CMS Analyses in the 10%

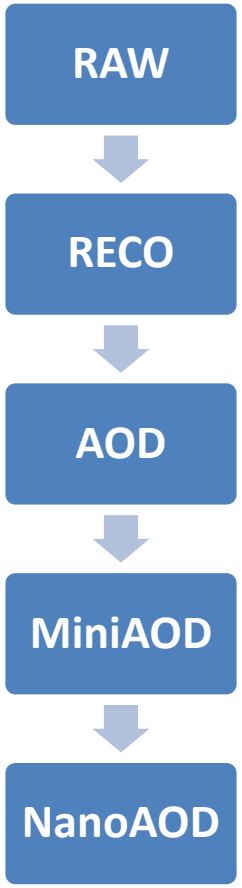
Juliette Alimena (The Ohio State University)

HSF DAWG Meeting

July 15, 2020

CMS Data Formats

Main event data formats for standard pp collision runs:



~1 MB/ev

~3 MB/ev

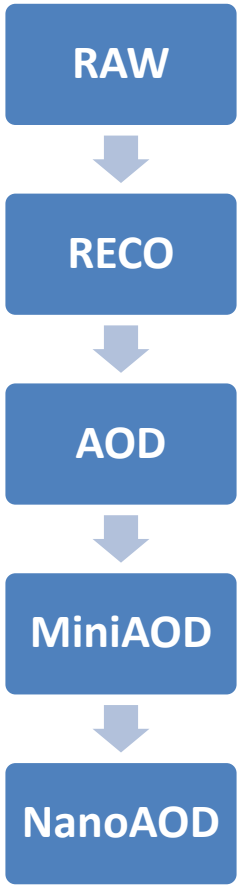
~500 kB/ev

~50 kB/ev → **Basis for most run 2 user ntuples**

~2 kB/ev → **Increasingly being used by physics analysis groups and physics object groups**

Adapted from M. Peruzzi (ACAT 2019) and J. Pata (HSF March 2020)

The ~10%



~1 MB/ev

~3 MB/ev

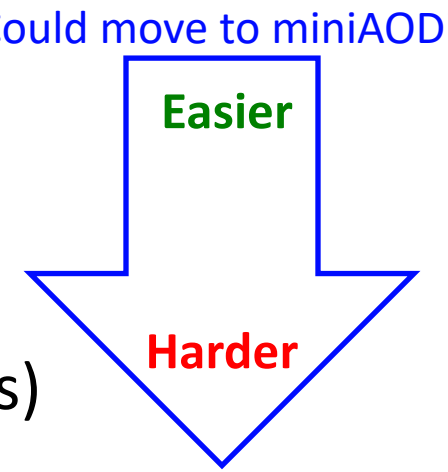
~500 kB/ev

~50 kB/ev

~2 kB/ev

Analyses that need AOD, RECO, or even RAW:

- Heavy ions
- B-physics
- **Long-lived exotic particles**
- “Data scouting” (trigger-level analyses)



Event Content for Heavy Ions

- Currently using AOD for PbPb data and MC
  - HI data now a sizeable contribution to AOD
  - Remote reads to one Tier 2 can saturate network bandwidth
  - Dataset size limits replication
- Moving to miniAOD for Run 3
  - First step: reMiniAOD 2018 PbPb data, to clear up disk resources
  - Modifications of MiniAOD content and algorithms underway
- Focus on PbPb, any customization needed for pp reference and pPb will be modest

Moving to MiniAOD for HI

- Additional collections:
  - Low pT pixel tracks
  - HI-specific info: centrality and reaction plane
  - Event selection filters
  - Key information for general tracks, including 4-vector info with reduced precision
- Additional variables:
  - For tracker muons (soft ID)
  - For cut-based track selection
  - More precise chisq/dof

- Porting algorithms:
  - Different jet underlying event subtraction
  - HI-style photon isolation
- Threshold changes:
  - Change low pT muon selection
  - Change Egamma H/E, pT selections
- Migration from forward hadron calorimeter tower-based variables to particle flow candidates for centrality and reaction plane

Final miniAOD size will still be ~10x smaller than AOD

Event Content for B-Physics

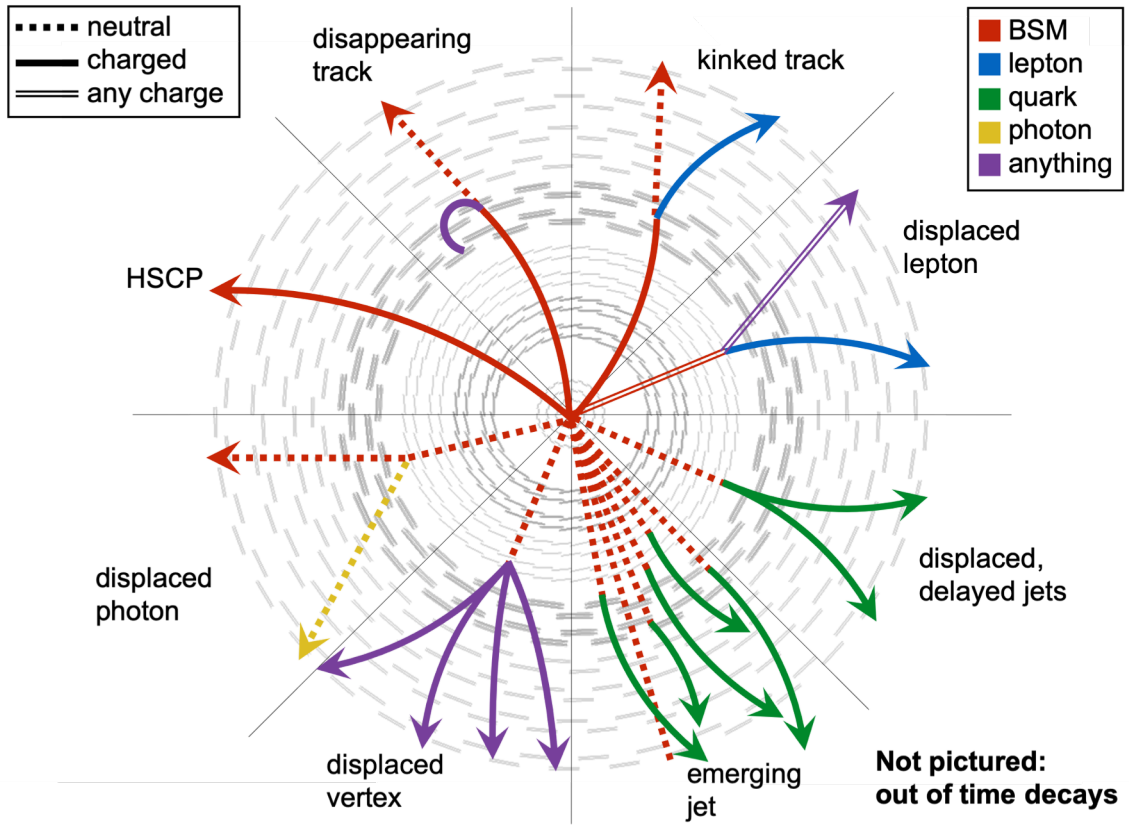
- In most of the published analyses AOD has been used
  - Because low pT tracks (pt between 0.55 and 0.95 GeV) are included in miniAOD but without full parameter set including covariance matrix

1. Analyses requiring **very soft tracks from pions** (pt < 0.5 GeV)
   - These tracks are included in miniAOD, hard to include all of them due to computing time/memory limitations. Also, they would need to be refitted
2. Analysis that uses AOD to do the **primary vertex refit**, to improve the mass resolution
   - Not all of the low pT tracks used to fit the PV are included in miniAOD

- Example: **excited Lambda-b baryons** (*Phys. Lett. B 803 (2020) 135345*)
  - The signal efficiency **decreases by a factor of ~2** when requiring pt(pion)>0.55 GeV
  - The significance of the Lambda-b(5912)0 peak in the data is **5.7σ**, thanks to AOD usage. If only tracks with pt>0.55 GeV were used, the significance would be lower than 5σ.
  - Also refit the primary vertex → **improved the mass resolution by ~50%**
- **For ongoing analyses, currently working on:**
  - Get around these effects with the current event content
  - Investigate adding to miniAOD, collaborating with those working on heavy ions

LLP Searches at CMS

- Long-lived particles: BSM particles with measurable lifetime
- Wide variety of LLP signatures and strategies
- Often require **unusual and innovative techniques**
- Some challenges:
  - Dedicated triggers
  - Atypical backgrounds
  - Unique object reconstruction
  - Unusual discriminating variables



Often not in mini or nanoAOD!

Event Content for LLP Analyses

- The vast majority of LL analyses currently rely on AOD
- Only 2 or 3 LL analyses (out of ~25) can use miniAOD
- Some analyses even have dedicated skims or special needs (see next slides)

Disappearing Tracks

- Disappearing tracks analysis (arxiv:2004.05153) done with miniAOD+ a few collections in AOD
- **Additional content needed on top of miniAOD:**
  - generalTrack-based isolation sum
  - E_{calo} (sum of all calo rechits nearby candidate track)
- Improvements in progress for Run 3:
  - **Adding these two variables as floats in miniAOD**
  - Determining if cuts applied to save dE/dx info for are ok for this analysis
  - Plan to make proposal for additions to miniAOD soon
- Need to **access some (not all) RAW data:**
  - Access electrons for tag-and-probe study with pt<50 GeV, and all rechit info available
  - Some rechit info dropped even at AOD level
  - Need to accurately reconstruct the whole event, or will overestimate electron background
  - Only need to access ~1k events per year of data taking
  - **Will probably need to continue in this manner for Run 3**

Stopped Particles

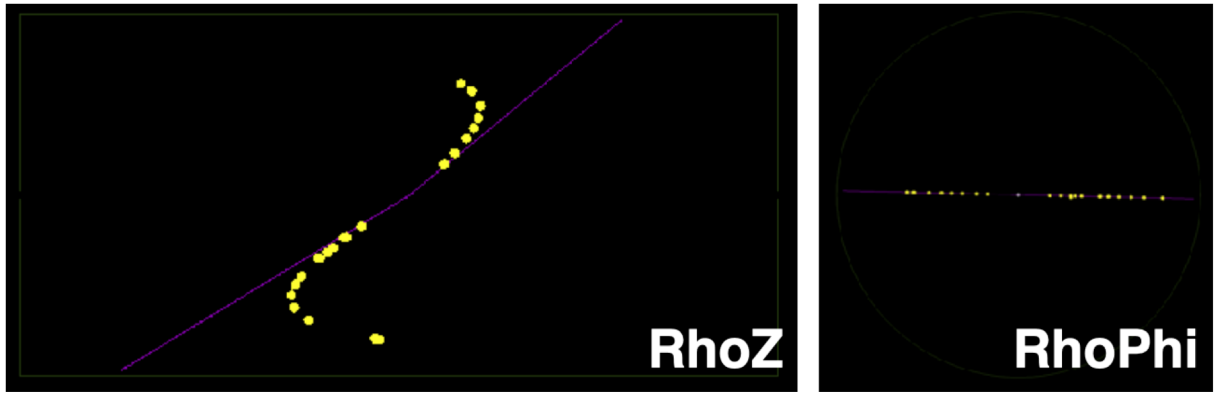
- Stopped Particles (10.1007/JHEP05(2018)127) searched for long-lived exotic particles that stop in the detector and decay to jets or muons
- Uses the NoBPTX dataset
  - Only 8 HLT paths (4 unprescaled), all of which veto on the collision BX and +/-1 BX
  - This dataset contains cosmics, beam halo, and noise events (and LL signal?)
  - 2016 RECO dataset was 20 TB (as opposed to say, SingleMuon, which was 1150 TB)
- Skim idea:
  - **Skim on NoBPTX dataset, keep only the event content needed for this analysis, but make no event selections**
    - Calo jets
    - Hcal noise variables
    - CSC and DT segments
    - RPC hits
    - Displaced standalone muons
  - → **~30% of RECO** (2017 NoBPTX RECO data)
  - Aim to have skim for Run 3 also

Displaced Muons

- Several LL searches involving muons use the reco::Track collection **displacedStandAloneMuons**, which is not available in miniAOD
- displacedStandAloneMuons are tracks made from only hits in the muon system and do not use the beamspot constraint in the track fit
  - Borrows seeds from cosmic muon reconstruction
  - Most appropriate collection for most displaced muons
- Used by searches for displaced dimuons, inelastic DM, etc.
- However, the muon POG has plans to improve displaced muon reconstruction in general for Run 3, and making displacedStandAloneMuons a reco::Muon available in miniAOD is on the list

Monopole Reconstruction

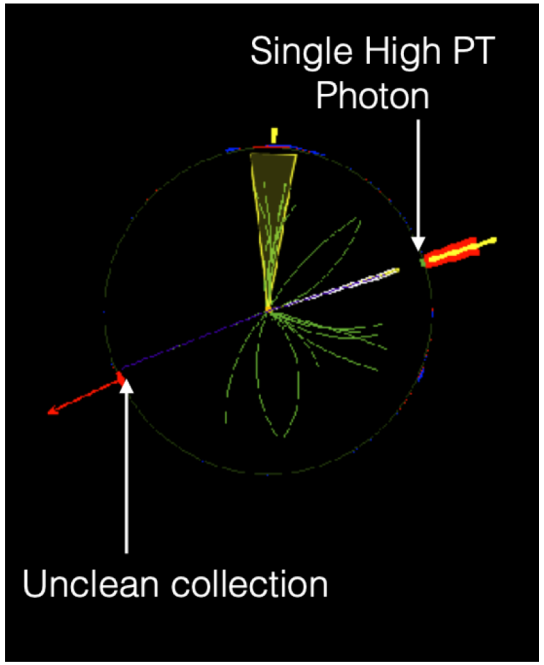
- The search for monopoles requires a **dedicated reconstruction**
- Two key pieces:
  - **Tracking:**
    - The monopole would experience a force along the magnetic field → track curves toward the ± z direction
    - The usual charged particles curve in the xy direction



  - **Electromagnetic calorimeter (Ecal) clusters:**
    - The monopole energy loss is dominated by ionization in the Ecal
    - The cluster shape will track the monopole through the Ecal

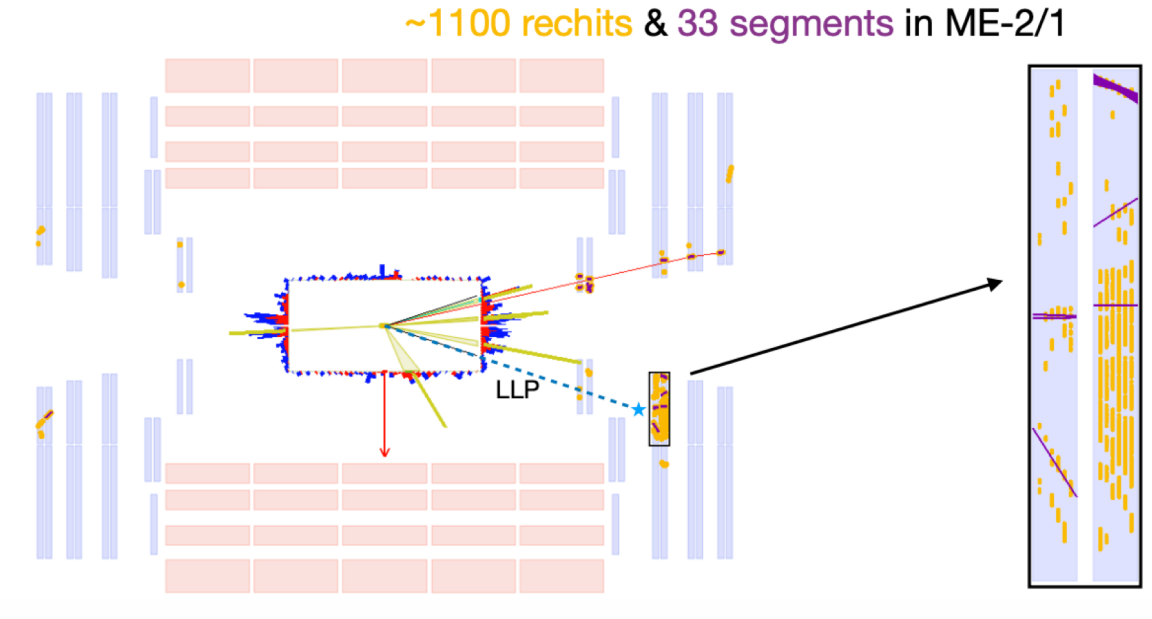
Monopole Skim

- **Additional collections needed on top of AOD:**
  - Track segments
    - In order to reconstruct the unusual monopole tracks
  - Both cleaned and uncleaned Ecal cluster collections
    - in order to avoid spike rejection filters
- Run 2 skim was run over **SinglePhoton** and **MET** datasets
- **Filter on unprescaled single photon and MET triggers**
- Input data format: RECO
- Output data format: AOD + additional collections
- Aim to have skim for Run 3 also



Hadronic Decays in the Muon System

- Neutral LLPs that decay beyond the calorimeters can leave a signature with just a **large cluster of rec hits (>200) in the muon system**
- Need RPC, DT, and CSC hits
- DT and CSC segment reconstruction eliminates signal
  - Too many hits in the same chambers not pointing at the IP



Hadronic Decays in the Muon System

- Currently a few options to do the analysis:
  - **Current plan for Run 2 analysis:** Use the MET > 200 GeV RAW-RECO skim to do a MET-based analysis
  - **Possible idea for Run 2 or 3:** Skim single/double lepton (+ potentially other) datasets filtering on the number of CSC/DT hits in the event
  - **Another option:** Add a new object to AOD that is a cluster of CSC/DT hits
- But also, essentially no background MC in RECO format
  - Could apply a skim on background MC?
  - Background MC may not be very useful anyways for this analysis, depending on which backgrounds end up mattering the most

What is Data Scouting?

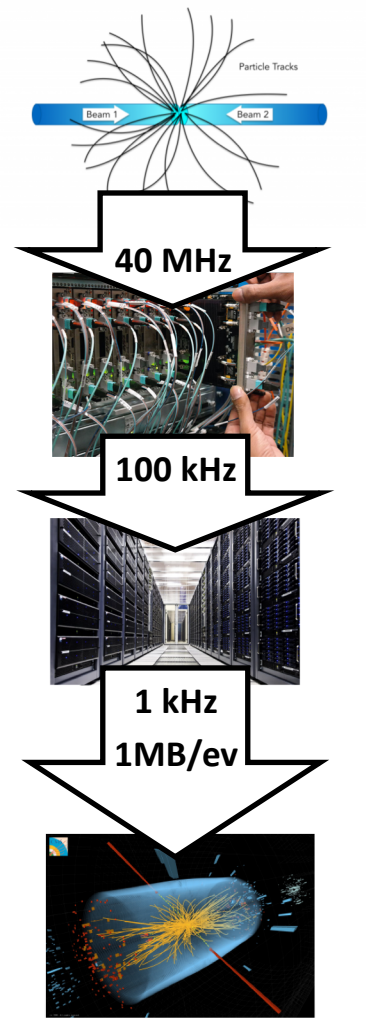
From Collisions to Analysis

LHC

Level 1 trigger

High Level Trigger

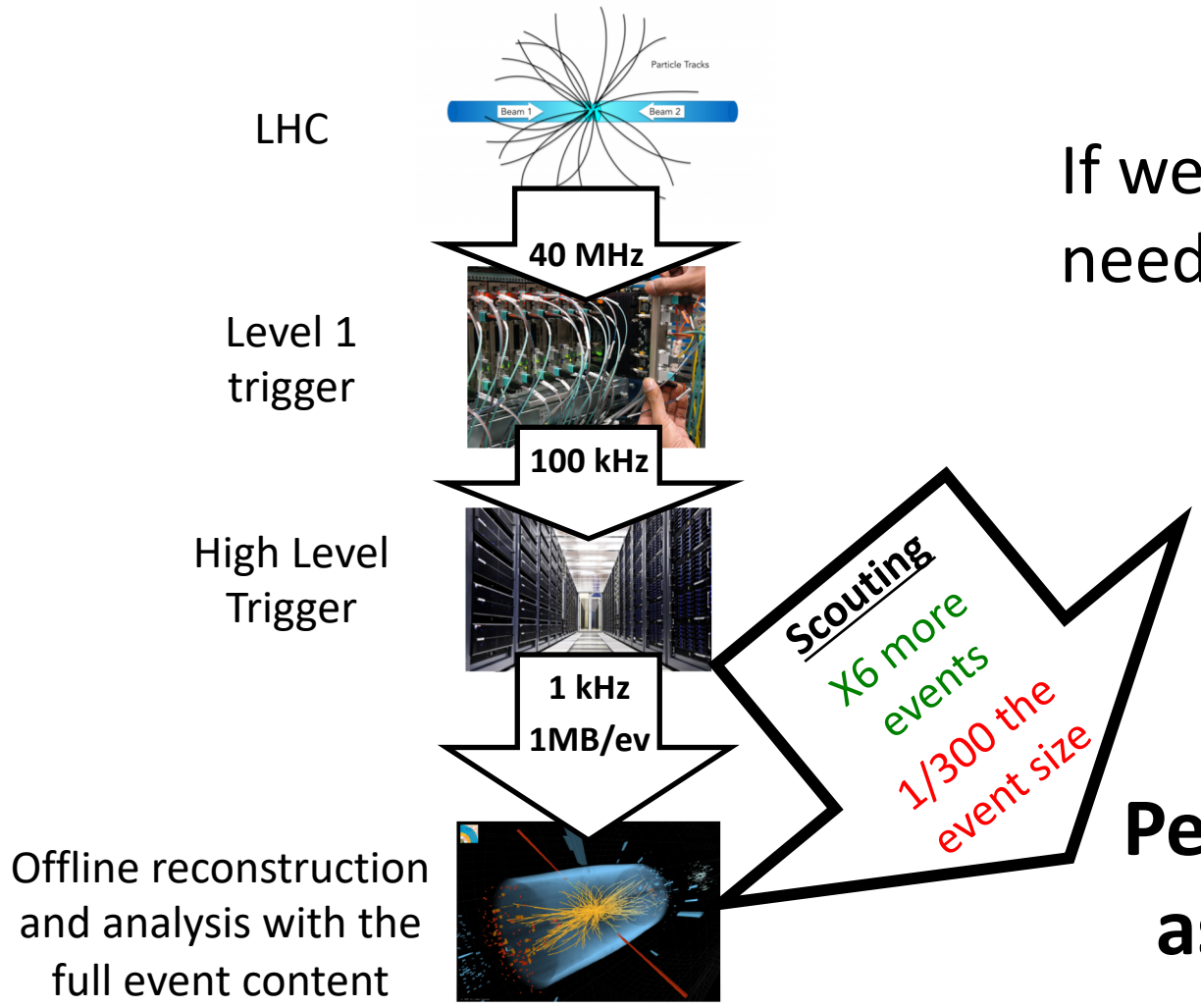
Offline reconstruction and analysis with the full event content



- The LHC delivers a huge amount of data, and we need to filter it in real time: trigger
- Generally store events with high p_T objects, but triggers have theory/pheno bias
- Huge reduction in rate: might be losing good events
- Actual limitation is trigger bandwidth = event rate X event size

What is Data Scouting?

Main idea: do more with less



If we want to increase the rate, we need to decrease the event size

Perform analysis with objects as reconstructed by the HLT

Data Scouting Event Content

Two scouting streams in Run 2:

- Calo Scouting
  - Vertices
  - Muons
  - Calo-based jets
  - Calo-based MET
- Particle Flow Scouting
  - Vertices
  - PF muons
  - PF jets
  - PF MET

Many ideas being worked on for Run 3:

- Photons
- Displaced particles
- GPU and machine learning developments

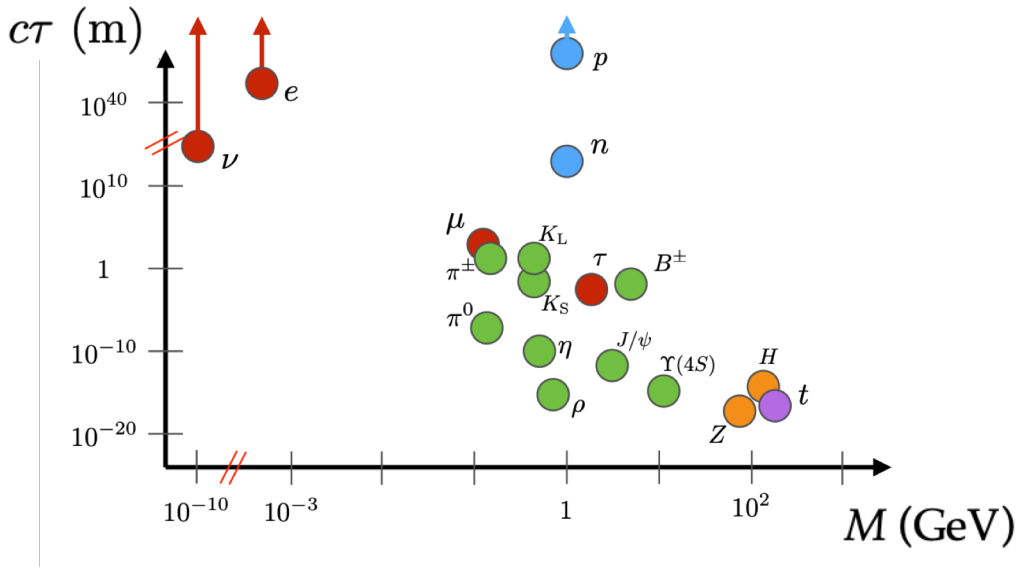
Summary

- Most Run 2 CMS analyses use mini or nano AOD
- Some exceptions: All heavy ions analyses (but moving to miniAOD), many B-physics analyses, most LLP analyses, all scouting analyses (HLT event content)
- LLP analysis needs are very specific: unlikely to find 1 or 2 skims that work for most LL analyses
  - That being said, might have 1 skim that works for a few LL analyses
- LLP analyses sometimes need RECO or even RAW event content
- Moving forward, the work will happen on several fronts:
  - Develop common skims for related LL analyses and/or analyses looking at the same dataset
  - Add variables to mini/nanoAOD so that some LL, B-physics, and heavy ions analyses can use them
  - Maintain this conversation among LL analyses and between LL and Physics Performance and Dataset (PPD) group, to maximize the efficiency and stay within operational limits
- The experience and input from the HSF analysis working group is extremely welcome!

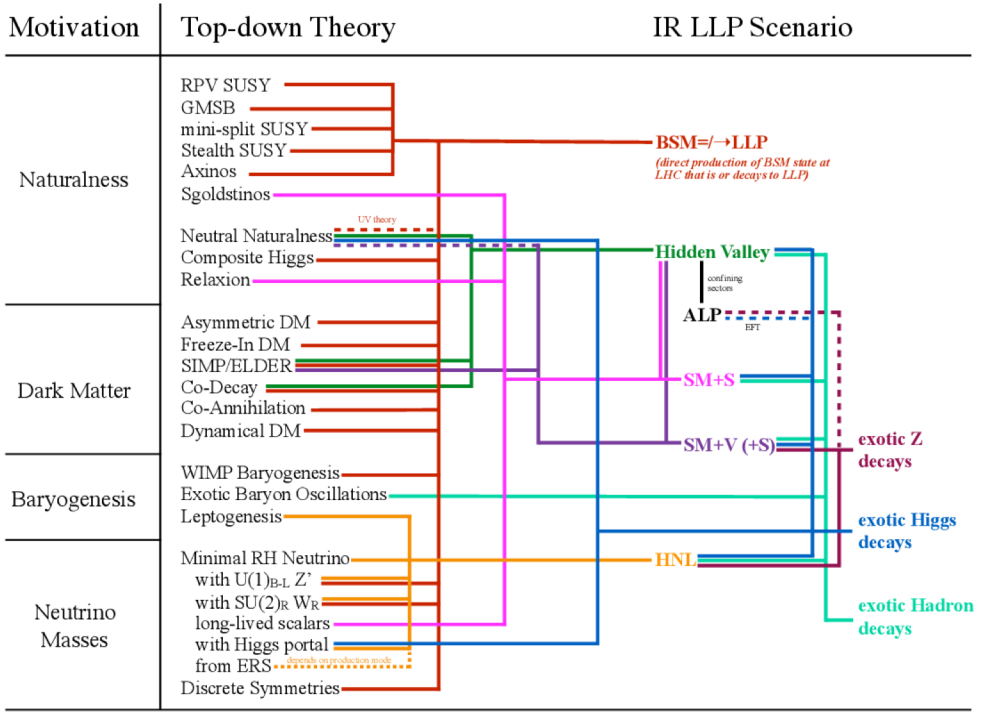
Backup

Why look for new long-lived particles (LLPs)?

Standard model particles span a **wide range of lifetimes (τ)**



LLPs appear in **many scenarios beyond the standard model**



Sexaquarks

- Uses **all events in all datasets** (accounting for overlaps)
  - Using the number of primary vertices, calculate the number of pp collisions
- Needs **AOD** to rerun the V0 reconstruction with relaxed cuts
- Large number of datasets to analyze! **Maybe a dedicated skim could be useful**
  - The details need some thought, once other questions about how this analysis will be done in Run3 are answered