August 17, 2018

Background studies for long-lived particles searches in the ATLAS muon spectrometer

David Aguillard
Wofford College

I. OVERVIEW

The search for long-lived, neutral particles which decay in the ATLAS muon spectrometer (MS) is one of the many searches for physics beyond the standard model (SM) undertaken at CERN. The idea is simple: a neutral particle that is not described by the SM will not interact with regular matter, but may eventually decay into SM particles. To search for these processes, we look for decays with no tracks in the inner detector (ID) and no activity in the hadronic calorimeter (HCal) or electromagnetic calorimeter (ECal) pointing toward its vertex. We call such a vertex a displaced vertex, because it appears displaced from the interaction point (IP).

My work at CERN involved the identification of background sources for long-lived particle (LLP) searches in the MS - anything that is reconstructed as an LLP in the MS, but is not an LLP. Background arises when vertex reconstruction algorithms falsely identify a vertex as coming from the IP and having no inner detector or calorimeter activity pointing toward it. The main source of background to LLPs decaying to hadronic jets in the MS is from hadronic or electromagnetic showers not contained in the calorimeter volume (*punch-through jets*) resulting in tracks reconstructed in the MS. Multijet events that contain vertices in the MS would have ID tracks and jets that point towards the displaced MS vertex as well as inwards to the IP. To reduce the acceptance of fake vertices from multijet events, vertices are required to be isolated with respect to ID tracks and calorimeter jets. Additional background, referred in this paper as *non-collision background*, can be generated by electronic noise in the MDT and RPC/TGC chambers, by cosmic ray muons, by multijet events with mis-measurement of jets and by machine induced background [1]. This last contribution, usually referred as *beam-induced background*, is composed of particles produced in the hadronic and electromagnetic showers caused by beam protons interacting with collimators or residual gas molecules inside the vacuum pipe.

Since the LLPs considered in my analysis are paired produced, we can look for events with one or two displaced vertices in the MS. The simplest strategy requires at least two MS vertices and it is inclusive to any other activity in the event. Thus, the two-vertex search has a very low level of background (very small probability of having an event with two fake displaced vertices passing the signal selection), but has also very little sensitivity to longer lifetimes. The second strategy requires only one displaced vertex in the MS. It can extend the sensitivity at very long lifetimes, but the background is much higher than in the two vertex search and has to be carefully estimated.

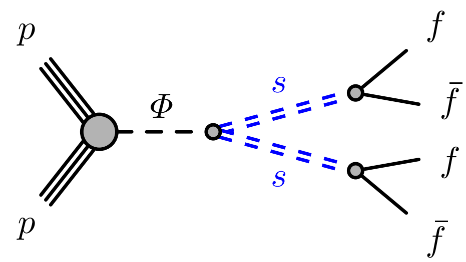

FIG. 1. Feynman diagram of the a scalar boson (Φ) decay to two long-lived scalars (s), which decay into SM fermions (f).

A. Scalar Portal Model

A novel way of introducing long-lived, neutral particles to the SM is through a hidden sector that weakly couples to the SM. For example, scalar (Higgs) portals [2–4], where the Higgs boson weakly mixes with a hidden sector scalar, can result in pair production of hidden sector scalars or pseudoscalars that carry no SM quantum numbers. The branching fraction limit for SM Higgs boson decays to undetected particles is currently at the 25% level [5] (assuming SM Higgs boson production and width), potentially allowing sizable branching fractions for decays to non-SM particles. The mechanism for LLP production in scalar decays is shown in Figure 1. Here, a scalar boson Φ decays with some effective coupling to a pair of long-lived scalars, s. The scalars s subsequently decay to SM particles. Since this model assumes that the coupling of the scalar to SM is the same as the SM Higgs, each long-lived scalar decays predominantly to heavy fermions, $b\bar{b}$, $c\bar{c}$, and $\tau^+\tau^-$. The ratio among the different decays depends on the mass of the scalar and for $m_s \gtrsim 25$ GeV it is almost constant and equal to 85:5:8.

The signature of a long-lived scalar decay is an isolated vertex with a high number of associated hits. Depending on the lifetime, the s decay vertices could occur in the ID, ECal, HCal, MS, or outside of the detector. Our search focuses on a lifetime such that the decay is in the MS, but regardless, it is much more likely only one of the vertices associated with the s decay occurs inside of the detector. Thus, the search for events which align with this model focuses on single, isolated vertices with a high number of associated hits in the MS.

II. SINGLE VERTEX SEARCH AND ABCD METHOD

An ABCD method is used to estimate the background for the single vertex search. In this section, I will give a general outline of the baseline event selection, then introduce the ABCD procedure, and finally present the problem my team encountered in the application of this procedure.

1. Baseline Selection

Our team uses 2015 and 2016 LHC data in its search for events containing LLP decays. The first step in cutting down this massive amount of data is an event level selection process, which begins by ensuring the event has a good primary vertex and has no detector malfunctions. Next, events which do not pass the Muon RoI Cluster trigger [6] are discarded.

The next step is to apply selection criteria on individual MS vertices, which are our candidates for s decays. The vertex is required to be matched to the triggering muon RoI cluster trigger and vertices in the barrel/endcap overlap region are not considered in the analysis. Moreover, vertices pointing to the HCal and ECal crack regions are not considered in the analysis, since in these regions, the amount of material is lower and the level of background higher. The HCal and ECal crack regions almost overlap the MS transition region, therefore the signal efficiency losses are very limited. In order to further reduce the background level, we require the vertex has a minimum number of MDT, RPC, and TGC hits (a vertex from a signal decay is expected to have more hits associated with respect to a vertex from background), and also a maximum on MDT hits that helps to reduce fake signal due to noise bursts.

A. ABCD Method

Consider a plot of two uncorrelated variables which form a plane. We fill the plane with entries, and divide the plane into four quadrants. Let A be the total number of entries in quadrant 1, B the total number of entries in quadrant 2, C the total number of entries in quadrant 3, and D the total number of entries in quadrant 4. Then:

$$\frac{A}{B} = \frac{C}{D} \tag{1}$$

is true as long as no source of background unevenly contaminates the quadrants (i.e. as long as the two variables are not correlated).

ABCD analysis is a powerful tool to estimate background events because, given that there is no unevenly distributed source of background, it allows us to predict the number of entries in a certain region through simple multiplication. Now if we consider a y-axis which measures total number of hits per vertex, and an x-axis which measures vertex isolation (i.e. the distance between the MS vertex and the nearest track or jet), and plot all MS vertices which pass the baseline selection, we can distinguish displaced vertices with a high number

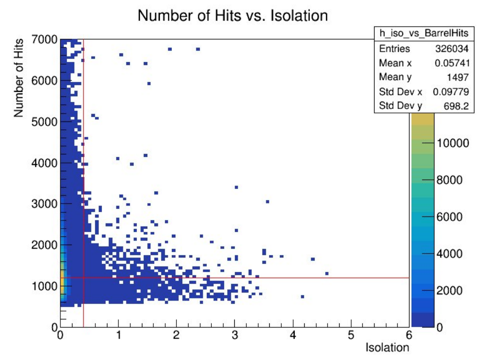

FIG. 2. ABCD Plane for single vertex search defined by the isolation and number of hits associated to the vertex. The top right quadrant contains entries which are candidate LLP decays (Region A).

of hits from vertices which result from phenomena originating from the ID and calorimeter activity. For my studies, a vertex is considered isolated if the angular separation between the vertex and a ID track or calorimeter jet is greater than 0.4 radians, and is considered to have a sufficient number of hits if it is associated with 1200 MDT + RPC hits (barrel), or 1200 MDT + TGC hits (endcaps).

These kinds of analyses are usually performed blinded, i.e. the signal region is not looked at during the analysis development. Using a variable that can discriminate between signal and background, it is possible to create a background dominated region, where the ABCD method is tested (the ”validation region”), and a signal dominated region that is kept blinded (the ”signal region”). This is how the analysis of the 2015 and 2016 data was performed. During the analysis development, the ABCD defined by the isolation and number of hits associated to the vertex, as described above, was tested and found to not work properly in the validation region (i.e. $A \neq \frac{BC}{D}$). Instead, the analysis used a different ABCD plane defined by isolation and the $\Delta\phi$ between the vertex and the missing energy in the event. Now, after the analysis has been unblinded and completed, we return to the question of why the original plane did not work as expected. It is likely that there is a source of background which acts like signal entering the ABCD plane.

In the following sections, I will discuss the work I did to understand the analysis in general, and the different ways I attempted to improve the ABCD performance.

III. MY ANALYSIS

1. Getting Started

Searching for sources of background begins with understanding how to deal with large amounts of data efficiently. Thus, my first task at CERN was to become comfortable working with C++ and ROOT. I began this process by creating my own skimmed data tree - a ROOT file containing only events which pass baseline selection, which allowed me

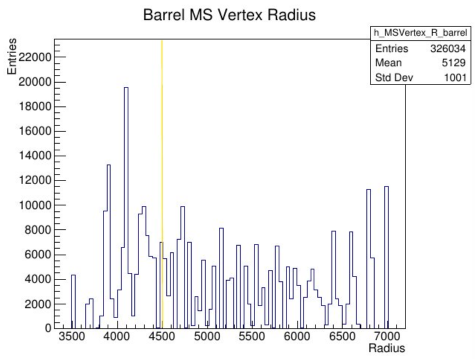

FIG. 3. Histogram of MS vertex radial position. All entries pass baseline selection.

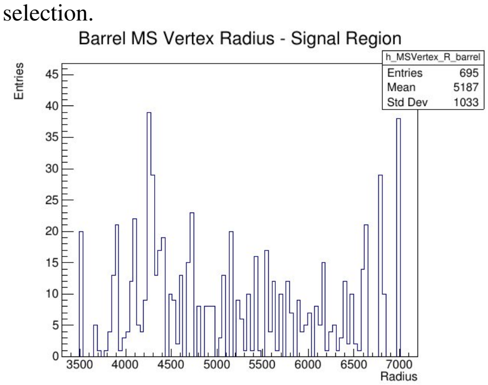

FIG. 4. Histogram of MS vertex radial position. in region A (where we expect signal). These events have a minimum angular isolation of 0.4, and at least 1200 associated hits.

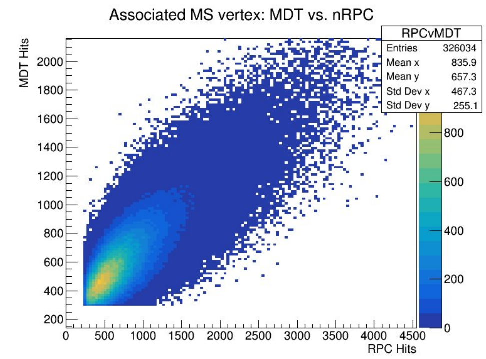

FIG. 5. 2D Histogram of the number of hits in the MDT vs. RPC associated with vertices. A strong correlation is expected.

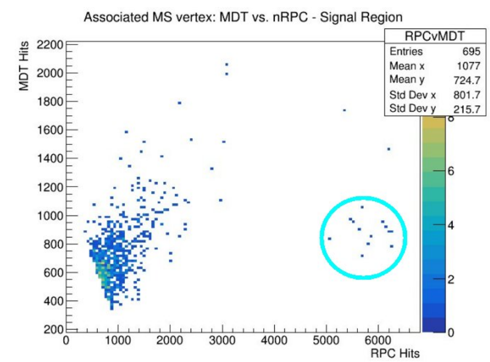

FIG. 6. Same 2D histogram, but only includes events in the region A. Note the cluster of entries with high RPC and low MDT hits.

to run on the 2015 and 2016 data in about 5 minutes instead of 6 hours. After creating trees for the data and a few Monte Carlo (MC) signal simulations, I was able to plot the event variables, as well as variables associated with the MS vertices. With these plots, I began experimenting with different cuts to see how they effected the ABCD performance. For example, there appeared to be an excess of MS verticies decaying in the HCal (radius < 4.5 m), and so I ran an experimental cut which excluded these hits.

These plots also allowed me to look for strange behavior in variables. Since background events due to noise, or other sources not coming directly from proton–proton collisions or beam-induced background, may leave telltale signatures, simply looking at the variables' distributions can lead to their identification. For example, if we plot number of MDT vs. RPC hits associated with each barrel vertex, we ought to see a strong correlation since these detectors are back-to-back. If we see clusters of entries with high MDT and low RPC hit numbers, or vice versa, this could be a sign of the presence of some non-standard background - like a noise burst or cosmic rays. While the plot in Figure 4 shows a very positive correlation between MDT and RPC hits for all events passing baseline selection, if we only consider signal events, we see a suspicious cluster (Figure 5). While this small cluster is rather insignificant, it inspired us to view the Atlantis Event Display (which I will discuss in the final section) that proved to be very helpful.

2. *Efficiency Considerations*

During my studies on possible variables that could characterize the background in region A of the ABCD plane, I have found that the $\mathrm{HT}_{\mathrm{miss}}$ variable (minus the vectorial sum of the p_{T} of the jets in the event) can effectively separate signal from background. Though it may be enticing to cut out vertices with properties that seem to be characteristic of background, one has to be sure that making the cut does not kill a significant portion of possible signal events. In order to ensure this, my next step was to study efficiency curves for the $\mathrm{HT}_{\mathrm{miss}}$ variable for data and some of the MC samples as a function of the cut on $\mathrm{HT}_{\mathrm{miss}}$.

Figures 7 and 8 show the histograms and efficiency plots of data and MCs, respectively, in region A for the $\mathrm{HT}_{\mathrm{miss}}$ variable. Looking at the histograms, we see that up to 50 GeV, there is a significant amount of data entries where the MCs predict very few. Looking at the efficiency, we see the cutting at 50 GeV retains about 0.95 signal efficiency.

Applying a selection on the $\mathrm{HT}_{\mathrm{miss}}$, the standard deviation between events counted in region A and the prediction $\frac{BC}{D}$ decreases from $\sigma = 18.5$ to $\sigma = 9.5$. Looking more closely, we find the improvement occurs partly because this cut removes events in region B at a higher rate than the other regions. Since

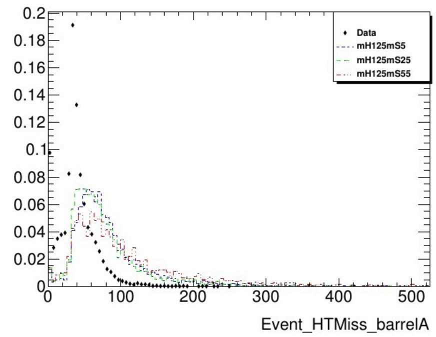

FIG. 7. Histogram of HT_{miss}. Notice the large amount of data just before the MCs begin to show signal at 50 GeV.

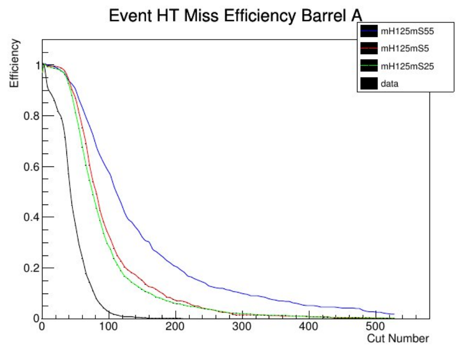

FIG. 8. A graph of efficiency as a function of the cut on HT_{miss}. If we cut at 50 GeV, we see we remove background while maintaining about 0.95 efficiency.

removing low HT_{miss} events from region B improve ABCD performance, we conclude events with low missing momentum populate the high isolation, low hit number section of our plane at an inordinate rate. Since we know this cut does not hurt signal, introducing it is a good step to improve the ABCD analysis.

3. *Other Cuts, without considering efficiency*

Comparing distributions for the data and MC simulations lead to a few other ideas for cuts. For example, when viewing the distributions of vertices in the transverse plane measured by the angle ϕ (Figure 9), I noticed there is a high concentration of hits when $\phi = \frac{n\pi}{4}$.

This is due to the high concentration of material associated with each barrel toroid tube, which increases the likelihood of fake vertex reconstruction from tracks and jets with enough energy to generate a particle shower in the material. To see if this causes additional background in our ABCD plane, I attempt a material veto. Knowing the distance from the IP and the width of the tube, I found that the $\Delta\phi$ associated with each tube is 0.2 radians. These cuts are easily seen in Figure 10. Of course, this removes a large amount of possible signal, but ignoring this issue for a moment, we do get an improvement in ABCD performance; our standard deviation drops from $\sigma =$ 18.5 to $\sigma = 13.5$. If we combine this cut with the previous cut on HT_{miss}, σ drops even further to 6.9. For fun, we can cut a $\Delta\phi$ of 0.5 radians around the vertex. If we do this along with the HT_{miss} cut (Figure 11), we find that $\sigma = 0.045$, and the ABCD is fixed! Unfortunately, this also cuts out a lot of signal events.

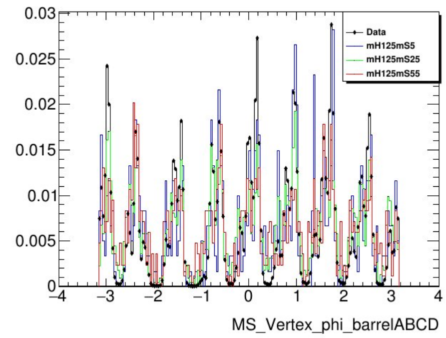

FIG. 9. Histogram of the angle in the transverse plane ϕ of MS vertices. Notice the equally spaced peaks associated with the toroid tube positions.

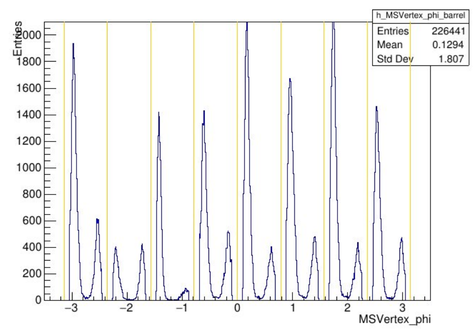

FIG. 10. Histogram of ϕ position, this time with the material veto on the toroids.

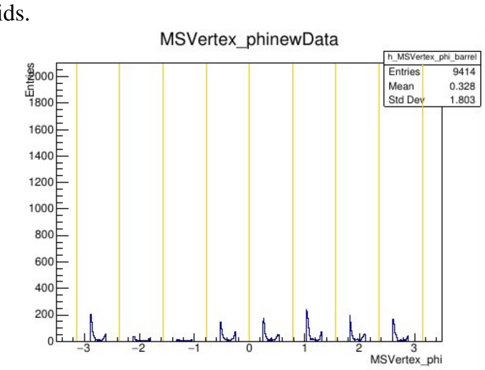

FIG. 11. Histogram of ϕ position, this time with the material veto on the both peaks, and a cut on HT_{miss}.

IV. EVENT DISPLAY

Analyzing the data in its entirety through key variables is not the only way to search for unaccounted background in the ABCD plane. Another approach is to view event display, and determine if there are any signatures in individual events which point toward a systematic issue.

After reviewing a number of the signal events, three possible sources of background in region A emerged:

1. Reconstructed MS vertices which lie on bent combined muon tracks

   The current analysis only isolates with respect to ID tracks. This means that if a track bends past the interaction point for any reason, a vertex which is not isolated with respect to the bent part can be considered isolated in our analysis. In event display, there were multiple cases where a reconstructed MS vertex was not isolated from a bent combined track. Thus, we have a source of signal background which can be eliminated by isolating with respect to the entire combined track.

2. Reconstructed MS vertices with the same trajectory as muons reconstructed in the MS only

   There were also multiple vertices which were reconstructed directly in the path of detected muons. Since we do not isolate with respect to these muons, a cut on vertices of this type may also improve ABCD performance. However, we need to be very careful, since tracks that do not point to the IP might come from signal events.

3. Reconstructed MS vertices which lie on extrapolated tracks

   In all cases, reconstructed vertices were found on extrapolated muon tracks. This is expected, since extrapolated tracks are reconstructed using MS only tracks extrapolated back to the IP (with additional corrections to take into account of the energy losses in the calorimeter). As for the MS only tracks, a possible isolation on these objects should be carefully validated.

Recall that the events with these characteristics are all in region A, meaning that there should be no activity surrounding them. Yet there is activity, which means the vertices belong in regions C or D. Unfortunately, I will not be able to explore any of these cuts, as the information on combined tracks, muon spectrometer tracks, and extrapolated tracks are not available in the current analysis ntuples. If I did have access, the first point addressed would have to be efficiency after performing each of these cuts. While an efficient cut on combined tracks

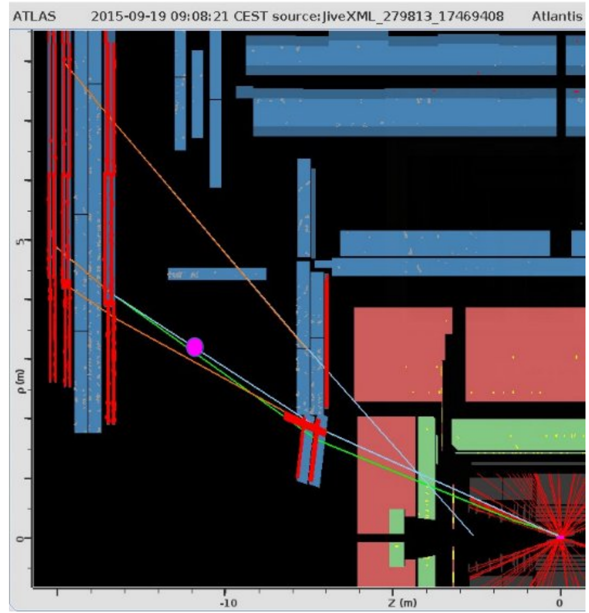

FIG. 12. Event display of background in region A. Notice the bent combined track (Green line) which points to the vertex (Purple circle). Ideally, this vertex would be an entry in region C or D instead of A.

(i.e. muons) is plausible, it is less plausible for a cut on extrapolated tracks or MS only tracks, since anything which generates hits in the MS, including LLP signal, is liable to produce an extrapolated track.

V. CONCLUSION

While it is unfortunate that I will be leaving, and therefore unable to run tests on the new containers, I am proud of my accomplishments over my two months here at CERN. My task, to search for background and improve the ABCD plane, was accomplished. I also increased my competence in many areas of computing, data analysis, and presentation skills, all while experiencing a brand new culture.

[1] R. Bruce et al., *Sources of machine-induced background in the ATLAS and CMS detectors at the CERN Large Hadron Collider*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **729** (2013) 825-840.

[2] Strassler, Matthew J. and Zurek, Kathryn M., *Echoes of a hidden valley at hadron colliders*, Phys. Lett. B **651** (2007) 374-379. arXiv:hep-ph/0604261.

[3] Strassler, Matthew J. and Zurek, Kathryn M., *Discovering the Higgs through highly-displaced vertices*, Phys. Lett. B **661** (2008) 263-267. arXiv: hep-ph/0605193.

[4] Strassler, Matthew J., *Possible effects of a hidden valley on supersymmetric phenomenology*, arXiv: hep-ph/0607160.

[5] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s} = 7$ *and* 8 *TeV*,JHEP **08** (2016) 045, arXiv: 1606.02266 [hep-ex].

[6] ATLAS Collaboration, *Triggers for displaced decays of long-lived neutral particles in the ATLAS detector*, JINST **8** (2013) P07015, arXiv: 1305.2284.