A Search for Sexaquarks in Parked 2018 Data at CMS

Wren Vetens[1] (she/her & they/them)
In collaboration with: Steven Lowette[2] Kevin Black[1]

[1]University of Wisconsin - Madison
[2]Vrije Universiteit Brussel

USLUA Lightning Talk, Fermilab, 14 December 2023

Premise

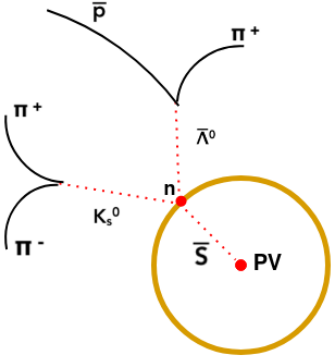
Occam's Razor appeal: "Sexaquark" (S): uuddss bound state. Search for $\bar{S}$ production at LHC, annihilating with a neutron in the CMS beampipe, and reconstruct charged final state in the tracking system.

As a deep QCD Bound State, could potentially **resolve muon g-2 anomaly** [4].

Very unique properties

- Electrically neutral, dibaryonic ($B = 2$), doubly strange ($S = -2$), spin–0
- No need for final state particles to come from the beamspot
- Stable (or very long-lived) if $m_S \lesssim 2\,\text{GeV/c}^2$ [3].

Within a mass range of ≈ 1.7 to 2 GeV/c^2, S is a **potential baryonic Dark Matter candidate.**



$$\bar{S}(\bar{u}\bar{u}\bar{d}\bar{d}\bar{s}\bar{s}) + n(udd) \rightarrow K_s^0\left(\frac{d\bar{s}-\bar{d}s}{\sqrt{2}}\right) + \bar{\Lambda}^0(\bar{u}\bar{d}\bar{s})$$

$$\bar{\Lambda}^0(\bar{u}\bar{d}\bar{s}) \rightarrow \pi^+(u\bar{d}) + \bar{p}(\bar{u}\bar{u}\bar{d})$$

$$K_s^0\left(\frac{d\bar{s}-\bar{d}s}{\sqrt{2}}\right) \rightarrow \pi^+(u\bar{d}) + \pi^-(\bar{u}d)$$

Strategy

Train Boosted Decision Tree (BDT) on Simulated $\bar{\mathrm{S}}$ signal and data–driven combinatorial background. Aim for a **zero–background** final result.

Simulation

Monte Carlo simulation of $\bar{\mathrm{S}}$ at 5 masses characterizing likely range for Dark Matter $\bar{\mathrm{S}}$: 1.7, 1.8, 1.85, 1.9, and 2.0 GeV/c^2[5].

- Production at LHC: custom-tuned verson of EPOS–LHC event generation software [7].
- $\bar{\mathrm{S}} - \mathrm{n}$ annihilation in beampipe and simulation of (grand) daughters in CMS detector: modified version of GEANT4.

Data

Use the 2018 CMS "B-Parking" Dataset for both **search region** and **background**.

- ≈ 300 Billion unique inelastic p–p collisions.
- Events passing loose trigger requirement "parked" to tape, processed later during LS2.

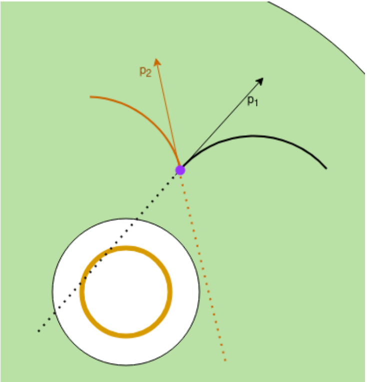
Fiducial cuts and a tight preselection must be applied for reasonable storage of analysis samples. The full dataset would otherwise require ≈ 3.5 petabytes to store.

Data Parking

- Limitation of standard data taking: **prompt reconstruction** requires a lot of computing resources.
- Raw “Parked” data instead saved directly to tape, reconstructed during a lull in computing infrastructure use (i.e. during a shutdown)
- More than the standard 1 kHz of physics events can be recorded! Rate now constrained by DAQ bandwidth and tape storage space.
- Since there is **No Trigger** associated with our final state, we search for our signal in the **pileup** of triggered events.
- What matters to our search is not the luminosity, but the total number of unique inelastic p-p collisions.
- Since our only source of background is combinatorial, pileup is also the greatest contributor.
- Want: Low Pileup, high degree of collisions $\Rightarrow$ B-Parking

Reconstruction

- Λ and K_s^0 formed from oppositely charged track pairs are fit to a common vertex using a *Kalman* vertex fitter as part of an adapted version of the CMS V^0 algorithm
- Reconstructed K_s^0 and $\Lambda^0/\bar{\Lambda}^0$ are kinematically refitted to their world average masses.
- The two then undergo a *Kalman* vertex fit to a common vertex. (For signal this will represent the annihilation vertex of the $\bar{\mathrm{S}}$).
- Common vertex required to have a good fit.
- The object is a "S" when it has a p from Λ^0, and a "$\bar{\mathrm{S}}$" when it has a $\bar{\mathrm{p}}$ from a $\bar{\Lambda}^0$.
- Low reconstruction efficiency ($\approx 0.006\%$) dominated by low final state acceptance and low reco. efficiency for $\pi^{\pm}$ from Λ ($\approx 13\%$ in accepted events).



Reconstruction Strategy

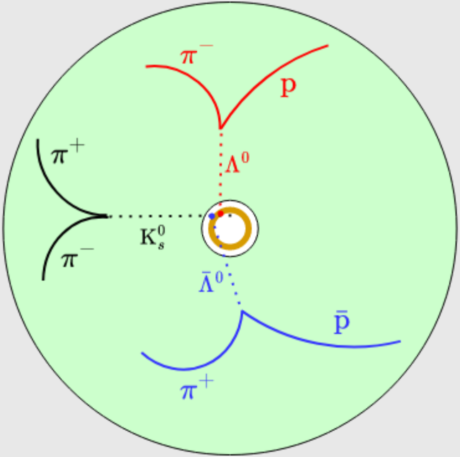
Search Region

$\bar{S}$ reconstructed from $\bar{\boldsymbol{\Lambda}}^{\mathbf{0}}$ and K_s^0 fit to common vertex.

Background

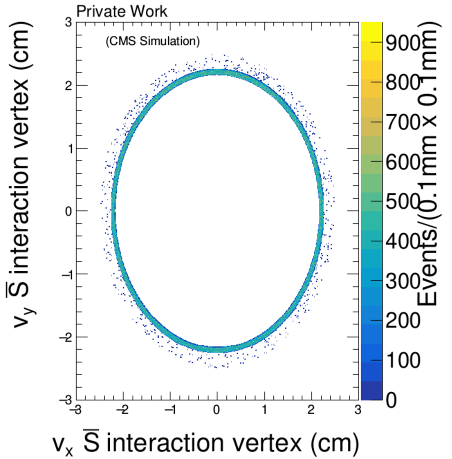
S reconstructed from $\boldsymbol{\Lambda}^{\mathbf{0}}$ and K_s^0 fit to common vertex.
Not a real S, since S does not annihilate with ordinary matter.

Reconstruct $\Lambda^0/\bar{\Lambda}^0$ with K_s^0 from **previous 50 events** (for higher statistics in the signal region) to compute correction factor for modeling $\bar{S}$ background as data S.



Simulation

- Annihilation with neutron, decay of $\bar{\Lambda}^0$ and K_s^0, and propagation of their daughters through the CMS tracker was simulated in a modified version of the GEANT4 simulation software.
- $\bar{\text{S}}$–n Interaction region consists of the ^{9}Be beampipe and surrounding air molecules.
  - Could in principle be extended to throughout the CMS tracker.
- Very low interaction cross section requires the introduction of a *looping* mechanism to ensure an economical simulation.
- Reintroduce expected angular dependency with angular weight factor - accounts for more beampipe material crossed at higher η



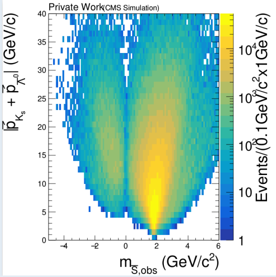
Neutron Fermi Motion

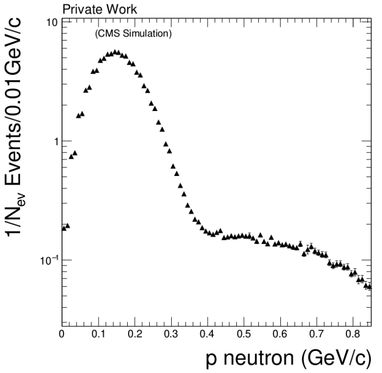
Wren Vetens

Nonzero neutron momentum in the ^{9}Be nucleus means that $\bar{\mathrm{S}}$ with high momenta have a greater spread in their "observed" mass:

$$\mathrm{m}^2_{\bar{\mathrm{S}},\mathrm{obs}} = \mathrm{m}^2_{\bar{\mathrm{S}}} + 2\left(\sqrt{\mathrm{m}_\mathrm{n}^2 + p_\mathrm{n}^2} - \mathrm{m}_\mathrm{n}\right)\left(\sqrt{\mathrm{m}^2_{\bar{\mathrm{S}}} + p^2_{\bar{\mathrm{S}}}} + \sqrt{\mathrm{m}_\mathrm{n}^2 + p_\mathrm{n}^2}\right) - 2p_\mathrm{n}p_{\bar{\mathrm{S}}}\cos\theta + p_\mathrm{n}^2.$$

Here, $\mathrm{m}_{\bar{\mathrm{S}}}$ is the true mass of the $\bar{\mathrm{S}}$ and $\mathrm{m}_{\bar{\mathrm{S}},\mathrm{obs}}$ is the mass calculated under a zero neutron momentum hypothesis. The "negative" mass occurs in the figure below when $(E_{\mathrm{K}^0_s} + E_{\bar{\Lambda}^0})^2 < \|\vec{p}_{\mathrm{K}^0_s} + \vec{p}_{\bar{\Lambda}^0}\|^2$.





Simulated neutron Fermi momentum

Parameterized BDT

Train *parameterized* BDT on data S background vs simulated signal $\bar{S}$:

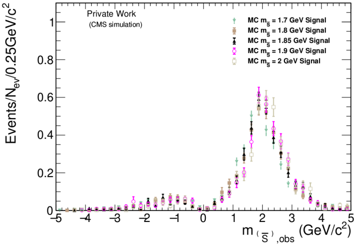
Parameterized BDT

BDT Training parameterized by $m_{\bar{S},gen}$ – a physical value for signal but not for background:

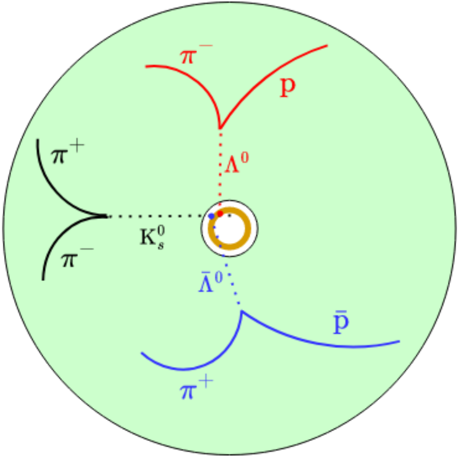
$m_{\bar{S},gen} = 1.7, 1.8, 1.85, 1.9, \text{and } 2.0\ \text{GeV}/c^2$

For background training sample, randomly assign each point a "$m_{\bar{S},gen}$" from these values. When classifying, evaluate model for each $m_{\bar{S},gen}$ as a mass *hypothesis*.

- Allows for model interpolation between different particle masses [1].
- $m_{\bar{S},gen}$ not straightforwardly distinguishable based on $m_{\bar{S},obs}$ due to smearing from neutron Fermi momentum.
- In the event of a discovery, S mass could be inferred through likelihood estimator [2] based on our model.



Recall...

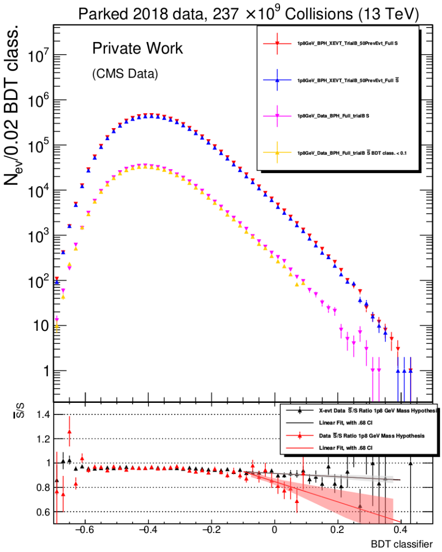


Data–driven $\bar{S}$ background “S” Reconstructed from p from Λ^0.
True search region, “$\bar{S}$,” is Reconstructed from a $\bar{p}$ from a $\bar{\Lambda}^0$.
Need a correction factor for this difference.
Factor must be extrapolated into the Signal Region: $\bar{S}$ with BDT $>$ 0.1

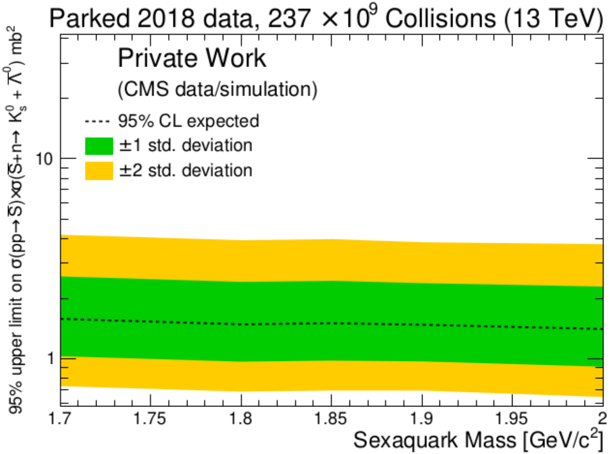
$S : \bar{S}$ Correction factor

Signal Region: $\bar{S}$ with BDT > 0.1

- Extract $\bar{S}/S$ ratio from $\bar{S}(S)$ with high BDT Classifier in combinatorial sample.
- Linear fit on ratio for BDT $\in$ (-0.1, ∞)
- Low statistics in data near signal region $\Rightarrow$ Large contribution to uncertainty from extrapolation.
- In data: Linear fit on ratio for BDT $\in$ (-0.1, 0.1) extrapolated into signal region. Difference with ratio in combinatorial sample computed as uncertainty ($\approx 40\%$ at BDT 0.35).



Anticipated Blind Asymptotic Limits



- Anticipate sensitivity down to 1/2000th of normal hadronic cross sections.
- Due to model parameterization, extrapolation factor and optimal BDT cut must be taken as functions of mass hypothesis.
- Aiming for Moriond 2024, planning to work through pre-approval checklist early January.

References I

[1] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *The European Physical Journal C*, 76(5), apr 2016.

[2] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, May 2020.

[3] Glennys R. Farrar. Stable sexaquark, 2018. arXiv:1708.08951[hep-ph].

[4] Glennys R. Farrar. The muon g-2 and lattice qcd hadronic vacuum polarization may point to new, long-lived neutral hadrons, 2022. arXiv:2206.13460[hep-ph].

[5] Glennys R. Farrar. A stable sexaquark: Overview and discovery strategies, 2022. arXiv:2201.01334[hep-ph].

[6] Greg Landsberg. B physics parking program at cms, 2018. Talk https://indico.cern.ch/event/754760/contributions/3127694/attachments/17141
parking-RDMS-2018.pdf.

References II

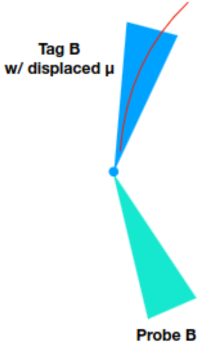
[7] T. Pierog, Iu. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. Epos lhc: Test of collective hadronization with data measured at the cern large hadron collider. *Physical Review C*, 92(3), Sep 2015.

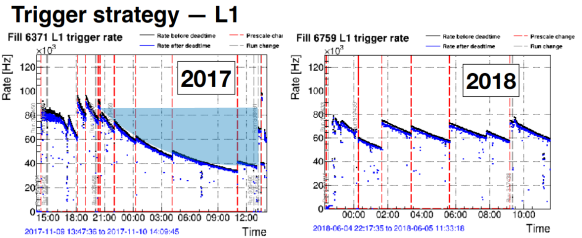
Backup

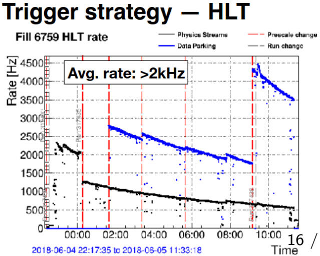
2018 B Parking

- Use existing triggers for $B^0 \to K^*\mu^+\mu^-$ combined with the fact that B-mesons are often produced in pairs to get a sample of $B^0 \to K^*e^+e^-$ for measuring $R(K^*)$
- Recorded 10^{10} unbiased B hadron events, average pileup approx 30

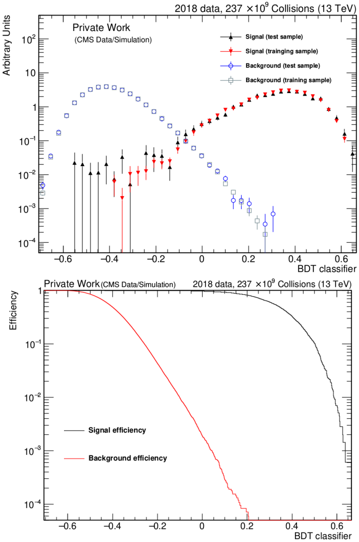
Figures on this slide from [6]




Trigger strategy — L1


Trigger strategy — HLT

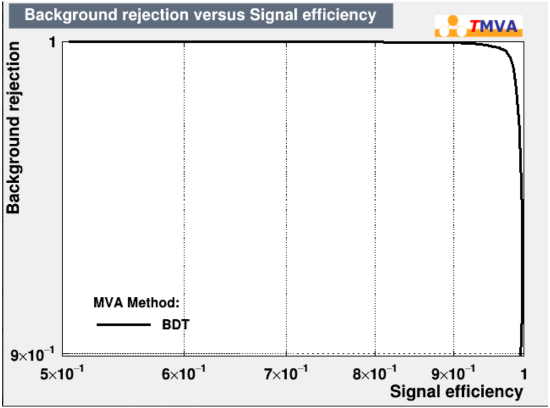
Training



BDT trained using the AdaBOOST algorithm, Sensitive variables include:

- Locations and of $\bar{\mathrm{S}}/\mathrm{S}$ interaction vertex and $\bar{\Lambda}^0/\Lambda^0$ and K_s^0 decay vertices.
- Longitudinal impact parameters of $\bar{\mathrm{S}}/\mathrm{S}$, $\bar{\Lambda}^0/\Lambda^0$, and K_s^0 with respect to the primary vertex (PV).
- Whether or not the $\bar{\mathrm{S}}/\mathrm{S}$, $\bar{\Lambda}^0/\Lambda^0$, and K_s^0 momentum points away from the PV.
- Number of PV present in the event.
- Fit quality of the $\bar{\mathrm{S}}/\mathrm{S}$ vertex.
- Reconstructed and generated S mass.
- K_s^0 p_T and η, and $\bar{\mathrm{S}}/\mathrm{S}$ η.
- Angular differences between the $\bar{\mathrm{S}}/\mathrm{S}$, $\bar{\Lambda}^0/\Lambda^0$, and K_s^0

Training Pt. II



BDT is precise enough to cut away **all** background while still having reasonable signal efficiency.

Need a handle on uncertainty in the high BDT–classifier signal region (BDT > 0.1), and must take into account the fact that all background is cut out in our optimization.

Both problems addressed with S : $\bar{\text{S}}$ Correction factor