Search for Long-Lived Particles with HCAL Segmentation in CMS at the Large Hadron Collider

Katia Avanesov
Mentor: Harvey Newman
Co-mentor: Kiley Kennedy

1 Introduction

Long-Lived Particles (LLPs), often classified as particles with lifetimes greater than 10^{-10} seconds, feature in both the Standard Model (SM) – such as charged pions, neutrons, electrons, and protons – as well as in many Beyond the Standard Model (BSM) theories, such as those relating to dark matter, matter anti-matter asymmetry, and supersymmetry [1]. In this paper, we study the process, $H \to XX \to b\bar{b}b\bar{b}$, in which it is hypothesized that a Higgs boson first decays to an unknown long-lived particle (LLP), X, which in turns decays to the commonly observed final state consisting of pairs of bottom and anti-bottom quarks. The process is illustrated below in Figure 1. Our analysis will also provide insight into heavy twin Higgs boson models [2]: given that the masses of the Higgs boson and LLP are left as free parameters within our model, we may scan across various mass combinations, including heavier Higgs boson masses. The Higgs boson itself has been a focal point of study at the LHC ever since its discovery in 2012 [3], so the possibility of its coupling to an exotic particle, or the existence of its heavier twin, offers exciting potential for the discovery of New Physics.

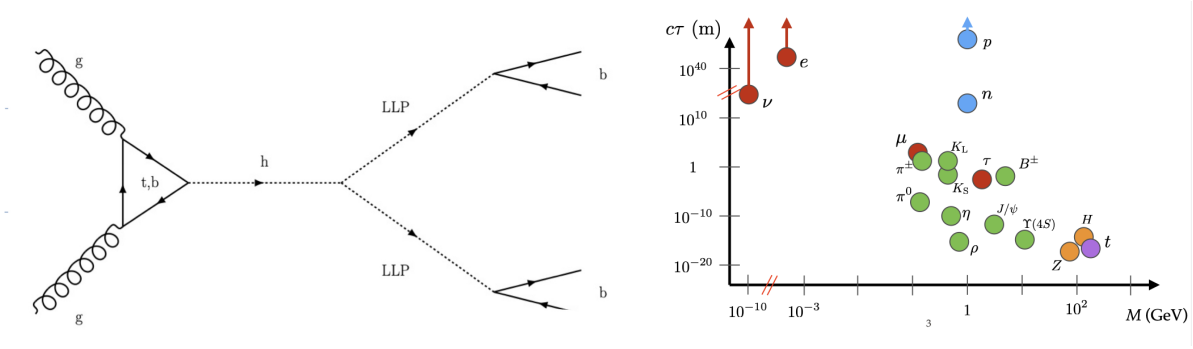

Figure 1: [Left] Feynman diagram showing the formation of a Higgs boson via gluon-gluon fusion and its subsequent hypothesized decay into a pair of LLPs which in turn decay to $b\bar{b}$ pairs. [Right] Plot of SM particles' lifetimes (τ) in units of $c\tau$ as a function of their mass in GeV. [4]

Many of the particles that CMS physicists have been interested in decay almost instantaneously after they are produced, such as the Higgs boson itself with a lifetime on the order of 10^{-22} seconds. A short-lived particle would therefore decay effectively at the center of the proton-proton interaction point, subsequently producing many decay products such as leptons, photons, and hadrons. The subdetectors of the CMS experiment, shown in Figure 2, are optimized to detect the many signatures that the original particle leaves behind. For instance, the CMS silicon-based inner tracker records the trajectory of charged particles through the readout of electrical signals, while the electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) record the energies of electrons and photons, and hadronic particles, respectively. The information from these many sub-detectors is then used to reconstruct the masses, momenta, and charges of the decay products, from which the identity of the original particle can then be deduced.

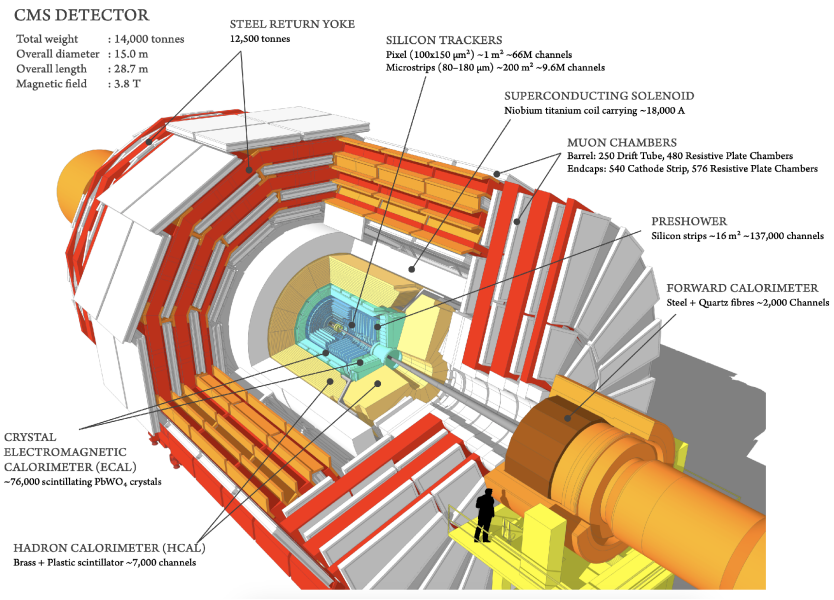

Figure 2: Diagram to show the components of the CMS detector [5].

By contrast, LLPs such as those hypothesized in the $H \to XX \to b\bar{b}b\bar{b}$ interaction could travel up to several meters before they decay, and thereby leave relatively fewer signatures within the detector, especially within its innermost parts such as the inner tracker or the ECAL. However, this absence of signal is a signature in and of itself. Indeed, this is the aim of the joint effort of the Caltech and Princeton group in their analysis of searching for Long-Lived Particles: to exploit their distinct topology, or lack-thereof.

One principle of our search strategy lies in selecting events with LLPs decaying within the HCAL because their signatures look very different compared to our main sources of background, such as QCD, $Z+$ jets and $W+$ jets, which form 'prompt' jets, as opposed to delayed ones. These differences are illustrated below in Figure 3. Firstly, we observe that a higher fraction of delayed jets' energy would be deposited in the outermost layers of the HCAL, since we require that the LLP decays inside of it. We expect a more even distribution of jet energy fraction from the prompt jets, if not more being deposited within the inner layers, given that the HCAL is made of very dense material and so jets produced before the HCAL are less likely to travel as far inside of it and deposit its energy far into its depths. Similarly, due to the dense HCAL material, we also see that the jets produced from within the HCAL are less 'spread out' compared to jets produced immediately within the LHC beam, as a result of their displaced vertices.

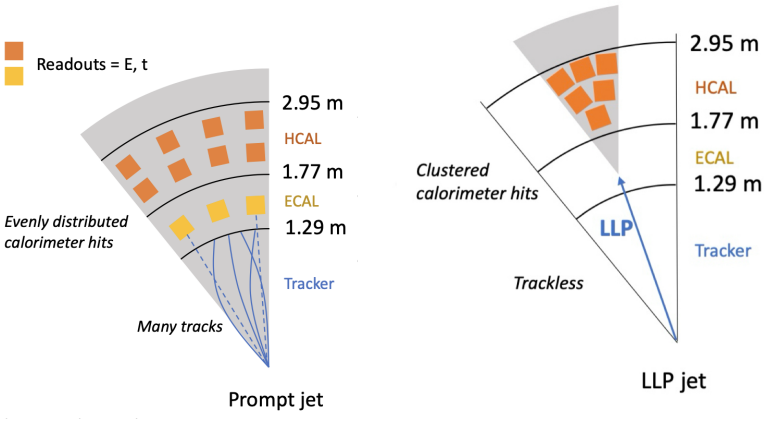

Figure 3: Illustration of a cross-section of the CMS detector to show the difference in signatures left by a prompt jet (left) and a delayed LLP jet (right). [4].

During my SURF program last year, I developed a prototype set of dense neural networks make selections of LLPs based on different metrics of their jets, trained on a Monte-Carlo generated signal of LLPs and a Run 3 dataset that was skimmed for a particularly signal-depleted region. The implementation of more advanced classification techniques is necessary for our analysis due to the minute cross-section of the Higgs boson decay at a mass of 125 GeV and at a total center of mass energy, $\sqrt{s} = 13.6$TeV, namely $59.2^{+5\%}_{-7\%}$ pb [6]. With an integrated luminosity of 160 fb^{-1} reported so far for Run 3 in September 2024, [7], we expect to find on the order of just 10 million Higgs boson decay events across the entirety of the Run 3 dataset. Thus, within our dataset of 100 million events that passed our L1 trigger, we would expect to be left with even fewer signal events. Moreover, although the Higgs boson to LLP decays are not predicted by the standard model, the Higgs to 'unknown' branching ratio remains as high as 11%, [6]. This effort has since evolved into forming the foundation for the 'assymetric double-tag' search strategy, which is outlined as follows. In this approach, we use two separate classifiers to analyze the two LLP jets produced within a single Higgs boson decay event (as illustrated in the Feynman diagram in Figure 1). Firstly, we select our event-space to be those events which have at least one jet decaying within the HCAL. This offers a good balance between rejecting noise yet still retaining a large number of events in our dataset. The jet that decays within the HCAL is then passed through the "depth" tagger, which is trained on information including the fraction of energy deposited in the layers of the HCAL. The output of this depth classifier for a given jet is then a score; using some pre-defined threshold, we deem the jet an LLP if its score exceeds this threshold. Similarly, the jet that is free to decay anywhere within the detector is passed to an 'inclusive' tagger. Then, we can represent our events within a 2D space, with each event consisting of a pair of jets assigned two numbers: its inclusive and depth score, as illustrated in the Figure 4 below.

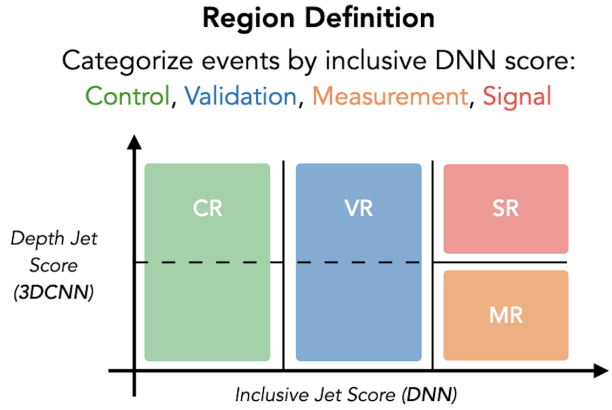

Figure 4: A visualization of the plot of the inclusive tagger score versus the depth tagger score. The dashed horizontal line represents a chosen fixed score threshold above which we treat events as signal, and below which, as background. CR: Control Regions. VR: Validation Region. MR: Measurement Region. SR: Signal Region.

In the simplest case, we assume that the depth and inclusive scores are uncorrelated— this is reasonable if the jets truly behave as independent objects. This then implies that within any specific bin of DNN scores, the fraction of events classified as signal or background by the depth tagger remains constant. We can validate this assumption by examining the control (CR) and validation (VR) regions (see the figure above): we check for the proportion of signal and background (based on the depth score threshold). If the proportions within the CR and VR bins are consistent, we can extrapolate this consistency into the signal region (SR), allowing us to predict how many events should appear there, based on the number of events within the measurement regions (MR). Comparing the actual number of events in the SR to this prediction then informs us whether there is a significant discrepancy that might indicate new physics. The goal of my SURF project is therefore to find an optimum set of thresholds that maximize the statistical significance of our counting experiment, by aiming to maximize signal selection and background rejection.

2 Current work

To approach this problem, I first began by implementing a script to carry out the necessary calculations, building off of existing code but updating and refactoring it to achieve more efficiency and applying the most up-to-date cuts and data. The workflow of the main body follows the procedure outlined below.

Initializing the regions Defining histogram cuts based on DNN depth and inclusive scores, using the desired/ inputted set of thresholds.

Applying physical cuts and binning data We apply physical cuts by requiring that the two jets in the event are tagged as an LLP candidate by the L1 trigger ("DepthTagCand" and "InclTagCand"). These cuts reduce the size of our background datasets by approximately two orders of magnitude, leaving on the order of 10^5 events. The remaining events are then binned according to their transverse momentum, eta, and phi metrics, each partitioned by 10 bins, resulting in 10^3 bins in total.

Calculation of Mistag Rate The mistag rate is calculated for each bin in the Control Region: it is the fraction of events in the bin that are 'mistagged', divided by the total number of events in the bin. The mistag rate is 0 for any empty bin.

Validating the Mistag Rate Using the calculated per-bin mistag rate, we perform per-bin multiplication to calculate the expected number of events in each bin. This prediction is compared with the true data: both in terms of total number of predicted events, and per-bin comparisons. The total predicted number of events for the 2023 Bv2 LLPSkim dataset (v3.13) was in agreement with the true number up to less than 1% error. Figure 5 below shows a comparison of the predicted and true mistag rates for the transverse momentum bins.

Predicting the SR region In the same way that we calculate the prediction for the Validation Region, we predict the number of events in the Signal Region based on the total number of events in the Measurement Region. However, we cannot check the true number of events in the Signal Region prior to the official un-blinding of our analysis.

Visualising the data To visualise how many events are predicted in the Signal Region, we may perform 'scans' by varying the combinations of depth and inclusive tag thresholds and visualising the data as a 2D histogram or heatmap. An example plot is shown below in Figure 6.

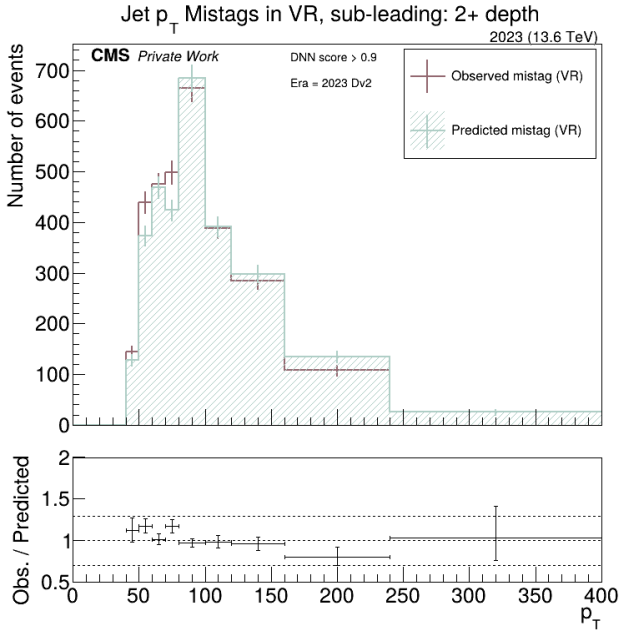

Figure 5: A plot showing a comparison between the predicted number of mistagged events in the Validation Region and the true number of mistagged events. The plot shows that the prediction captures the distribution of mistagged events well, though there are some discrepancies.

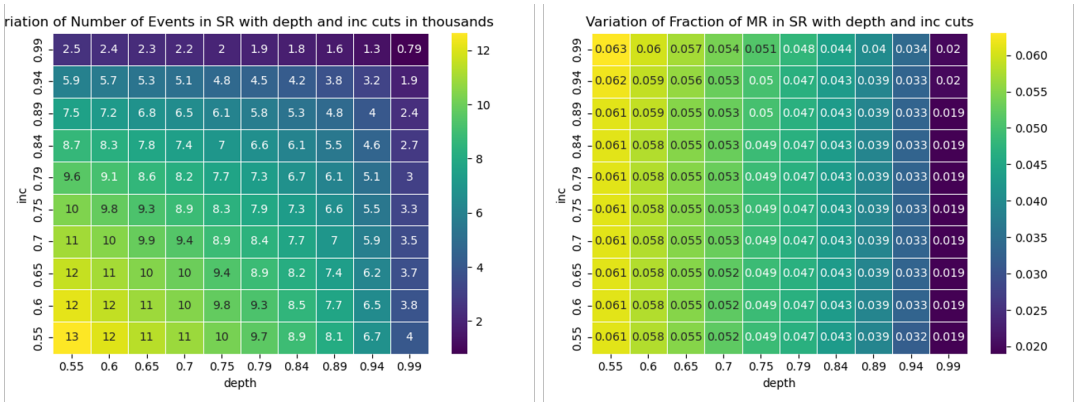

Figure 6: 'Threshold scans' for the 2023 Dv2 run era. Left: the number of events predicted in the SR. Right: the fraction of events in the SR relative to the total MR.

The two scans below reveal important information and distinct trends. Firstly, we see from the left plot that even with the most relaxed thresholds of 0.55 for both the depth and inclusive tagger, we only expect 13 events, which is about three orders of magnitudes less than the total number of events in the entire space. This is very promising given that we would ideally like on the order of 10 events in our signal region as this would highly maximise our statistical significance. The right plot shows a much more constant distribution vertically: i.e., the fraction of events in the SR relative to the total MR is constant even while changing the inclusive threshold. This acts as a sanity check and makes sense given that this should be equal to the total mistag rate calculated for the control region. However, we see that varying the depth threshold does change the fraction of events in the SR. This also aligns with what we would expect, given that the fraction decreases with more stringent limits.

Figure 7 below shows the scan for a signal file, in terms of the predicted number of events for each combination of thresholds. As we would hope, there are several times as many signal events retained for the same thresholds than there are background events: for comparison, consider the (0.75,0.75) thresholds: we see that while there are predicted 8.3 events for background, we have 61 events for signal. This is a promising start for our next steps as we look more into rigorously optimizing the threshold selection.

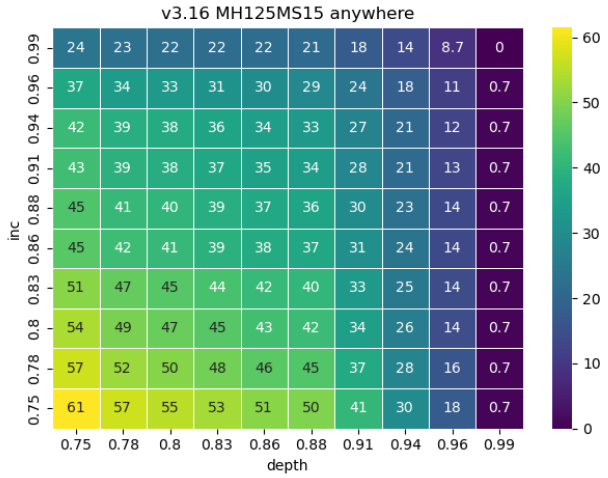

Figure 7: A threshold scan for the MH125-MS15 signal file.

As we look further towards optimizations we must consider several challenges that we need to address

Varying Mis-Tag Rates: In reality, we observe that the mis-tag rate varies with the physical parameters of the jets, such as its momentum, ϕ, etc. Therefore, it is necessary to study how the mis-tag rate varies as a function of each parameter. While we have incorporated binning the mist-tag rate for momentum, eta, and phi, we may consider binning on some other variables as well.

Uncertainties due to Neural Networks: Finally, we must account for and quantify the uncertainties introduced by the neural network models themselves due to their own mis-tag rates (i.e., assigning a high score to a background jet, given that background rejection will not be 100% at any reasonable threshold we choose).

References

[1] C. et al., "Long-lived particles at the energy frontier: the mathusla physics case," *Reports on Progress in Physics*, vol. 82, no. 11, p. 116201, Oct. 2019. [Online]. Available: http://dx.doi.org/10.1088/1361-6633/ab28d6

[2] Z. Chacko, H.-S. Goh, and R. Harnik, "Natural electroweak breaking from a mirror symmetry," *Phys. Rev. Lett.*, vol. 96, p. 231802, Jun 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.96.231802

[3] S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[4] G. Kopp, "Searching for llps in run 3 with a dedicated hcal timing trigger," Presentation at FNAL Pileup, February 2024. [Online]. Available: https://indico.cern.ch/event/1384246/#5-searching-for-llps-in-run-3

[5] "Cms detector introduction - cms open data workshop 2023," CMS Open Data Workshop, 2023. [Online]. Available: https://cms-opendata-workshop.github.io/workshop2023-lesson-cms-detector/01-introduction/index.html

[6] S. Navas *et al.*, "Review of particle physics," *Phys. Rev. D*, vol. 110, no. 3, p. 030001, 2024.

[7] CERN. (2023) Accelerator report: Lhc run 3 achieves record-breaking integrated luminosity. Accessed: 2024-09-21. [Online]. Available: https://www.home.cern/news/news/accelerators/accelerator-report-lhc-run-3-achieves-record-breaking-integrated-luminosity