Searching in CMS Open Data for Dimuon Resonances with Substantial Transverse Momentum

Cari Cesarotti[1]

with Yotam Soreq[2,3], Matthew J. Strassler[1], Jesse Thaler[2], and Wei Xue[3]

Jan 21, 2019

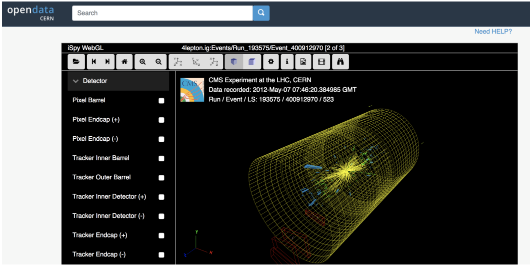
GGI Lectures on the Theory of Fundamental Interactions

[1] Department of Physics, Harvard University

[2] Center for Theoretical Physics, Massachusetts Institute of Technology

[3] Theoretical Physics Department, CERN

Open Data



CERN Open Data Portal: http://opendata.cern.ch/

What is Open Data?

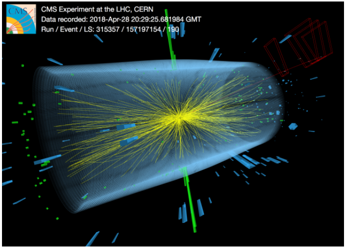
- LHC data publicly available
- Research grade data from CMS made public in 2014
  CMS Run 2011a: $\sqrt{s} = 7$ TeV, $\mathcal{L} = 2.11$ fb^{-1}

New Physics at LHC

Where can we look for new physics in old data?

- LHC is powerful probe in energy, intensity frontier
- Might have looked in the wrong place for new physics
  $\rightarrow$ How you look determines what you find
  $\rightarrow$ New physics signature buried in phase space
- First ever BSM search on Open Data

Consider class of models:
new physics with dimuon signatures
and substantial $p_T^{\mu\mu}$

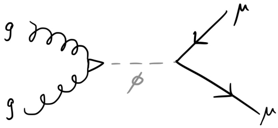


Previous searches at ATLAS & CMS with dimuons: $m_{\mu\mu}$ resonances

- High-mass: $m_{Z'} \in [120\ \text{GeV}, 3.36\ \text{TeV}]$
- Pairs of light bosons: $m_{Z'} \in [0.25, 8.5]$ GeV

LHCb searched for light new physics from direct production in *pp* collisions

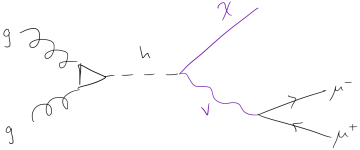
- $m_\phi < 70$ GeV



Models

We are considering a new vector particle V
produced via decay, therefore **with substantial** p_T^V

- Decays to dimuon pairs ($\mu^+\mu^-$)
- Produced by SM (W/Z/Higgs/t) or new particle
- Probing mass range of **14 GeV to 66 GeV**



Overview

We propose a search strategy that increases sensitivity for semi-inclusive searches with indirect production of new particle

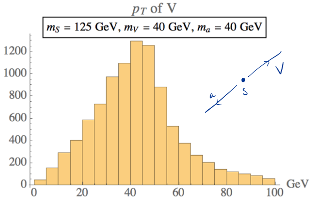
$$pp \to V + X, \quad V \to \mu^+\mu^-$$

Overview

We propose a search strategy that increases sensitivity for semi-inclusive searches with indirect production of new particle

$$pp \to V + X, \quad V \to \mu^+\mu^-$$

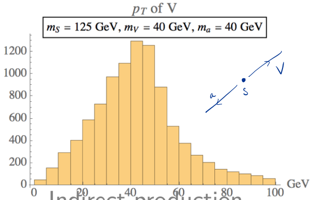
- Produced via decay: substantial p_T^V from recoil
- Cut on $p_T^{\mu\mu}$ preserves signal while reducing background



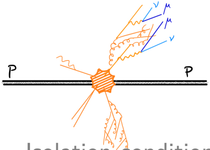
We propose a search strategy that increases sensitivity for semi-inclusive searches with indirect production of new particle

$$pp \to V + X, \quad V \to \mu^+\mu^-$$

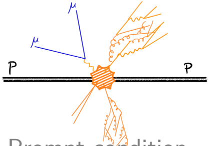
- Produced via decay: substantial p_T^V from recoil
- Cut on $p_T^{\mu\mu}$ preserves signal while reducing background
- Further reduce background: isolation and prompt cuts



Indirect production



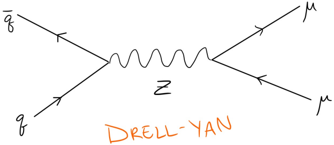
Isolation condition



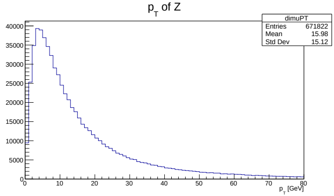
Prompt condition

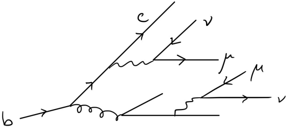
Improving sensitivity: remove background, keep signal $\left(\frac{S}{\sqrt{B}}\right)$

Backgrounds



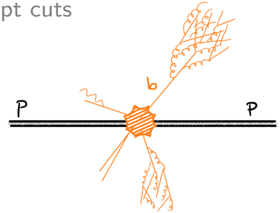
Suppressed by $p_T^{\mu\mu}$ cuts





QCD

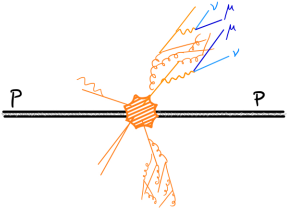
Further suppressed by isolation or prompt cuts



Isolated vs. Prompt Sample

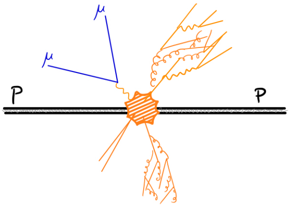
Isolated

- Both μ are isolated in detector
- Virtually eliminates QCD (heavy flavor) background



Prompt

- No isolation cut imposed
- $IP_{xy} < 100\mu$m
- QCD bgd. $\sim$ DY bgd.
- **Need Open Data to understand**



Isolated vs. prompt sample are sensitive to different classes of models

General Search Strategy

New, largely model-independent, analysis strategy for limits on:

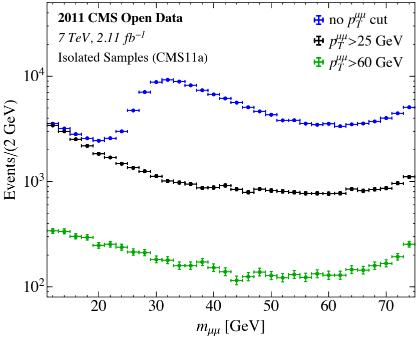
$$\frac{N_V}{\mathcal{L}} = \sigma\left(pp \to V + X\right)\mathcal{B}\left(V \to \mu^+\mu^-\right)A_V \epsilon_{\text{tr}}^V \epsilon_{\text{iso}}^V$$

Goal: Increase sensitivity by imposing $p_T^{\mu\mu}$ cuts.

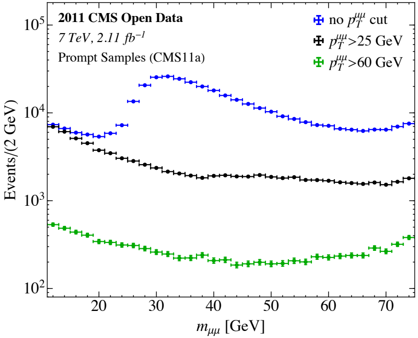
- Dimuon trigger ($p_T^{1,2} > 13, 8$ GeV)
- Only opposite-sign (OS) muons
- Define **isolated** and **prompt** samples
- Apply increasingly restrictive $p_T^{\mu\mu}$ cuts (substantial p_T^V)
- Look for a bump

Prompt vs. Isolation

Isolated Sample



Prompt Sample



Each $p_T^{\mu\mu}$ cut shrinks background by factor of ~ 6

NO P_T CUT

$P_T > 25$ GeV

$P_T > 60$ GeV

General Search Strategy

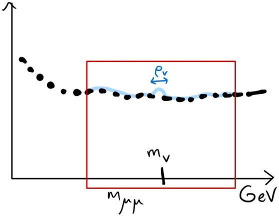
Cut flow for the $V \to \mu^+\mu^-$ search on CMS11a

	Dimuon Events	
Baseline Acceptance and Tight Muons Cuts ($p_{T,1} > 15$ GeV, $p_{T,2} > 10$ GeV, $\lvert\eta_\mu\rvert < 2.1$, $d_0 < 2$ mm, $z_0 < 10$ mm)	2,155,900	
Search Region (OS, $m_{\mu\mu} \in [11, 78]$ GeV, $d_0 < 250\,\mu$m, $z_0 < 2000\,\mu$m)	486,242	
	Isolated Sample ($I_{\rm comb} < 0.15$)	Prompt Sample ($IP_{xy} < 100\,\mu$m)
$p_T^{\mu\mu} > 0$	139,282	343,519
$p_T^{\mu\mu} > 25$ GeV	34,550	72,434
$p_T^{\mu\mu} > 60$ GeV	5,206	7,736

Resonance Search

Finding a resonance $\rightarrow$ 'bump hunting'

- Background fit with polynomial
- Center window around $m_{\mu\mu}$
- Width of window is 35 ρ_V spread over 140 bins
- Fit with fifth-order polynomial (x^0 to x^5)



Resonance Search

Place new bound on $pp \to V + X, V \to \mu\mu$

- 95% CL_s upper limits on

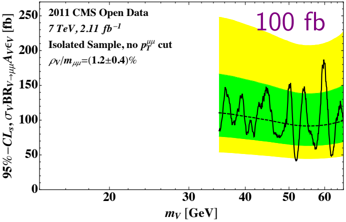
$$\sigma\left(pp \to V + X\right)\mathcal{B}\left(V \to \mu^+\mu^-\right)A_V\epsilon_{\mathrm{tr}}(\epsilon_{\mathrm{iso}})$$

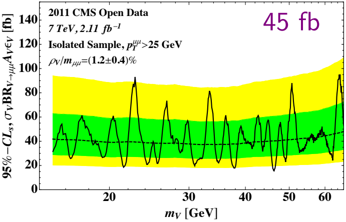
- Acceptance A_V and isolation ϵ_{iso} have some model dependence
- We only gain on cross section bound if we don't lose acceptance

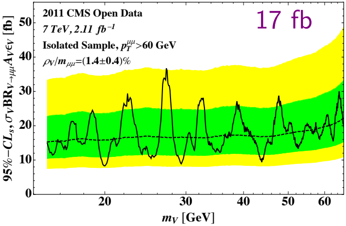
Resonance Search

95% CL$_s$ on $\sigma\,(pp \to V + X)\,\mathcal{B}\,(V \to \mu^+\mu^-)\,A_V\epsilon_{\text{tr}}\epsilon_{\text{iso}}$

Isolated Sample



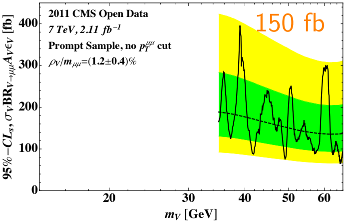


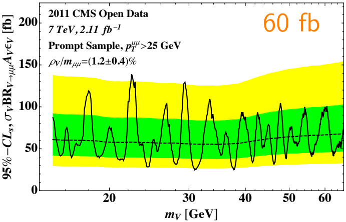


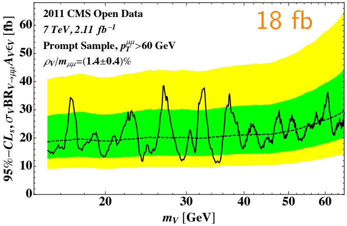
Resonance Search

95% CL$_s$ on $\sigma\left(pp \to V + X\right)\mathcal{B}\left(V \to \mu^+\mu^-\right)A_V\epsilon_{\mathbf{tr}}$

Prompt Sample







Resonance Search

Consider region $m_{\mu\mu} > 40$ GeV

$$\sigma\left(pp \to V + X\right) \mathcal{B}\left(V \to \mu^+\mu^-\right) A_V \epsilon_{\text{tr}} \epsilon_{\text{iso}}$$

$p_T^{\mu\mu}$ Cut	Isolated Upper Lim.	Prompt Upper Lim.
No $p_T^{\mu\mu}$ cut	100 fb	150 fb
$p_T^{\mu\mu} > 25$ GeV	45 fb (2x)	60 fb (2.5x)
$p_T^{\mu\mu} > 60$ GeV	17 fb (6x)	18 fb (8x)

Resonance Search

Consider region $m_{\mu\mu} > 40$ GeV

$$\sigma\left(pp \to V + X\right) \mathcal{B}\left(V \to \mu^+\mu^-\right) A_V \epsilon_{\text{tr}} \epsilon_{\text{iso}}$$

$p_T^{\mu\mu}$ Cut	Isolated Upper Lim.	Prompt Upper Lim.
No $p_T^{\mu\mu}$ cut	100 fb	150 fb
$p_T^{\mu\mu} > 25$ GeV	45 fb (2x)	60 fb (2.5x)
$p_T^{\mu\mu} > 60$ GeV	17 fb (6x)	18 fb (8x)

p_T cuts enhance sensitivity if A_V does not drop too rapidly!

Benchmark Scenarios

What models will retain acceptance across $p_T^{\mu\mu}$ cuts?

Consider models with scalar S and vector V

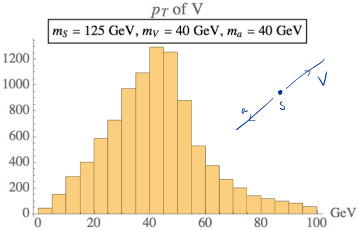
- Final state: $\mu^+\mu^- +$ jets $+$ no missing p_T
- No particle clusters near signal $\rightarrow$ use isolation cuts

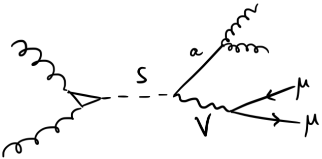
$$pp \rightarrow S \rightarrow V + a$$

a pseudoscalar

$$a \rightarrow gg, \quad b\bar{b}$$

$$V \rightarrow \mu^+\mu^-$$





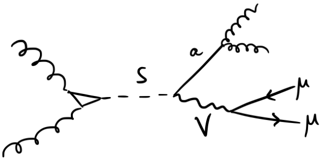
Benchmark Scenarios

In some regions of $m_{\mu\mu}$, our search strategy cut strengthens cross section limit by factor of $\gtrsim 2$

Consider $m_S = m_h = 125$ GeV, and $m_V = m_a = 40$ GeV

$A_V(0) = 54\%$, $A_V(25) = 47\%$, $\rightarrow A_V(0)/A_V(25) \sim 85\%$ (from Pythia8)
$\epsilon_{\text{tr}} \sim 85\%$, $\epsilon_{\text{iso}} \sim 85\%$ (from Open Data)

$$\mathcal{B}\left(h \to Va\right)\mathcal{B}\left(V \to \mu\mu\right) \lesssim 8 \times 10^{-3}$$



Benchmark Scenarios

Our search sets competitive if not stronger bounds on certain classes of models

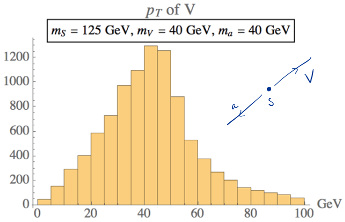
Improve limits in comparison to inclusive searches on Run I

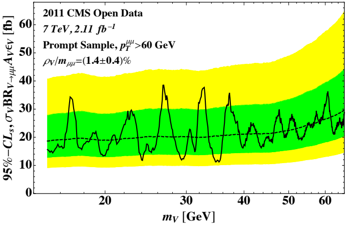
Conclusions

- Performed largely **model-independent** analysis for $V \to \mu^+\mu^-$
- Improve sensitivity on many models by exploiting moderately boosted kinematics

  $\to$ ATLAS & CMS can improve limits by an **order of magnitude** using Run II
- Prompt limits improve sensitivity to certain classes of BSM models

  $\to$ Analysis need knowledge of QCD background, impossible without Open Data





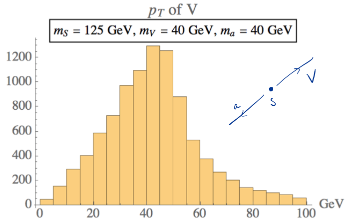
Conclusions

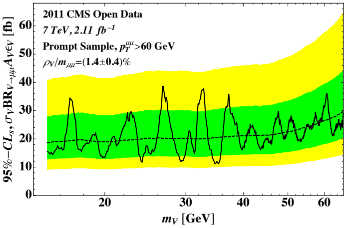
- Performed largely **model-independent** analysis for $V \to \mu^+\mu^-$
- Improve sensitivity on many models by exploiting moderately boosted kinematics

  $\to$ ATLAS & CMS can improve limits by an **order of magnitude** using Run II
- Prompt limits improve sensitivity to certain classes of BSM models

  $\to$ Analysis need knowledge of QCD background, impossible without Open Data

Open Data can be used to test BSM analysis strategies!

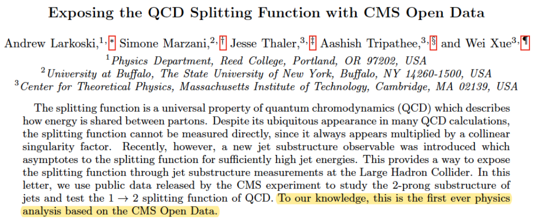




Back-ups

Open Data

Lots of utility in Open Data



Exposing the QCD Splitting Function with CMS Open Data

Andrew Larkoski,[1,*] Simone Marzani,[2,†] Jesse Thaler,[3,‡] Aashish Tripathee,[3,§] and Wei Xue[3,¶]
[1] *Physics Department, Reed College, Portland, OR 97202, USA*
[2] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*
[3] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

The splitting function is a universal property of quantum chromodynamics (QCD) which describes how energy is shared between partons. Despite its ubiquitous appearance in many QCD calculations, the splitting function cannot be measured directly, since it always appears multiplied by a collinear singularity factor. Recently, however, a new jet substructure observable was introduced which asymptotes to the splitting function for sufficiently high jet energies. This provides a way to expose the splitting function through jet substructure measurements at the Large Hadron Collider. In this letter, we use public data released by the CMS experiment to study the 2-prong substructure of jets and test the $1 \to 2$ splitting function of QCD. To our knowledge, this is the first ever physics analysis based on the CMS Open Data.

- Tests long-term storage framework of data
- Source of SM events poorly modeled by MC
  $\to$ QCD Splitting function (1704.05066)
- Study detector effects on new search strategies
- Search for Beyond-the-Standard-Model (BSM) phenomena

Open Data

Lots of utility in Open Data

Exposing the QCD Splitting Function with CMS Open Data

Andrew Larkoski,[1,*] Simone Marzani,[2,†] Jesse Thaler,[3,‡] Aashish Tripathee,[3,§] and Wei Xue[3,¶]
[1] *Physics Department, Reed College, Portland, OR 97202, USA*
[2] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*
[3] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

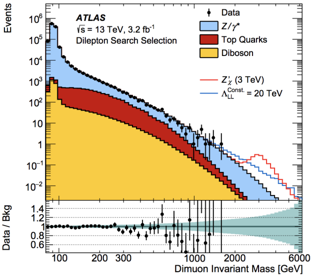
The splitting function is a universal property of quantum chromodynamics (QCD) which describes how energy is shared between partons. Despite its ubiquitous appearance in many QCD calculations, the splitting function cannot be measured directly, since it always appears multiplied by a collinear singularity factor. Recently, however, a new jet substructure observable was introduced which asymptotes to the splitting function for sufficiently high jet energies. This provides a way to expose the splitting function through jet substructure measurements at the Large Hadron Collider. In this letter, we use public data released by the CMS experiment to study the 2-prong substructure of jets and test the $1 \to 2$ splitting function of QCD. To our knowledge, this is the first ever physics analysis based on the CMS Open Data.

- Tests long-term storage framework of data
- Source of SM events poorly modeled by MC
  $\to$ QCD Splitting function (1704.05066)
- Study detector effects on new search strategies
- Search for Beyond-the-Standard-Model (BSM) phenomena

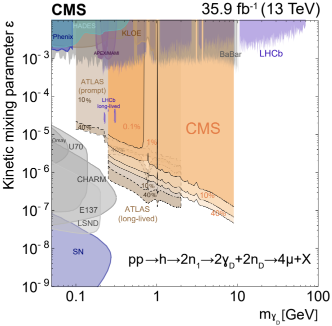
New Physics at LHC

Previous searches at ATLAS & CMS with dimuons: $m_{\mu\mu}$ resonances

- High-mass: $m_{Z'} \in [120\ \text{GeV}, 3.36\ \text{TeV}]$
- Pairs of light bosons: $m_{\gamma_D} \in [0.25, 8.5]$ GeV



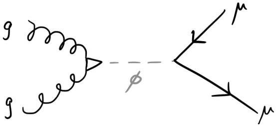
ATLAS, `arXiv:1607.03669`



CMS, `arXiv:1812.00380`

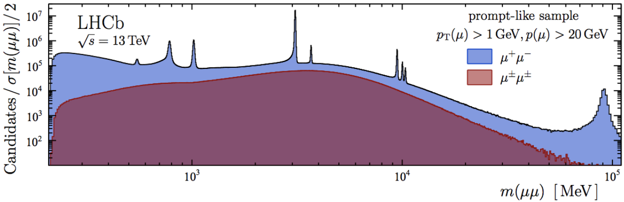
Previous Searches at LHC

LHCb searched for light new physics from direct production in *pp* collisions



- $m_\phi < 70$ GeV
- Excellent mass resolution, low p_T^μ trigger
- Inclusive search

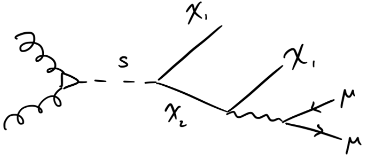
LHCb, `arXiv:1710.02867`



Model 2

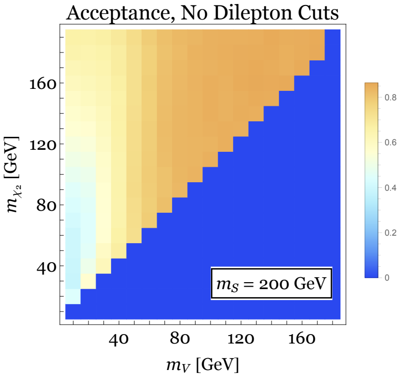
$$S \to \chi_1\chi_2, \qquad \chi_2 \to \chi_1 V$$

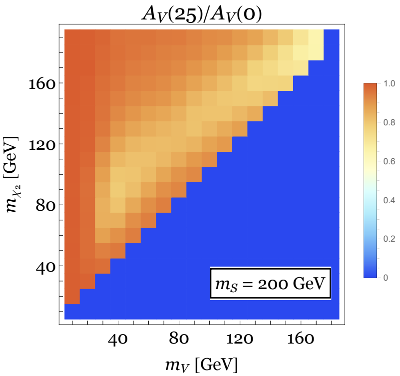
- Isolation or prompt cut depends on mass hierarchy
- For $\chi_2 \ll m_S$, χ_2 boosted, decay products are collimated
  $\to$ Ruin isolation, use prompt



Model 2

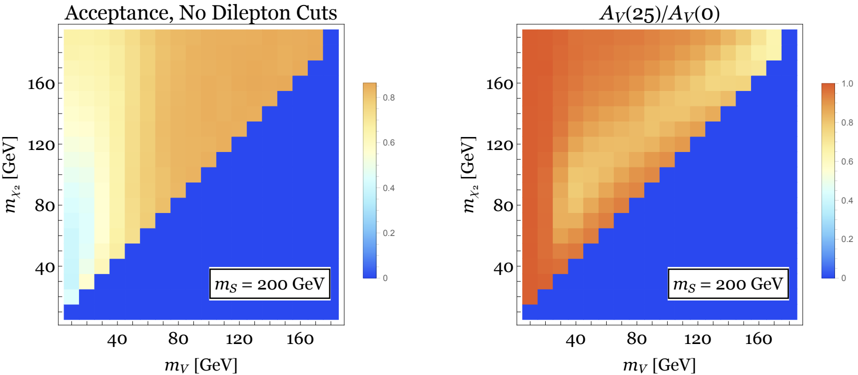
$$m_S = 200\ \text{GeV},\ m_{\chi_1} = 10\ \text{GeV}$$





Model 2

$m_S = 200$ GeV, $m_{\chi_1} = 10$ GeV



Acceptance ratio $\gg 50\%$ for most of kinematic range

Good candidate for p_T enhanced search

Validation Studies

All CMS11a results compared to CMS10 reported values

Trigger Efficiency

- Measured from MC truth vs. MC reconstructed
- Compared against CMS public 2011 muon results

Isolation Efficiency

- Compare tag & probe results from MC and Open Data
- Agree within 1%

Kinematic Acceptance

- Cuts listed previously
- Detector depedent

Resonance Search

Fitting Disclaimer

- Muon p_T resolution is a function of η, p_T
  $\rightarrow$ dimuon mass resolution is also a function of η, p_T
  $\rightarrow$ Resolution ρ_V is dependent on V kinematics
- Bump has non-Gaussian tails
  $\rightarrow$ Photon emission for $m_{\mu\mu} < m_V$
  $\rightarrow$ Mismeasuring, scattering
  $\rightarrow$ Gaussian smearing is function of η

Correct for these effects on bounds

Resonance Search

Why study this range and $p_T^{\mu\mu}$ cuts?

- At low mass m_V and/or large $p_T^{\mu\mu}$, muons are too collimated and triggering begins to fail
- At high mass m_V and/or large $p_T^{\mu\mu}$, fewer event counts $\rightarrow$ larger uncertainty

Resonance Search

Why study this range and $p_T^{\mu\mu}$ cuts?

- At low mass m_V and/or large $p_T^{\mu\mu}$, muons are too collimated and triggering begins to fail
- At high mass m_V and/or large $p_T^{\mu\mu}$, fewer event counts $\rightarrow$ larger uncertainty

Experimentalists with better trigger/reco knowledge and larger data set could extend kinematic regions!