have displaced signature – more likely to fake EMJs
- High track multiplicity allows for better EMJ identification



large error bars due to low statistics at high track multiplicities

Background Estimation Method

- Data driven background estimate
- Misidentification rate estimated in data using signal-free photon+jet control regions

- Two photon control regions used:
  - b-jet enhanced and b-jet suppressed CRs
  - Flavour composition in control regions determined using a template fit to the CSVv2 b-tag score distribution

- Apply misidentification rate from photon+jet control regions to signal-depleted QCD control regions to estimate yield in signal region

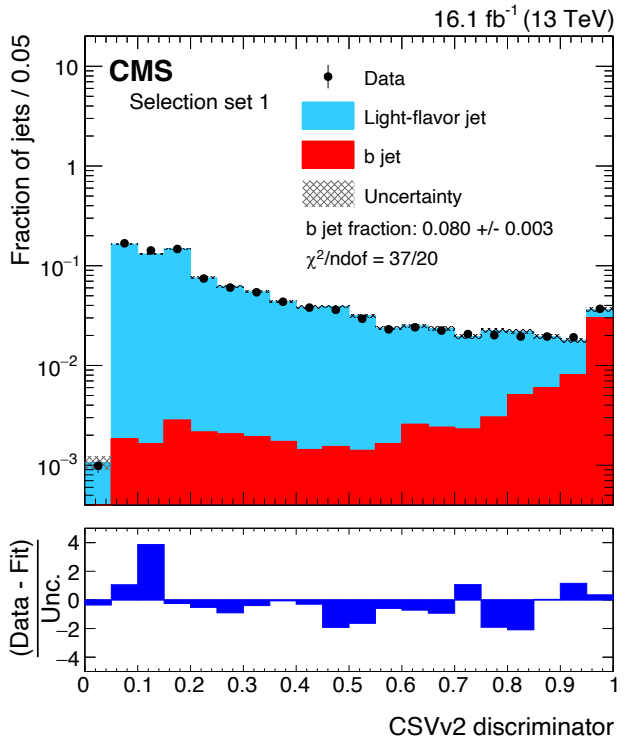
- Validated in QCD enhanced regions (cut sets 7 & 8)

Photon CRs

- $N_\gamma = 1$, $p_T > 175$ GeV, $|\eta| < 1.44$
- $N_{SMJ} \geq 1$, $p_T > 50$ GeV, CSVv2 > 0.8 OR < 0.2

QCD CRs

- $N_{EMJ} = 0$, $p_T^{miss} > 200$ OR $N_{EMJ} = 1$, $p_T^{miss} \geq 0$

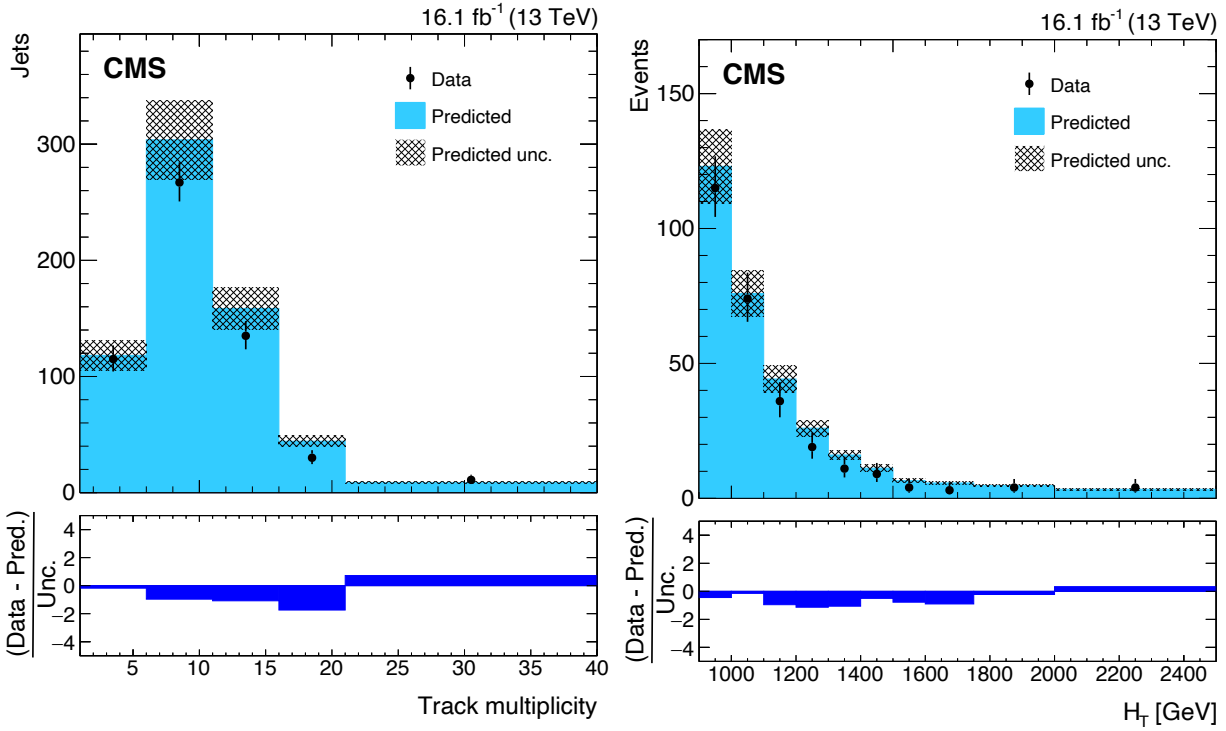


Background Estimate Validation

The accuracy of the background estimation method is evaluated in QCD enhanced validation regions 7&8.

Uncertainties from the background estimate from:
- b quark fraction, difference between CSVv2 fit and MC truth, 0.6–5%
- non-b quark composition, difference between photon and QCD control regions, (1.4-28.3%)

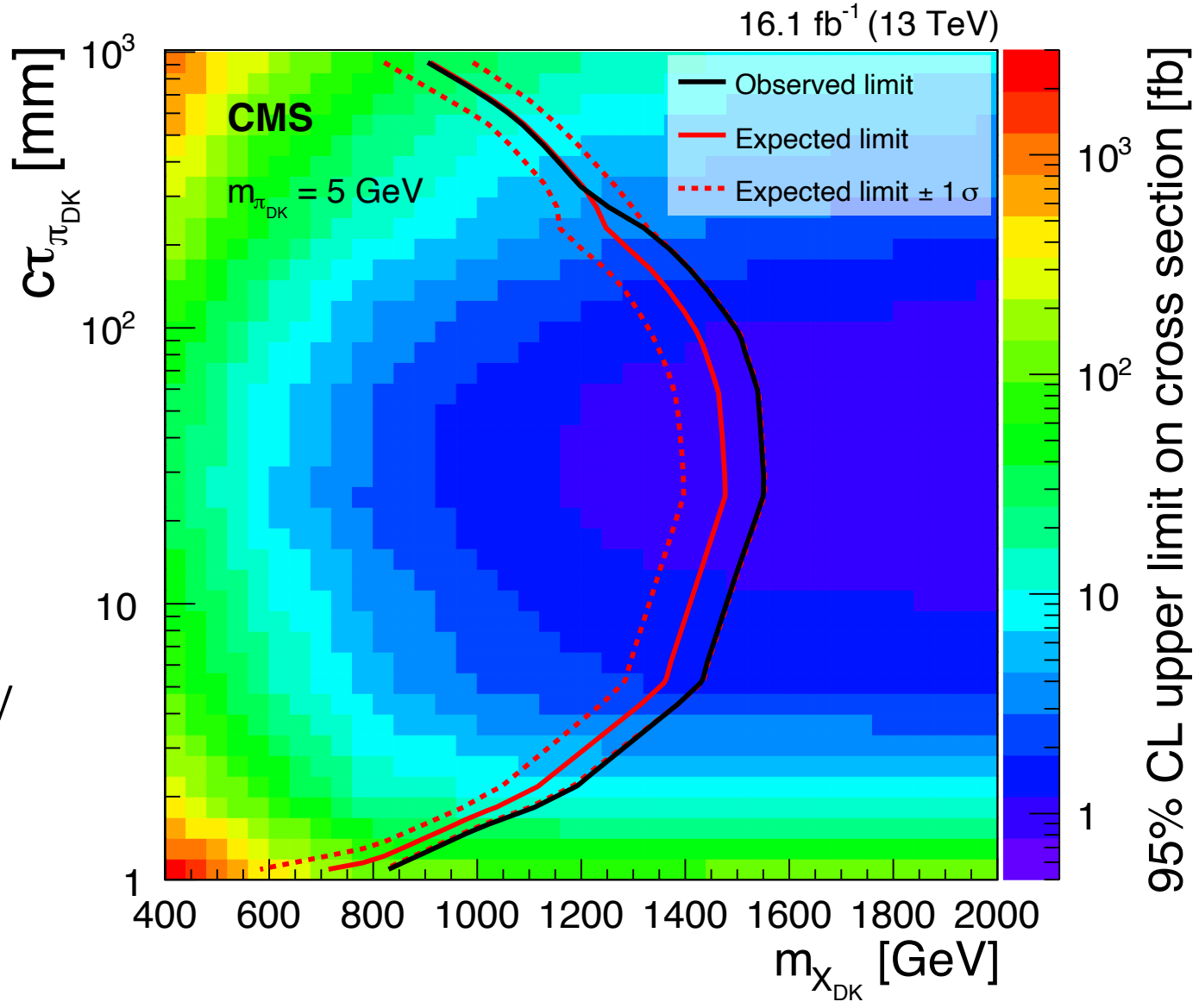
Set number	Source of uncertainty (%)	
	b quark fraction	non-b quark composition
1	2.8	1.4
2	0.6	4.4
3	2.9	28.3
4	5.0	4.4
5	0.9	4.0
6	1.6	2.1
7	1.0	6.3



Background estimate in SM QCD-enhanced sideband (selection set 8)

Results

- No significant excess above the SM expectation observed using 16.1 fb^{-1} of data collected at √s = 13 TeV in 2016
- Limits do not depend strongly on $m_{\pi,DK}$
- Exclude $m_{X,DK}$ from 400-1250 GeV for $c\tau_{\pi,DK}$ between 5-225mm



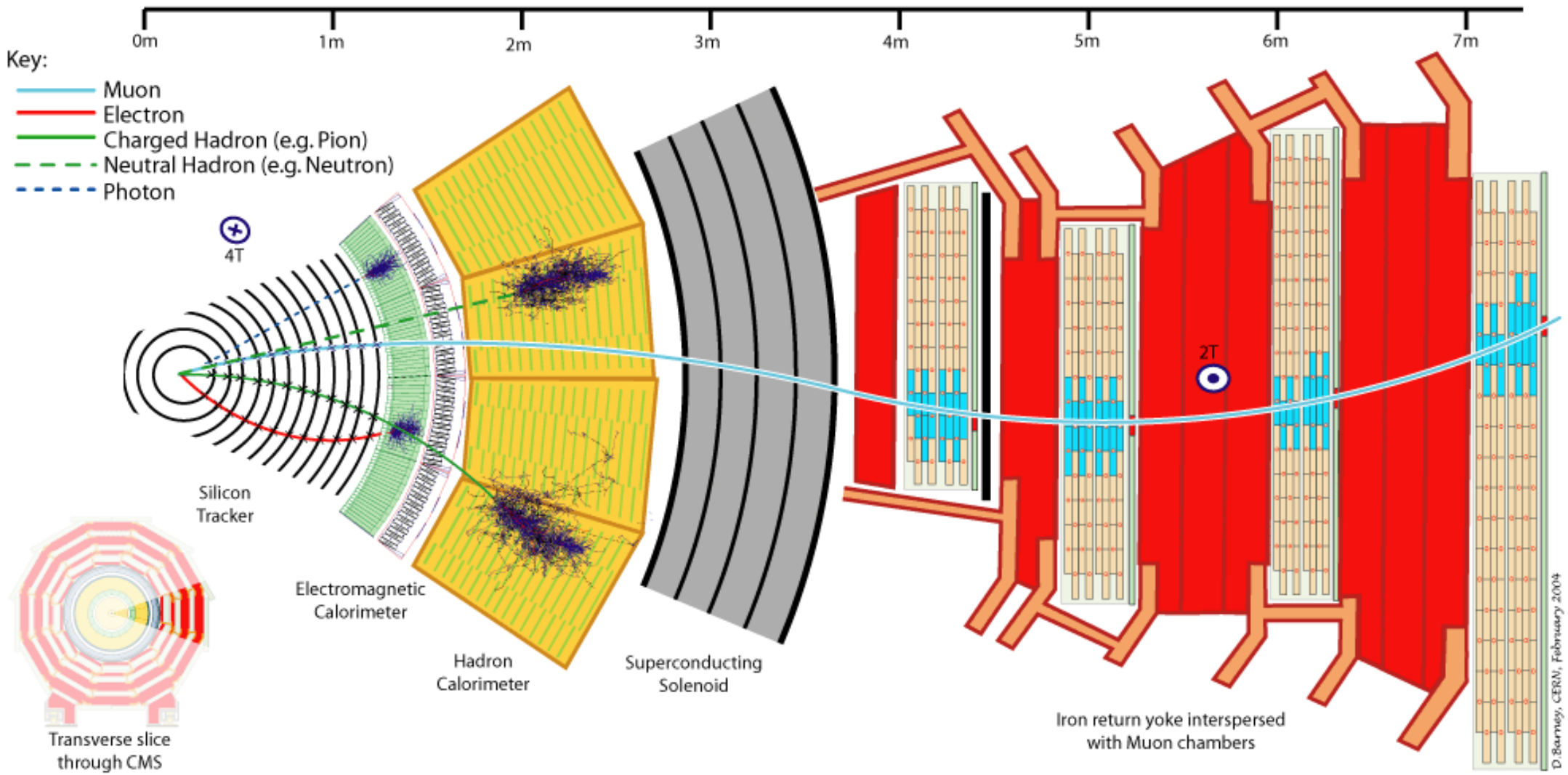
Outlook

- Early Run 2 CMS search found no statistically significant evidence for EMJs in parameter space studied
- Full Run 2 result with ~140 fb^{-1} will increase size of dataset by ~10x – *coming soon*

Future directions:

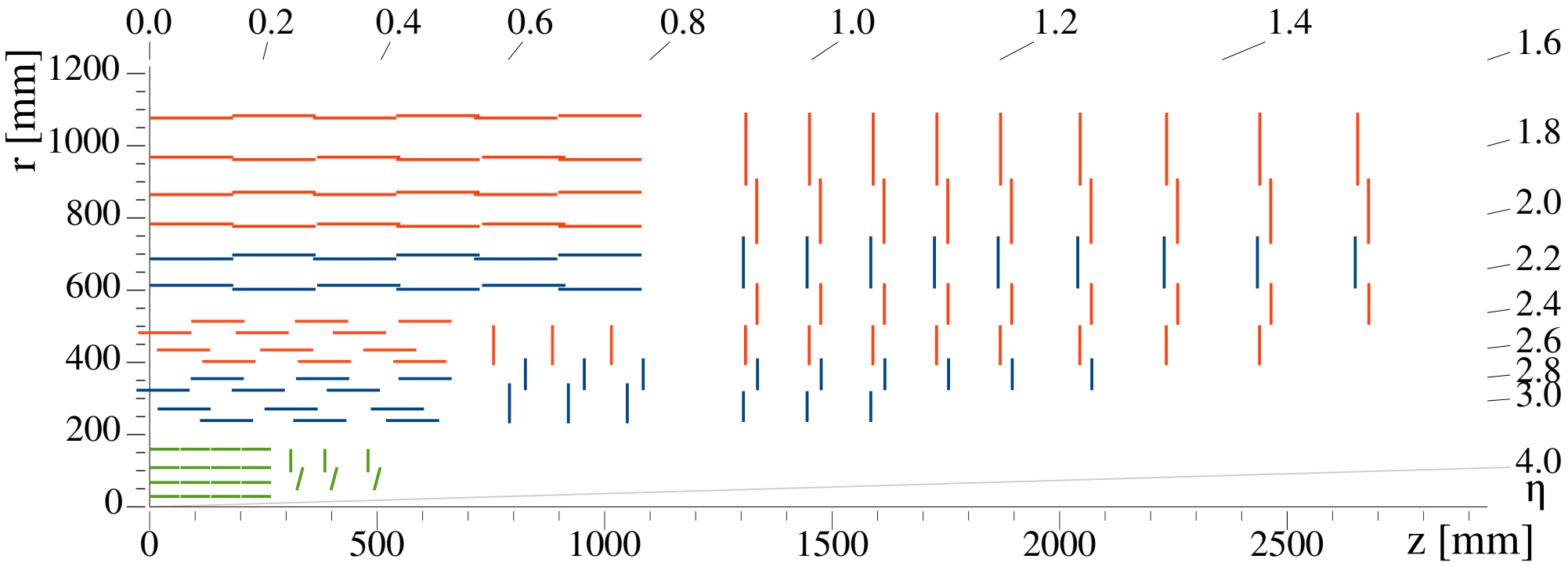
- Explore a larger variety of dark QCD models
- More targeted searches using machine learning
- Combination with other dark shower signatures (semi-visible jets & SUEPs)
- New triggers for Run 3 (displaced jet / anomaly detection triggers)
- Track trigger for CMS @ HL-LHC

Thank you!



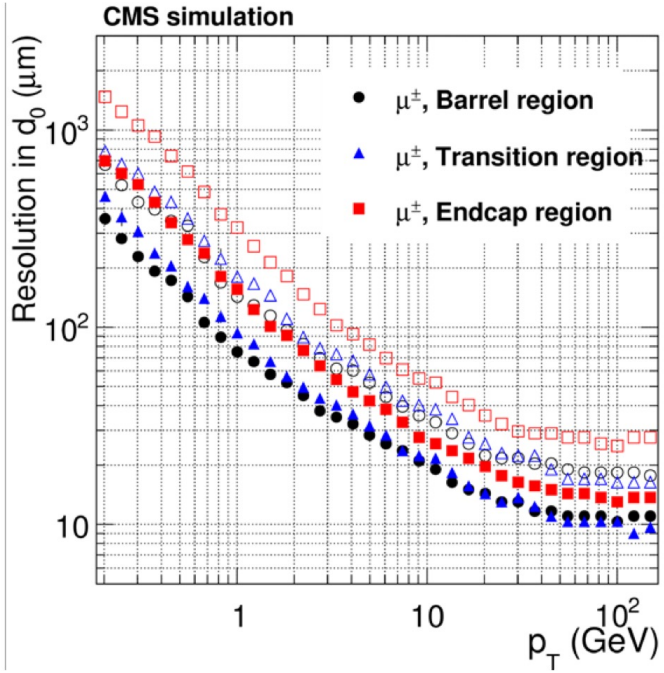
Source https://cds.cern.ch/record/2205172

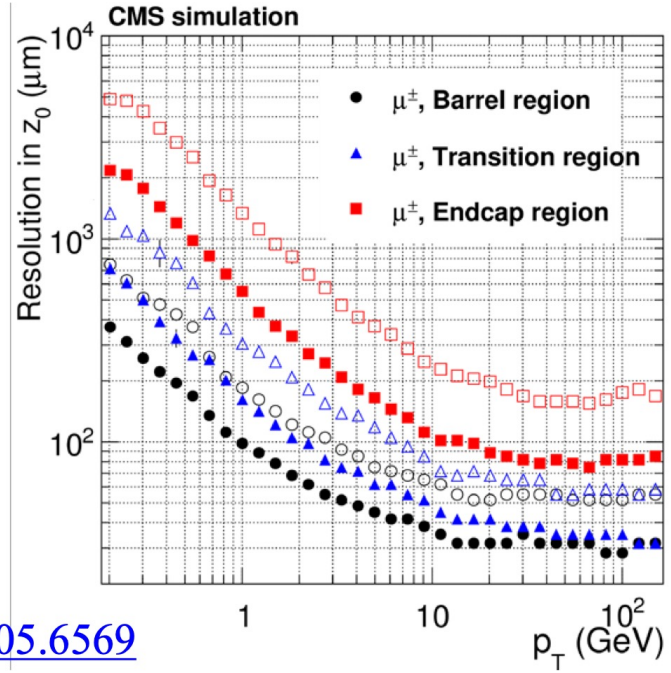
CMS Tracker Layout



https://twiki.cern.ch/twiki/pub/CMSPublic/DPGResultsTRK/Phase1_Tracker_1Quarter.pdf

CMS Track Resolutions





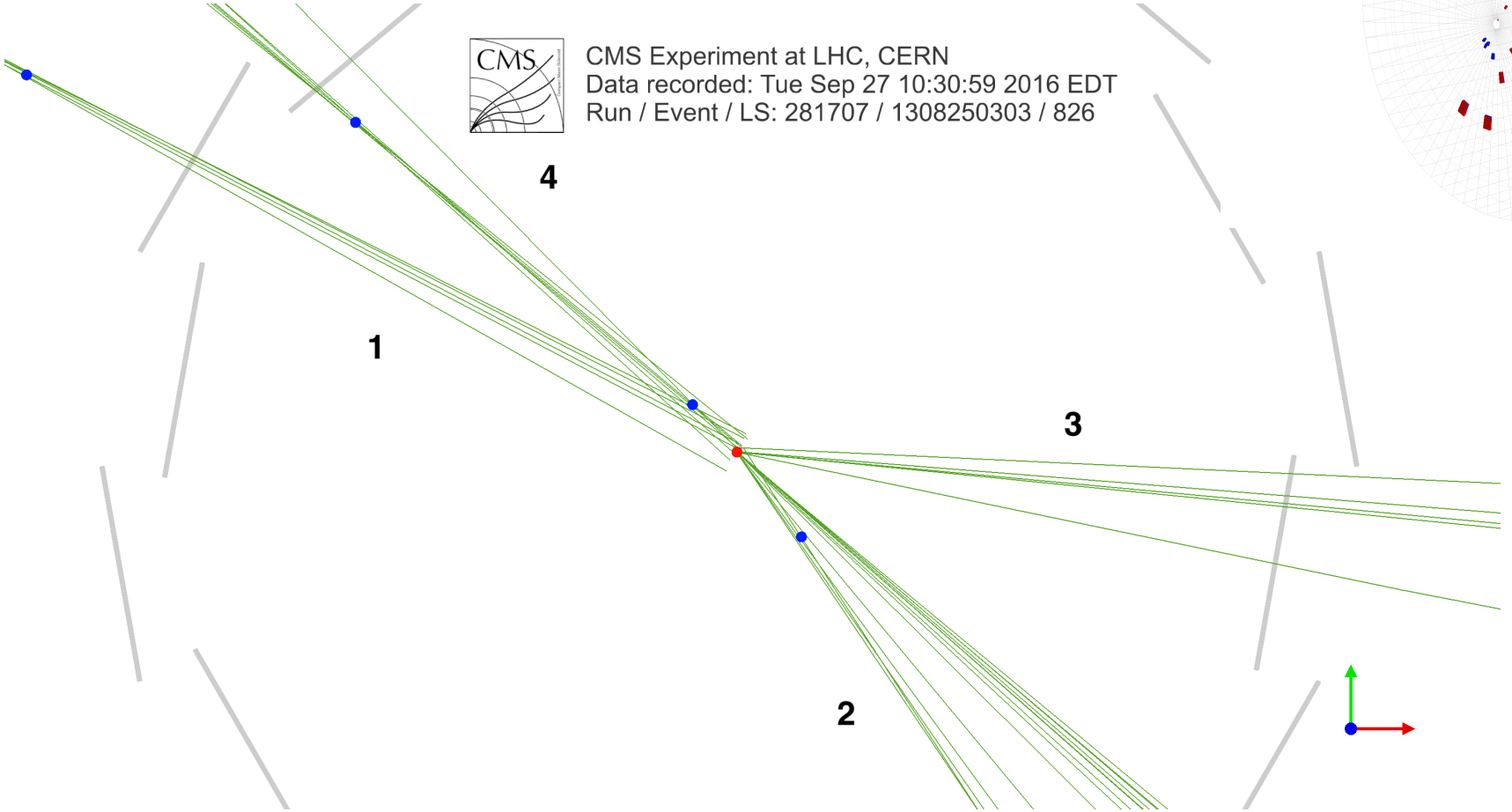
arXiv:1405.6569

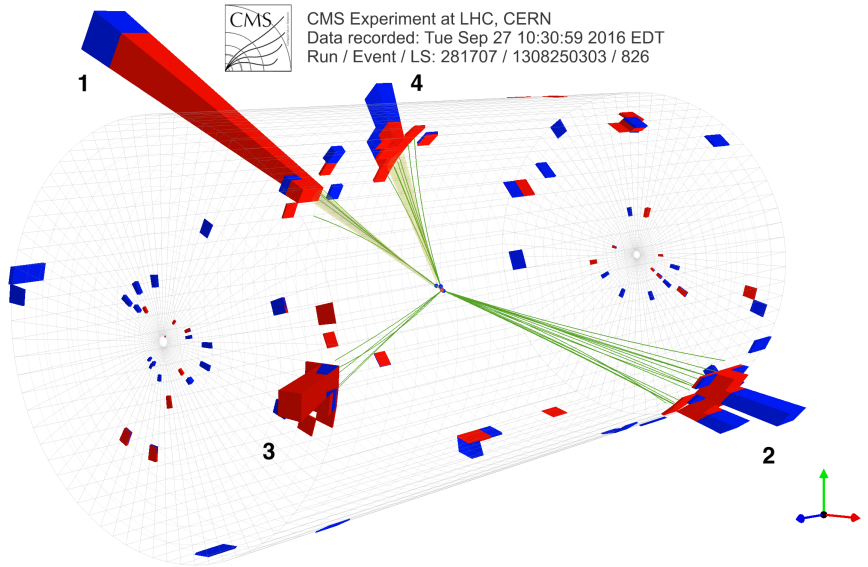
Resolutions

($1 < p_T < 10$ GeV, $|\eta| < 1.4$)

- p_T: 1.5%
- transverse: 25–90 μm
- longitudinal: 45–150 μm

Emerging Jet Selection





Event display of an event passing both selection set 1 and selection set 5. Jets 1 and 4 pass the emerging jet criteria.

Dark pion lifetime

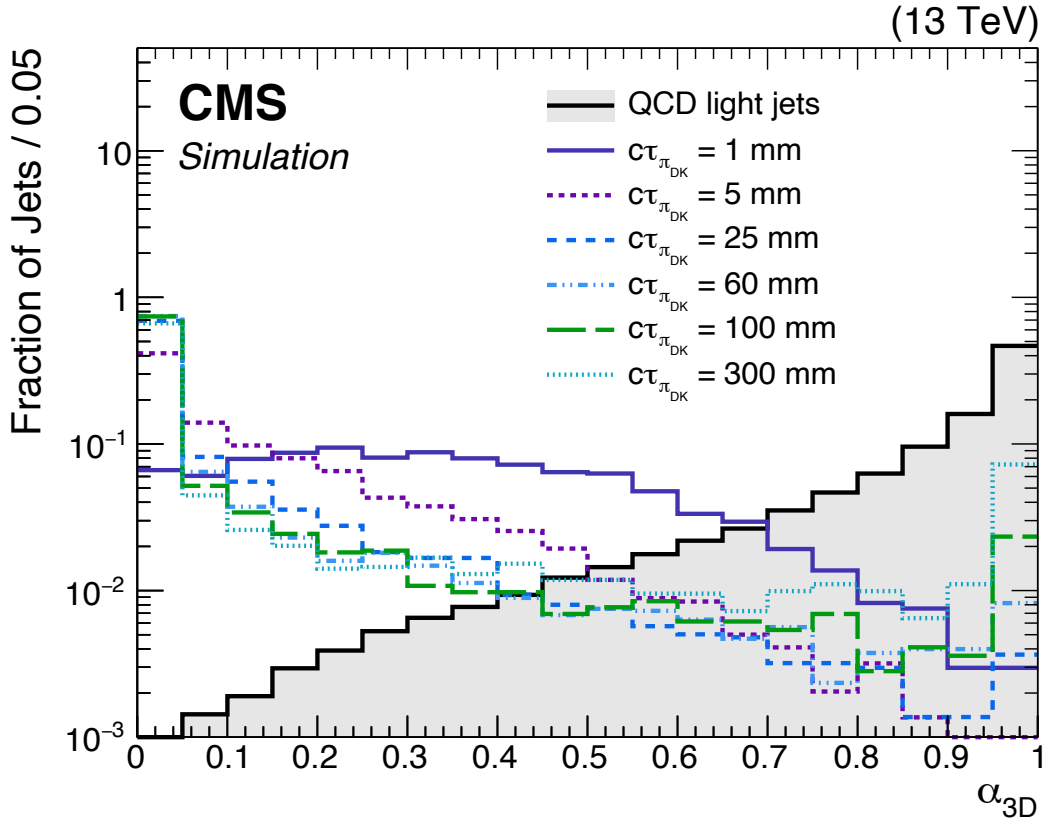
Higher m_X -> longer dark pion lifetime

$$c\tau \approx 80\,\text{mm}\left(\frac{1}{\kappa^4}\right)\left(\frac{2\,\text{GeV}}{f_{\pi_{\text{DK}}}}\right)^2\left(\frac{100\,\text{MeV}}{m_{\text{down}}}\right)^2\left(\frac{2\,\text{GeV}}{m_{\pi_{\text{DK}}}}\right)\left(\frac{m_{\text{X}_{\text{DK}}}}{1\,\text{TeV}}\right)^4$$

Heavier dark pion –> longer lifetime,
but not as strong of an effect

where κ is the appropriate element of the $N_{C_{\text{DK}}} \times 3$ matrix of Yukawa couplings between the mediator particle, the quarks, and the dark quarks; $f_{\pi_{\text{DK}}}$ is the dark pion decay constant; and m_{down}, $m_{\pi_{\text{DK}}}$, and $m_{\text{X}_{\text{DK}}}$ are the masses of the down quark, the dark pion, and the mediator particle, respectively.

Momentum Fraction from Prompt Tracks



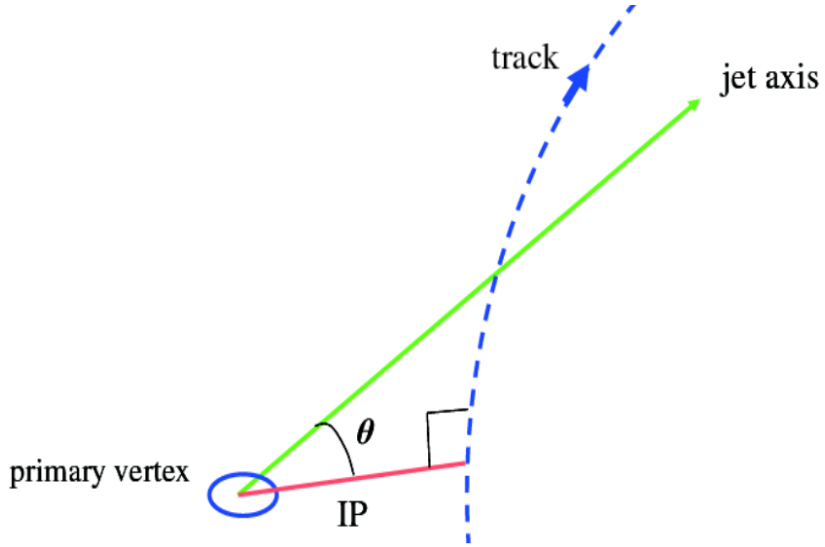
$$\alpha_{3D} = \frac{\sum p_T(D_N < D_{N,cut})}{\sum p_T}$$

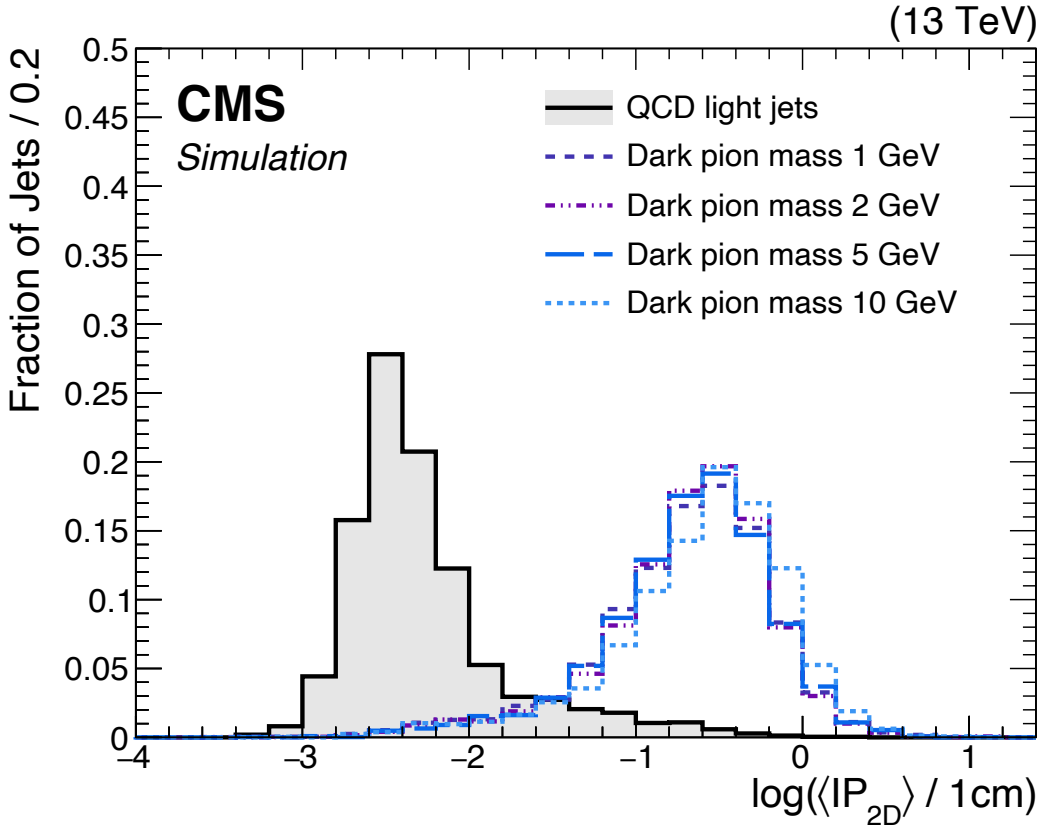
$$D_N = \sqrt{\left[\frac{z_{PV} - z_{trk}}{0.01\,\text{cm}}\right]^2 + [IP_{sig}]^2}$$

Fraction of momentum carried by prompt tracks will be high for QCD jets, but low for emerging jets.

Median Transverse Impact Parameter

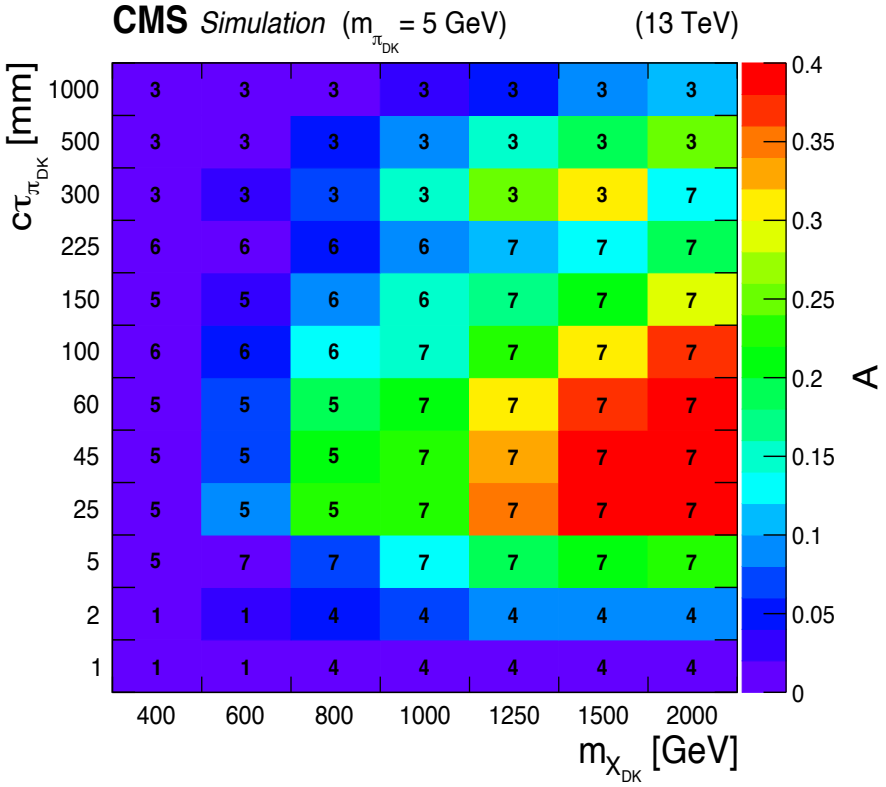
Emerging jets will contain more tracks with a large unsigned transverse impact parameter (IP_{2D})





Cut set definitions

Sets optimized to maximize: $$\frac{S}{\sqrt{S+B+(0.1B)^2}}$$



Set number	H_T	$p_{T,1}$	$p_{T,2}$	$p_{T,3}$	$p_{T,4}$	p_T^{miss}	$n_{EMJ}(\geq)$	EMJ group	no. models
1	900	225	100	100	100	0	2	1	12
2	900	225	100	100	100	0	2	2	2
3	900	225	100	100	100	200	1	3	96
4	1100	275	250	150	150	0	2	1	49
5	1000	250	150	100	100	0	2	4	41
6	1000	250	150	100	100	0	2	5	33
7	1200	300	250	200	150	0	2	6	103
8	900	225	100	100	100	0	2	7	SM QCD-enhanced
9	900	225	100	100	100	200	1	8	

targeting prompt EMJs

targeting longer life-times

targeting higher masses

used for validation of background estimate

CMS: Emerging jets background estimate

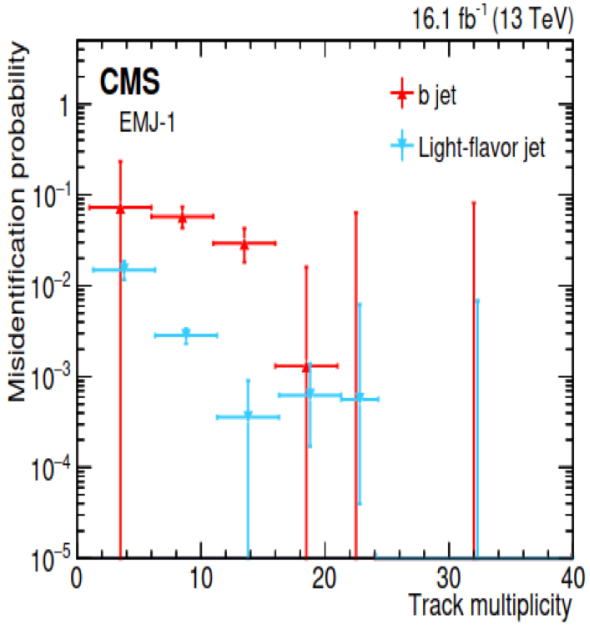
- Signal consists of tracks originating from several vertices at varying IP from PV
  - Sub-jet axes point out radially from PV
  - Heavy flavor jets could mimic the signature of short π_d lifetime
- Multijet QCD (light and b-jets) is the main background
- Estimate fake rate in two γ+jets CRs in data, one with heavy flavor and one without
  - The b-jet fraction of each CR (f_b) is fitted to two MC templates (light and b jets)

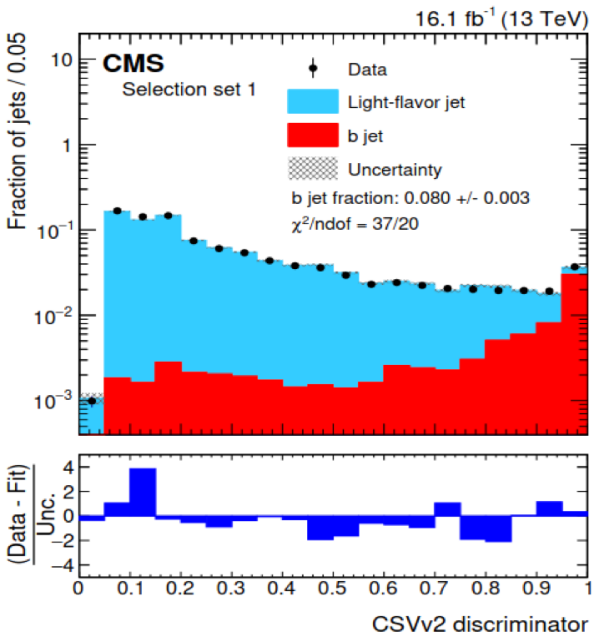
$$\begin{pmatrix}\epsilon_{f1}\\ \epsilon_{f2}\end{pmatrix} = \begin{pmatrix} f_{b1} & 1-f_{b1}\\ f_{b2} & 1-f_{b2}\end{pmatrix}\begin{pmatrix}\epsilon_{fb}\\ \epsilon_{fl}\end{pmatrix}$$

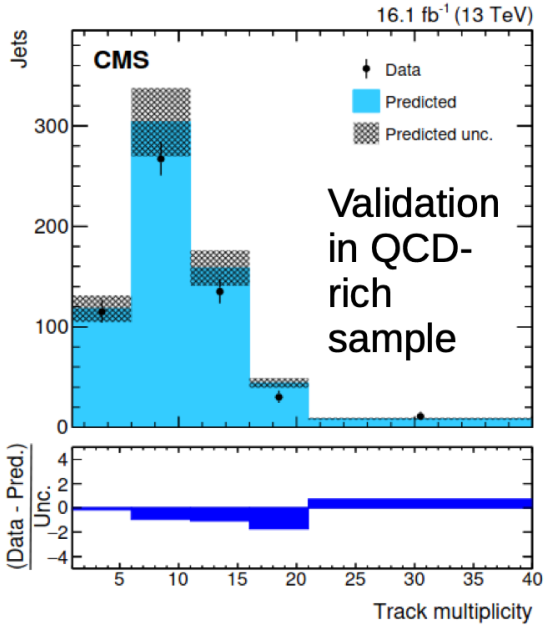
$$\begin{pmatrix}\epsilon_{\mathrm{fb}}\\ \epsilon_{\mathrm{fl}}\end{pmatrix} = \begin{pmatrix} \frac{1-f_{\mathrm{b2}}}{f_{\mathrm{b1}}-f_{\mathrm{b2}}} & \frac{-(1-f_{\mathrm{b1}})}{f_{\mathrm{b1}}-f_{\mathrm{b2}}}\\ \frac{-f_{\mathrm{b2}}}{f_{\mathrm{b1}}-f_{\mathrm{b2}}} & \frac{f_{\mathrm{b1}}}{f_{\mathrm{b1}}-f_{\mathrm{b2}}}\end{pmatrix}\begin{pmatrix}\epsilon_{\mathrm{f1}}\\ \epsilon_{\mathrm{f2}}\end{pmatrix}$$

the ϵ calculated bin-by-bin in n_{trk}

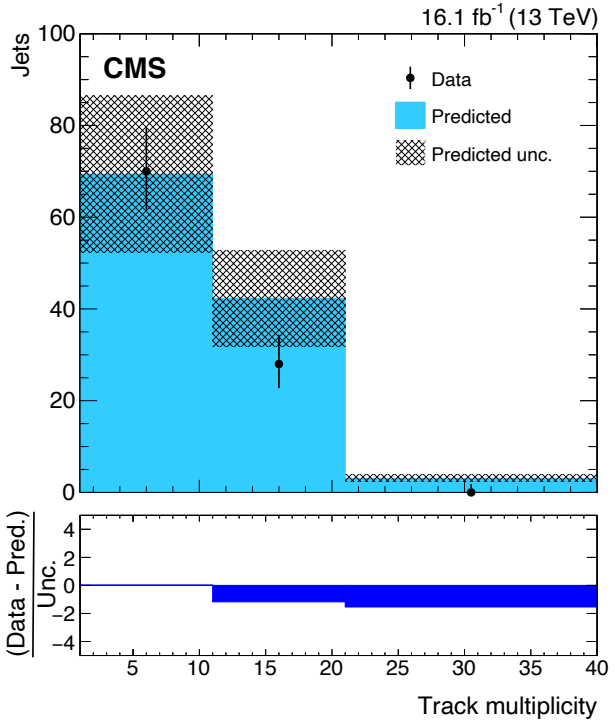
$$\epsilon_{\mathrm{f}} = \epsilon_{\mathrm{fb}} f_{\mathrm{b}} + \epsilon_{\mathrm{fl}}\,(1-f_{\mathrm{b}}) \qquad P(EMJ) \propto \epsilon_f \epsilon_f (1-\epsilon_f)(1-\epsilon_f) + \ldots$$

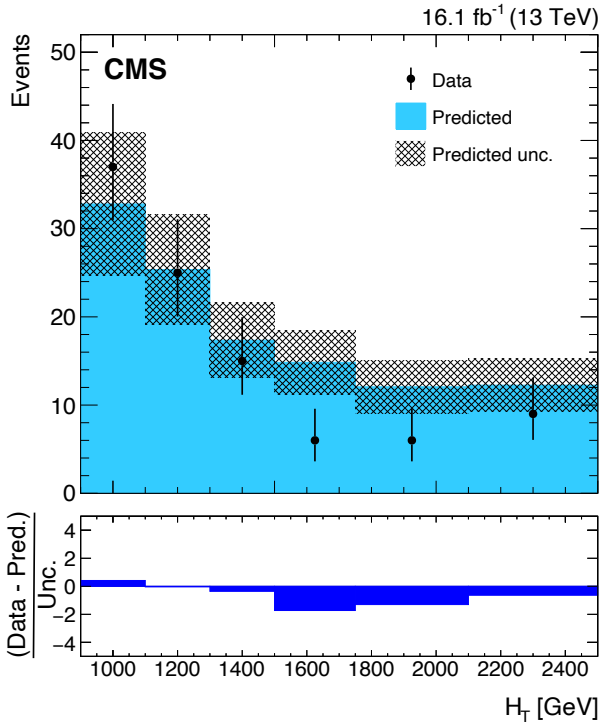




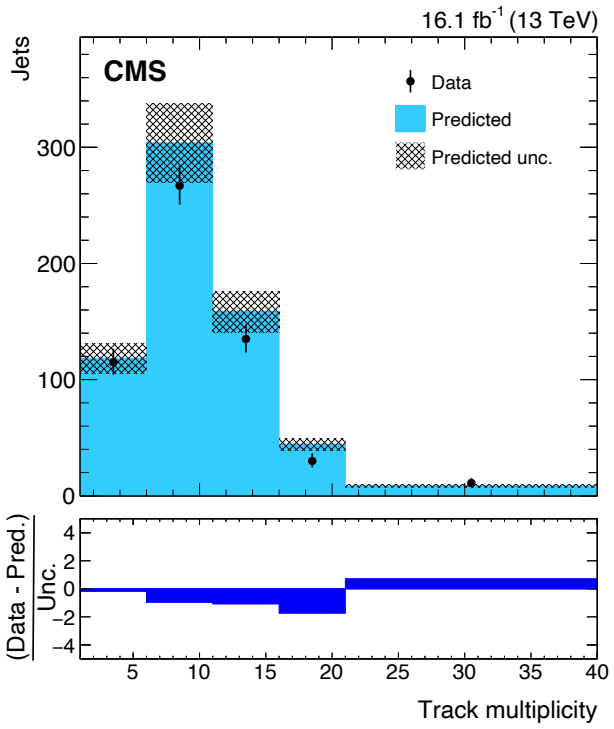


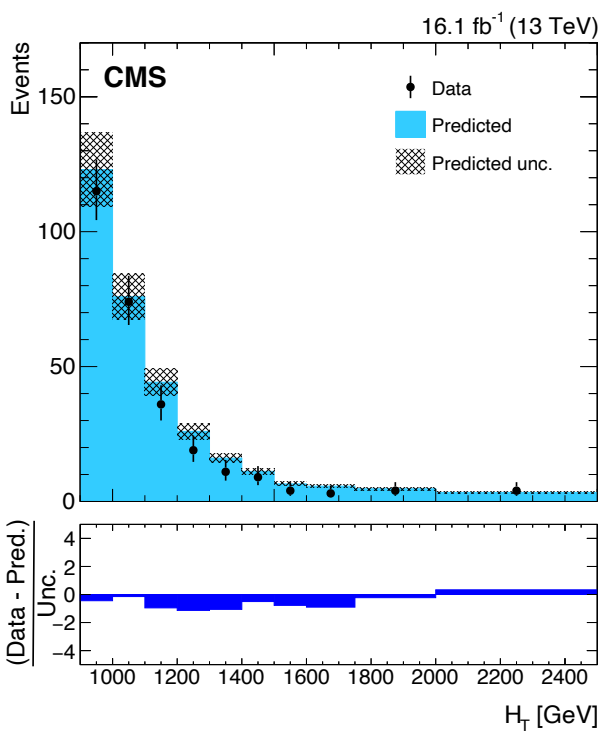
Background Estimate





Signal free region





SM QCD-enhanced (selection set 8)

Event Yields

Table 6: Expected (mean $\pm$ syst$_1$ $\pm$ syst$_2$) and observed event yields for each selection set. Uncertainties due to the limited number of events in the control sample and statistical uncertainties in the misidentification probabilities are denoted by "syst$_1$", while "syst$_2$" combines the systematic uncertainty sources discussed in Table 4. The "Signal" column shows the expected event yield for the heaviest mediator mass that can be excluded for each set, with the systematic uncertainties from sources discussed in Table 5 summed in quadrature. The associated model parameters are specified in the last three columns.

Set number	Expected	Observed	Signal	Model parameters		
				$m_{X_{DK}}$ [GeV]	$m_{\pi_{DK}}$ [GeV]	$c\tau_{\pi_{DK}}$ [mm]
1	$168 \pm 15 \pm 5$	131	36.7 ± 4.0	600	5	1
2	$31.8 \pm 5.0 \pm 1.4$	47	$(14.6 \pm 2.6) \times 10^2$	400	1	60
3	$19.4 \pm 7.0 \pm 5.5$	20	15.6 ± 1.6	1250	1	150
4	$22.5 \pm 2.5 \pm 1.5$	16	15.1 ± 2.0	1000	1	2
5	$13.9 \pm 1.9 \pm 0.6$	14	35.3 ± 4.0	1000	2	150
6	$9.4 \pm 2.0 \pm 0.3$	11	20.7 ± 2.5	1000	10	300
7	$4.40 \pm 0.84 \pm 0.28$	2	5.61 ± 0.64	1250	5	225

Systematic Uncertainties

Source	Uncertainty (%)
Track modeling	<1 – 3
MC event count	2 – 17
Integrated luminosity	2.5
Pileup	<1 – 5
Trigger	6 – 12
JES	<1 – 9
PDF	<1 – 4

Set number	Source of uncertainty (%)	
	b quark fraction	non-b quark composition
1	2.8	1.4
2	0.6	4.4
3	2.9	28.3
4	5.0	4.4
5	0.9	4.0
6	1.6	2.1
7	1.0	6.3

Main sources of uncertainty due to few events in control sample & statistics in misidentification estimate.