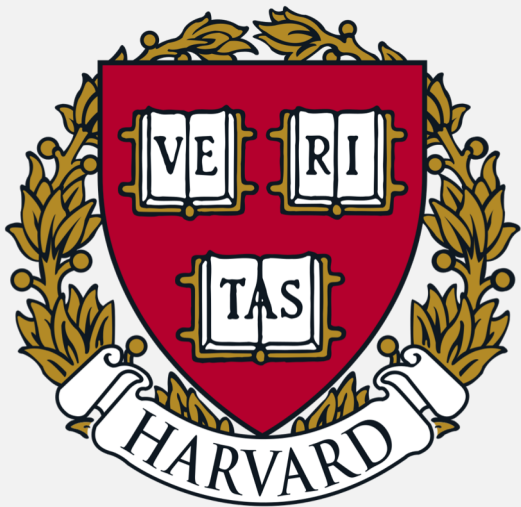# OPEN DATA & BSM PHYSICS

Cari Cesarotti
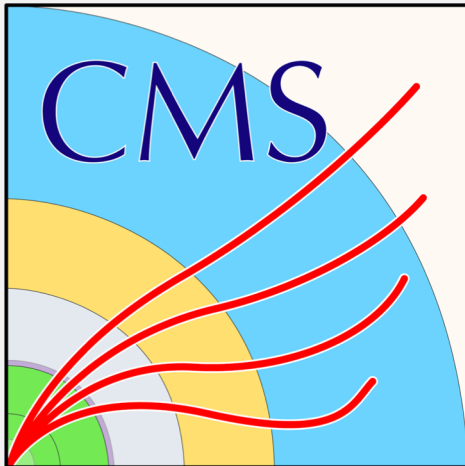
Harvard University

Snowmass Computational Frontier Workshop

August 11, 2020

# CMS OPEN DATA IS RESEARCH-GRADE COLLIDER DATA MADE PUBLIC

http://opendata.cern.ch/

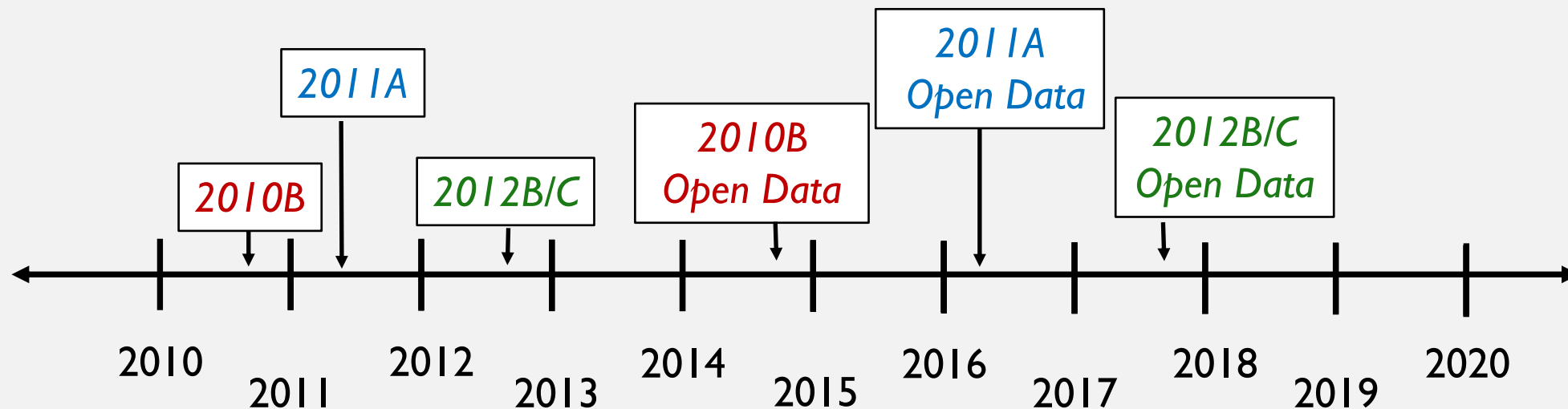# WHAT'S IN THE OPEN DATA?

Ready-to-analyze data is stored in *Primary Datasets:*

- High-level, reconstructed events
- Specified event selection
- Corresponding MC*

    (essential for understanding detector effects & acceptance)

Additionally, the Open Data portal provides derived datasets,

analysis tools, & instructions
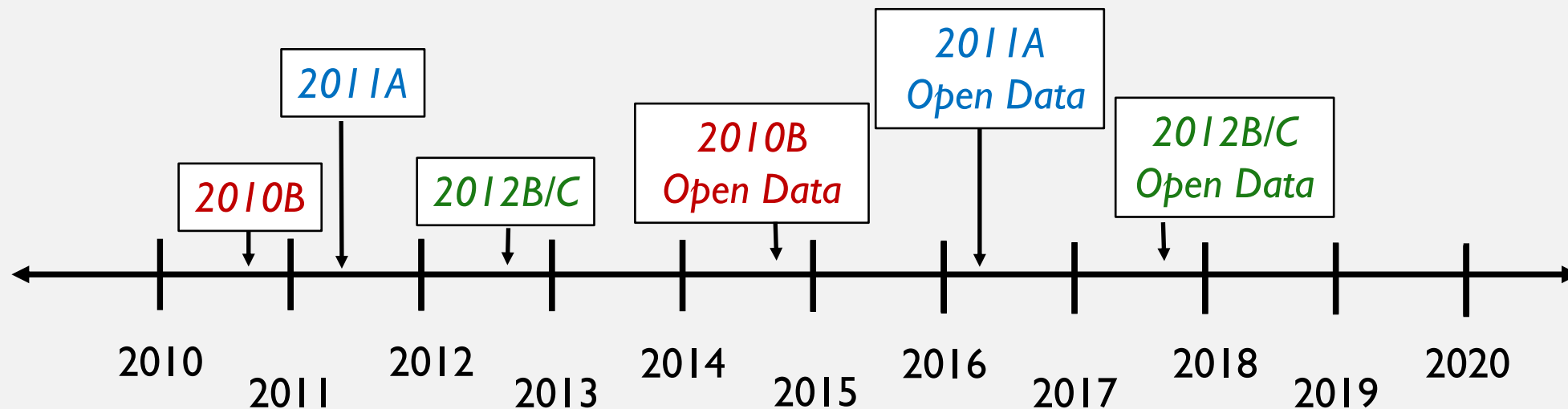
# TIMELINE OF DATA COLLECTION & RELEASE



November 2014:
Run 2010B
7 TeV, 32 pb⁻¹
No MC

April 2016:
Run 2011A
7 TeV, 2.5 fb⁻¹
MC

December 2017:
Run 2012B/C
8 TeV, 11.6 fb⁻¹
MC

# TIMELINE OF DATA COLLECTION & RELEASE



**2011A**

**2010B**

**2012B/C**

**2010B**
*Open Data*

**2011A**
*Open Data*

**2012B/C**
*Open Data*

2010    2011    2012    2013    2014    2015    2016    2017    2018    2019    2020

**November 2014:**
Run 2010B
7 TeV, 32 pb⁻¹
No MC

**April 2016:**
Run 2011A
7 TeV, 2.5 fb⁻¹
MC

**December 2017:**
Run 2012B/C
8 TeV, 11.6 fb⁻¹
MC

## ROUGHLY 5 YEAR DELAY

# SHOULD WE CARE ABOUT 5 YEAR OLD DATA?

## SHOULD WE CARE ABOUT 5 YEAR OLD DATA?

# YES

# WE SHOULD CARE ABOUT *N* YEAR OLD DATA

The goal of particle theory is to understand and predict the fundamental laws of nature.

What better way to do that than directly studying what nature gives us?

# OPEN DATA & THEORY

## QCD splitting function

(A. Tripathee, W. Xue, A. Larkoski, S. Marzani, J. Thaler '17)

## Jet Substructure

(A. Larkoski, S. Marzani, J. Thaler, A. Tripathee, W. Xue '17)

(P.T. Komiske, R. Mastandrea, E. M. Metodieve, P. Naik, J. Thaler '19)

## Machine Learning w/ Open Data MC

(C. F. Madrazo, I. H. Cacha, L. L. Iglesias, J. M. de Lucas '17)

(M. Andrews, M. Paulini, S. Gleyzer, B. Poczos '17)

## ML Anomaly Detection

(O. Knapp, G. Dissertori, O. Cerri, T. Q. Nguyen, J. Vlimant, M. Pierini '20)

## SM Cross Section Measurements

(A. Apyan, W. Cuozzo, M. Klute, Y. Saito, M. Schott, B. Sintayehu '19)

## Parity violation in Jets

(C. G. Lester, M. Schott '19)

## BSM Searches

(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue '19)

# OPEN DATA & THEORY

## QCD splitting function

(A. Tripathee, W. Xue, A. Larkoski, S. Marzani, J. Thaler '17)

## Jet Substructure

(A. Larkoski, S. Marzani, J. Thaler, A. Tripathee, W. Xue '17)

(P.T. Komiske, R. Mastandrea, E. M. Metodieve, P. Naik, J. Thaler '19)

## Machine Learning w/ Open Data MC

(C. F. Madrazo, I. H. Cacha, L. L. Iglesias, J. M. de Lucas '17)

(M. Andrews, M. Paulini, S. Gleyzer, B. Poczos '17)

## ML Anomaly Detection

(O. Knapp, G. Dissertori, O. Cerri, T. Q. Nguyen, J. Vlimant, M. Pierini '20)

## SM Cross Section Measurements

(A. Apyan, W. Cuozzo, M. Klute, Y. Saito, M. Schott, B. Sintayehu '19)

## Parity violation in Jets

(C. G. Lester, M. Schott '19)

## BSM Searches

(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue '19)

*IF THE DATA HAS BEEN OUT FOR >5 YEARS, SHOULDN'T WE HAVE ALREADY SEEN INDICATIONS OF NEW PHYSICS?*

Well, that depends

# THEORIST* PERSPECTIVE ON OPEN DATA AND BSM SEARCHES

*MY

Experimentalists will make the discovery, but theorists can guide them there.

To theorists, Open Data should be seen as a means to test feasibility of proposed new physics searches.
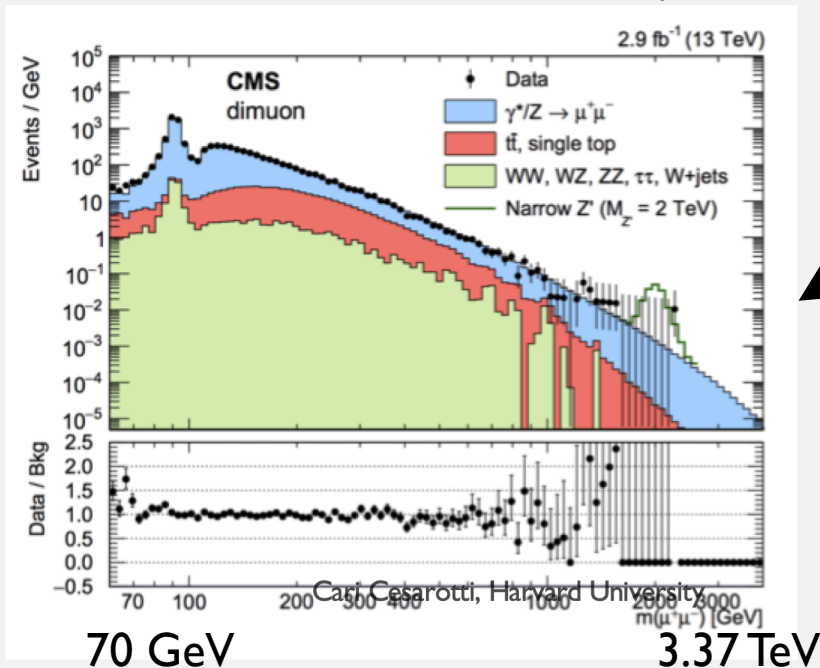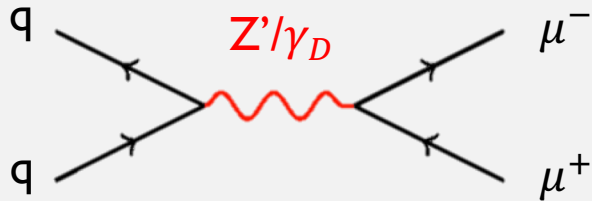
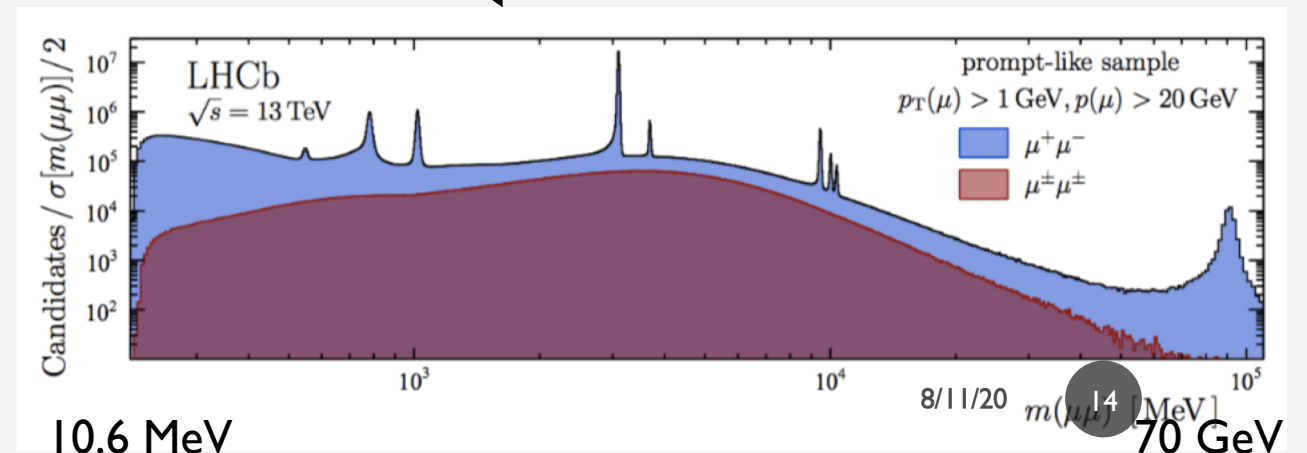# EXAMPLE: BSM SEARCH FOR DIMUON RESONANCES

(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue)

Seems likes an extremely well explored regime,
especially on a decade-old data set ($\sqrt{s} = 7$ TeV)…
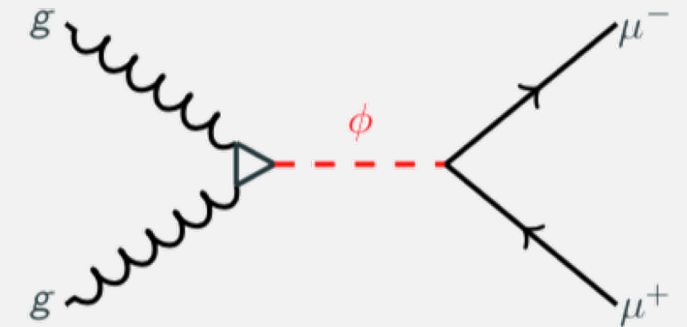
# EXAMPLE: BSM SEARCH FOR DIMUON RESONANCES

(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue)

Seems likes an extremely well explored regime,
especially on a decade-old data set ($\sqrt{s} = 7$ TeV)…



CMS 2016,
High Mass

LHCb 2017,
Low Mass

70 GeV                    3.37 TeV

10.6 MeV                    70 GeV
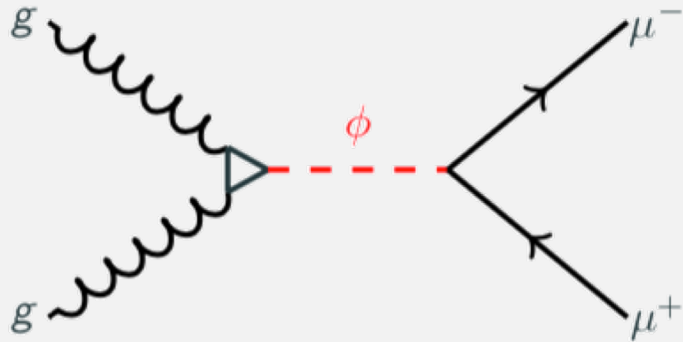
Carl Cesarotti, Harvard University

8/11/20

14

# EXAMPLE: BSM SEARCH FOR DIMUON RESONANCES
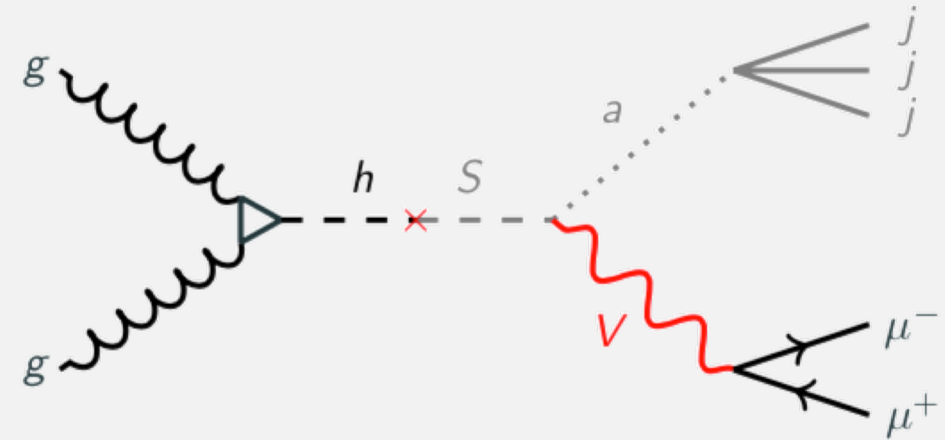
(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue)

…but is an excellent playground for theorists
proposing a new semi-inclusive search!

**Previous searches were** inclusive, direct production



→ $\phi$ produced with small $p_T$

**Propose search for** indirect production
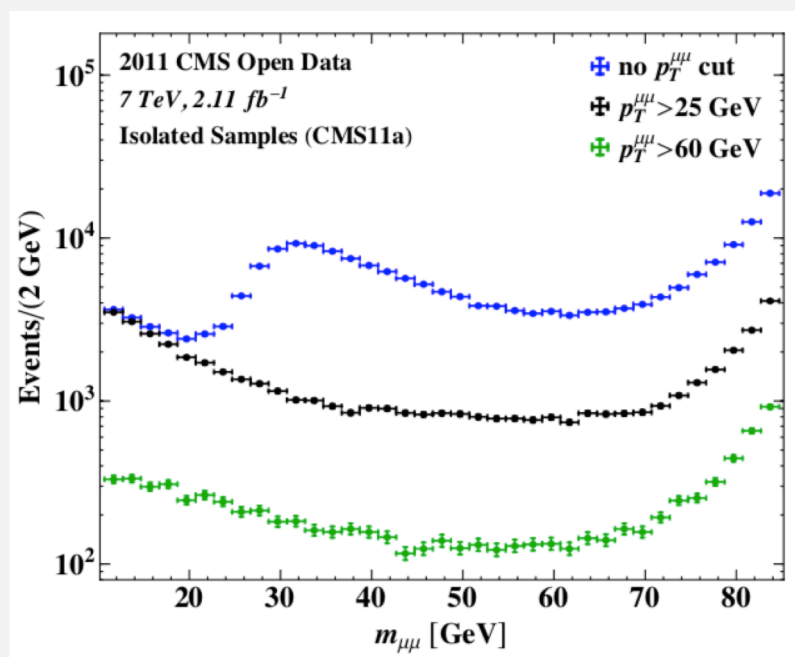


→ *V* produced with substantial $p_T$
→ Cut on $p_T$ of dimuon system, *then* bump hunt
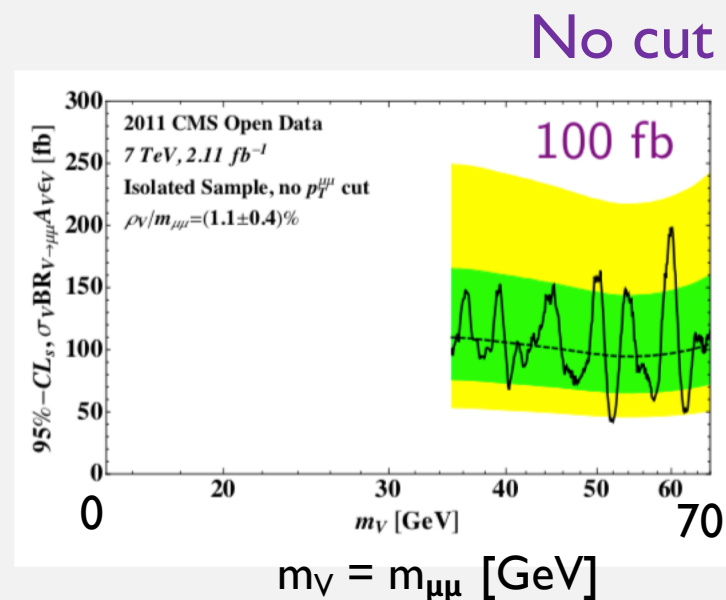→ Previously unexplored kinematic regime

# EXAMPLE: BSM SEARCH FOR DIMUON RESONANCES

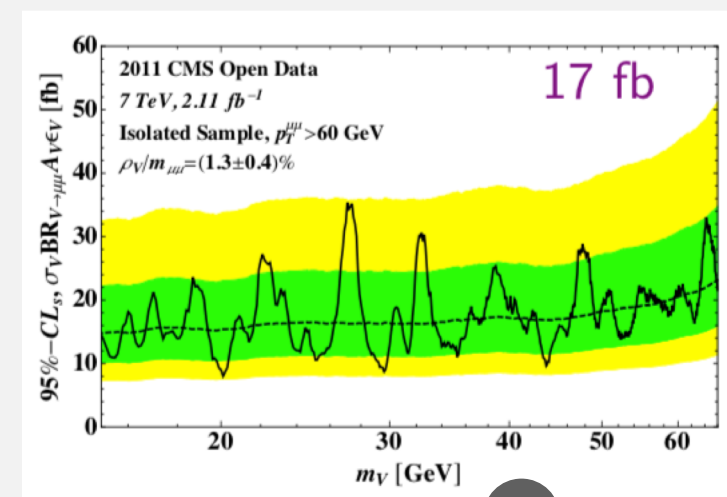(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue)

Using Open Data 2011A, we were able to illustrate how exploiting boosted kinematics can greatly enhance sensitivity



No cut

$P_T^{\mu\mu} > 60$ GeV

$P_T$ cuts reduce background by factor of 6

# EXAMPLE: BSM SEARCH FOR DIMUON RESONANCES

(CC, Y. Soreq, M. J. Strassler, J. Thaler, W. Xue)

Note that this study:

1. Provides a model independent bound

2. Would set tighter bounds with Run II data

3. Is absolutely feasible to experimentalists to implement

# WITH GREAT POWER COMES GREAT RESPONSIBILITY

Maintaining, formatting, distributing, etc. Open Data is nontrivial, so we as theorists should provide compelling reasons for CERN to do it!

Designing analysis, testing detector efficiencies, calculating acceptances is extremely nontrivial! Approach Open Data with diligence and respect.

Our goal as theorists is not to discover new physics in open data[*], but to inform future experimental efforts to facilitate the discovery of new physics.

*But it wouldn't be impossible!*

# WITH GREAT POWER COMES GREAT RESPONSIBILITY

Maintaining, formatting, distributing, etc. Open Data is nontrivial, so we as theorists should provide compelling reasons for CERN to do it!

Designing analysis, testing detector efficiencies, calculating acceptances is extremely nontrivial! Approach Open Data with diligence and respect.

Our goal as theorists is not to discover new physics in open data[*], but to inform future experimental efforts to facilitate the discovery of new physics.

*But it wouldn't be impossible!

## Thank you!