Noah Hinojos

Professor Whiteson

29 April 2022

Physics 121W

Searching for Z`; Boson or Bogus?

**Abstract**

This research investigates how the LHC would statistically verify the existence of a new elementary particle. The investigation contextualizes this verification process, using simulated proton-proton collision data to search for a hypothetical Z` boson. The invariant mass of the resulting electron-positron decay is first measured and plotted. Then, the statistical significance of each mass window will indicate the likelihood Z`’s is located there. The results demonstrate Z`’s mass most likely resides within the 200GeV invariant mass window, with a statistical significance of 62.437, weighing (201.3 ± 9.2) GeV. However, many of the final measurements are limited due to predetermined histogram weights and arbitrarily decided binning. A hypothetical particle could be detected by the LHC, however, the intensive statistical analyses needed to assert this extend far beyond the scope of this investigation.

**Introduction**

The Standard Model currently identifies 17 elementary particles: 12 fermions that comprise matter and 5 bosons that dictate its interactions. This model is widely accepted as the standard subatomic catalog yet it is nevertheless incomplete. It does not coincide with general relativity nor can it explain the universe’s continuous expansion. Many theorists therefore try expanding the Standard Model, incorporating new fundamental forces and particles. Many of these theories can be contextualized as hypothesizing some new elementary particle, such as a Z` boson (hereby referred to as Z`). Z` might be an additional, heavy-vector boson that momentarily exists as a result of proton-proton collisions. Unfortunately, these protons also assimilate into a virtual photon and, in either case, the resulting particle immediately decays into an electron-positron pair. This makes it difficult to determine if this hypothetical Z` exists because it could not be directly observed. However, the invariant mass of the electron-positron pair is determinable, hence, any particle prior-decay would have this same invariant mass due to conservation of energy. The invariant mass of virtual photons varies greatly yet the invariant mass of this hypothetical Z’ would be consistent. Therefore, consistent measurements of invariant mass from electron-positron decay would imply the pair emerged from a new, consistently-massed particle, Z`.

When investigating this new particle’s existence, the invariant mass of numerous electron-positron decay events must first be measured. They should span a broad range of mass windows to ensure all possible masses for the Z` are considered. The statistical significance of these measurements in respect to typical background events may be used to either support or reject the hypothesis that a new particle, Z`, exists. If the significance at a given window is notably large, then it may imply Z` is located there. However, if the statistical significance is too low or even equal across all windows, then the measurements are too ambiguous to proclaim discovery.

**Methods**

This search for hypothetical Z` is emulated by dissecting a simulated dataset of proton-proton collisions, analogous to what the LHC might record. There are 6 subsets of data within this dataset, each corresponding to a specific invariant mass window: 200 GeV, 300GeV, 400GeV, 500GeV, 750 GeV, and 1000 GeV (assuming natural units). Each subset consists of approximately 100,000 background-signature and 10,000 signal-signature events of electron-positron decay. Each event tabulates the number of detected electrons, their charges, as well as their four-momenta in modified polar coordinates.

The combined invariant mass of the electron-positron pair will be calculated using the following equation:

Where represents each particle’s three-momenta. The energy of an individual electron/positron will be determined from the length of the particle’s four-momenta vector:

The combined invariant masses at each window will then be modeled in Python using weighted histograms. At every mass window, the number of background entities are plotted against the number of background+signal entities. The weights for each window are predetermined, depicted on the following table:

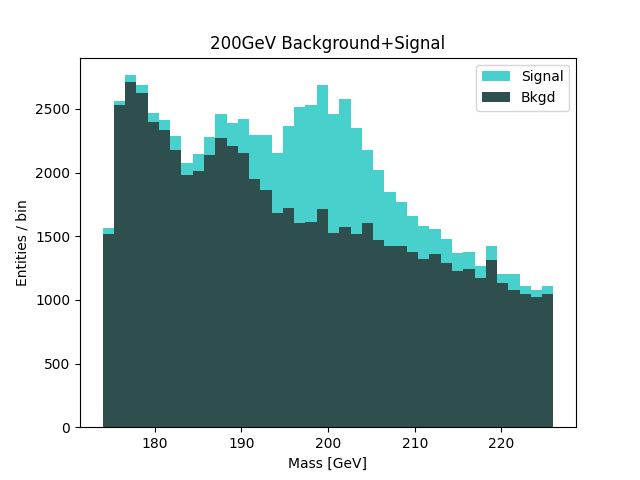
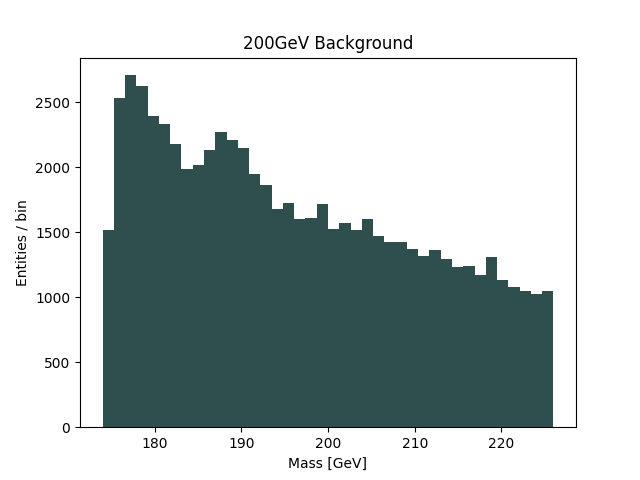
| **Invariant Mass** | **Signal Weight** | **Background Weight** |
| --- | --- | --- |
| 200 GeV | 3.0 | 2.0 |
| 300 GeV | 2.0 | 2.0 |
| 400 GeV | 1.5 | 2.0 |
| 500 GeV | 1.0 | 2.0 |
| 750 GeV | 0.75 | 2.0 |
| 1000 GeV | 0.50 | 2.0 |

Each histogram will also be formatted with 40 bins and span the boundaries:

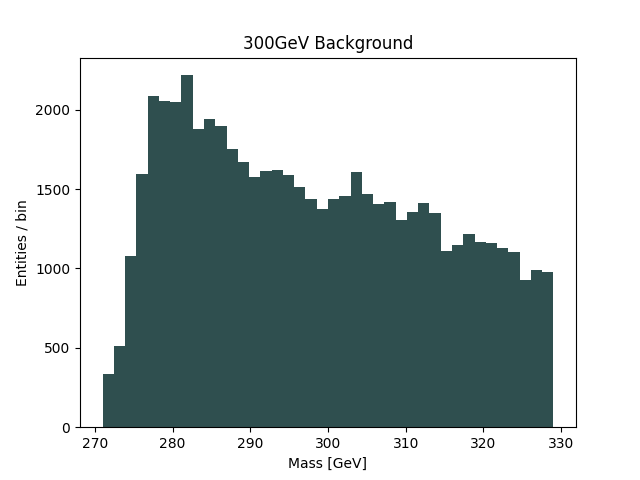
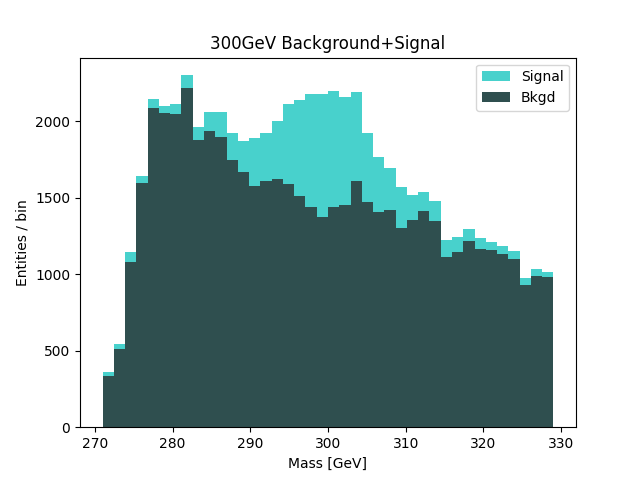
Next, the statistical significance should exemplify how many signal entities are detected in proportion to the number of background entities. Numerically, this significance is calculated as . Unfortunately, determining within what mass range to consider entities is much more challenging. The range cannot span the entire histogram because then the potential mass values would be too broad—recall that Z` is hypothesized as a consistently-massed particle. However, the range must be large enough to not overlook any significant mass values that would correspond to Z`. In other words, the most optimal mass range should yield the highest statistical significance. To determine this range, first denote all bins that represent peaks in the number of signal entities and then record their invariant masses . Then, use numerical analysis to find what mass range Δ about maximizes statistical significance. Note that Δ corresponds to the most optimal range of bins stemming from the bin, and Δ will denote the uncertainty of the precise mass of Z`, . The finalized significance should depict the optimal likelihood Z` exists within the window.

**Results**

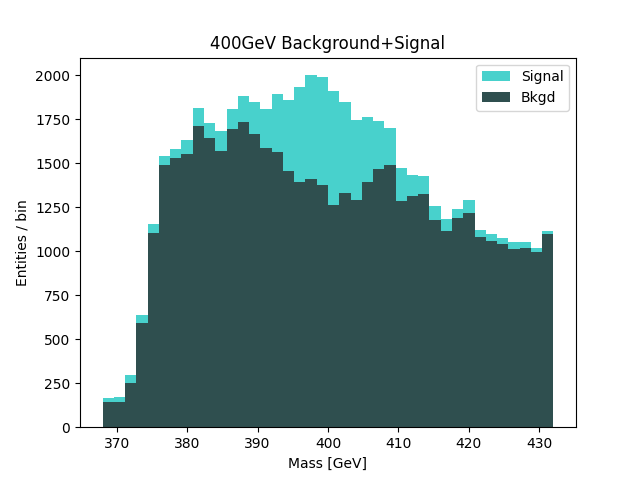
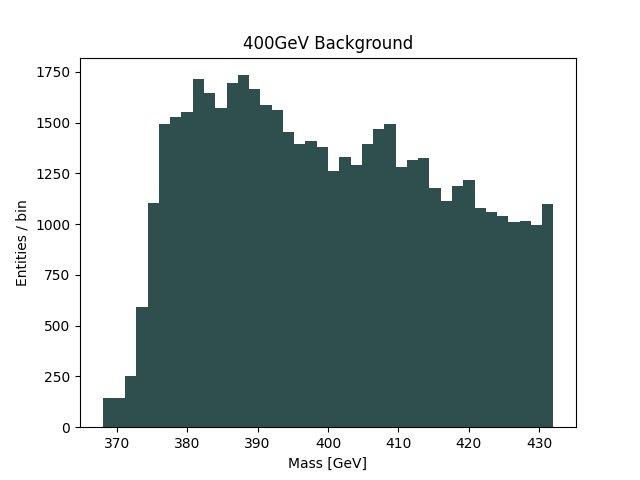
Of the 6 mass windows addressed, each demonstrated a substantial statistical significance of finding Z`. This is visible in every histogram as each graph has a noticeable peak in the number of signal entities about the central mass.



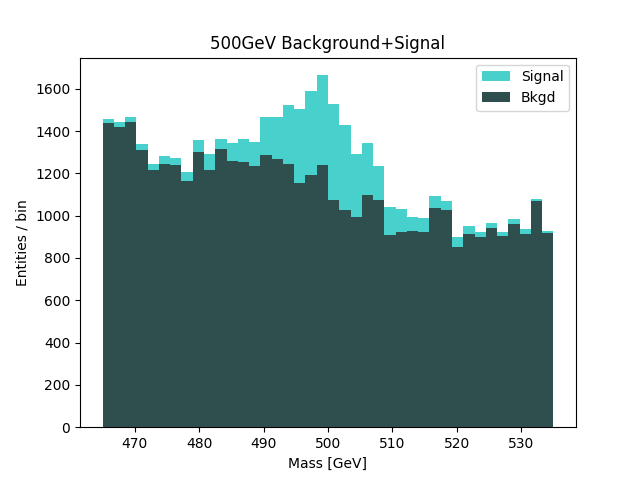
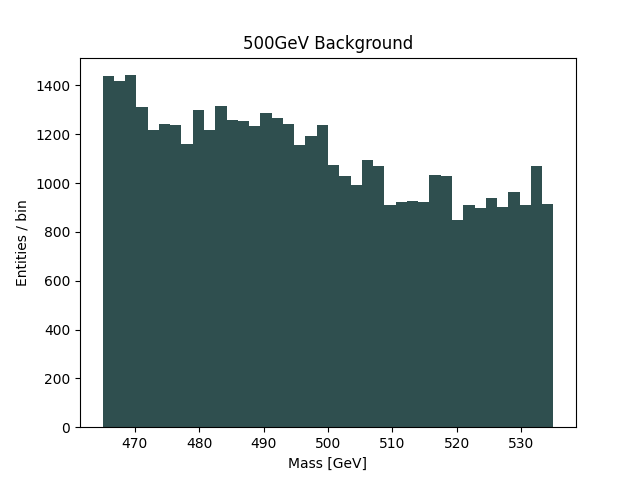
For the 200GeV invariant-mass window, the statistical significance of finding Z` is 62.437. Here, Z`’s invariant mass would be within the range of (201.3 ± 9.2) GeV.

****

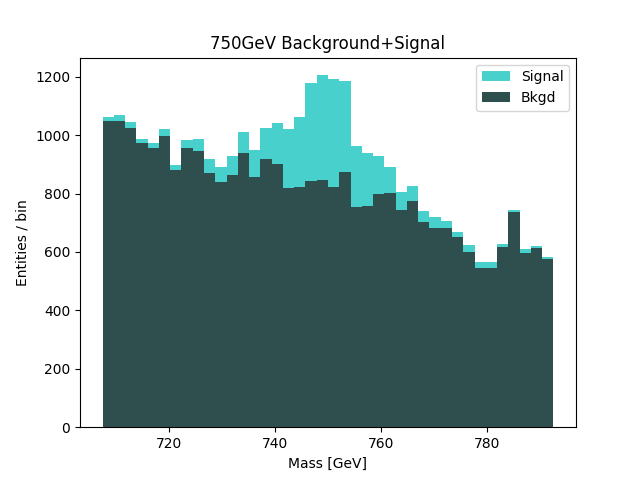
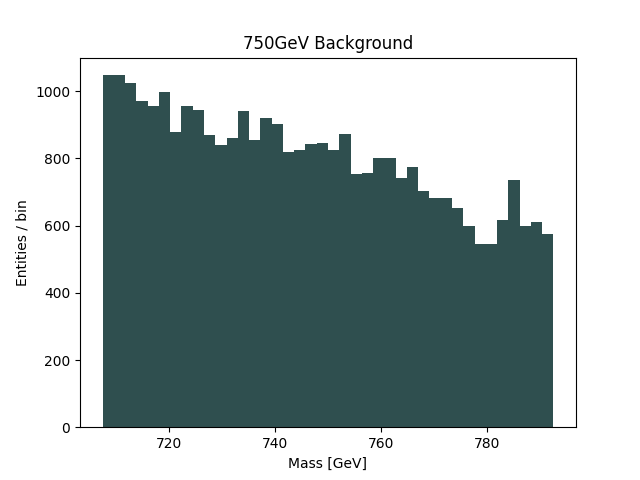
The statistical significance of finding Z` inside the 300GeV window is 48.961, and the particle would have a mass of (298.55 ± 8.7 GeV). This significance is approximately 21% smaller than the 200GeV case.

****

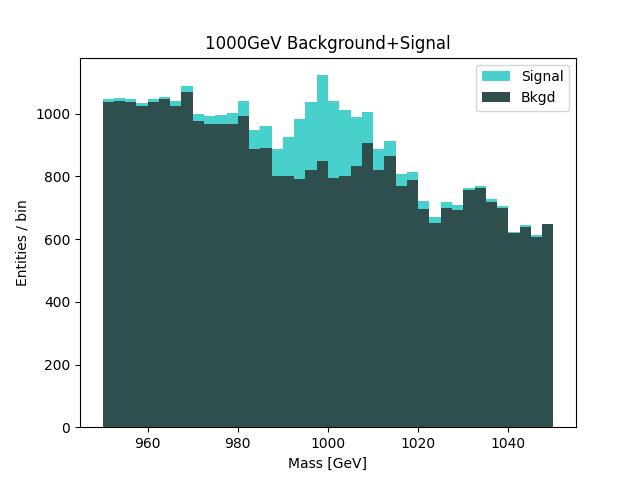
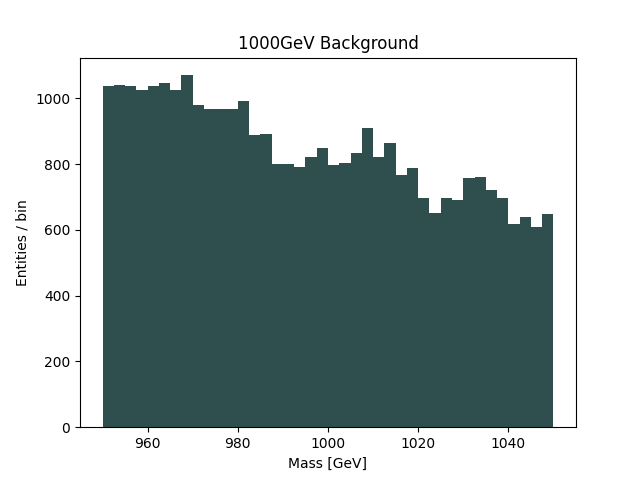
At 400 GeV, the statistical significance of Z` is 39.764 and would fall within a mass range of (400.00 ± 8.0 GeV). This significance is about 18% less than the 300GeV window.

****

The statistical significance of finding Z` within the 500GeV window is 30.26, approximately 23% smaller than the 400GeV window. Z`s invariant mass here at 500GeV would be (500.00 ± 8.8 GeV).

****

At 750 GeV, the statistical significance of finding Z` is 27.299, and Z`’s invariant mass would be (750.00 ± 8.5 GeV). Surprisingly, the statistical significance is only 10% less than the previous, 500 GeV window.

****

Finally, the statistical significance in the 1000 GeV case is 18.824, approximately 31% less than the 750GeV case. Here, Z’ would have an invariant mass of (997.50 ± 7.5) GeV.

The statistical significance of every mass window is substantial, ranging between 18 and 63. These significances are inversely proportional to the invariant mass of the window itself, where 200GeV displayed the greatest statistical significance and 1000GeV displayed the least. The peak number of signal entities is always located at each histogram’s central invariant mass, conveniently positioning there as well. The 200GeV window’s exhibited the highest Δ while the 1000GeV window exhibited the lowest. All other specific measurements in all other windows fall between the 200GeV and 1000 GeV measurements, therefore these specific measurements can be ignored.

**Discussion**

The probability Z` is located within any of the 6 invariant mass windows is high with the greatest significance at 200GeV. This is primarily because the sheer number of signal events is largest at 200GeV yet smallest at 1000GeV—this is visible by looking at the total area covered by signal data within each graph. Yet the histogram weights artificially scaled the number of signal events at each window, tripling the signal entities for 200GeV and halving them at 1000GeV. Since the statistical significance is based on the proportion of signal to background, the weights resultantly maximized the 200GeV significance and minimized the 1000GeV significance. This is problematic. Unfortunately, the histogram weights cannot be completely foregone; they are needed to surprisingly balance the number of signal and background entities in each histogram. The tripling of signal entities at 200GeV is intentional because, there, the number of background entities is greatest. Similarly, the 1000GeV window has the lowest number of background entities, hence, the signal is weighted to be lowest as well. The weights intentionally scale smaller as mass increases to account for the smaller number of background entities. Now it is uncertain whether or not the dataset’s simulation accounted for these specific weights, however, what can be certain is that the greatest statistical significance still remains at the 200GeV mass window.

Yet it is also at this same window the mass of Z` is most imprecise, as demonstrated by the largest Δ. This is due to Δ following the inverse trend of statistical significance, where Δ is maximized at 200GeV and minimized at 1000GeV. To understand why this trend in Δ occurs, first know that the difference between the largest and smallest Δ is merely 1.7GeV. Second, the size of each bin is approximately 2.3 GeV. Since Δ is based on the most optimal selection of bins, the largest and smallest Δ at each window only differs by one bin. The room for error here must be tremendous because a slight change in bin size or histogram scale should vastly change the corresponding Δ at that window. This issue is further compounded by the fact that histogram binning and scale were almost arbitrarily determined. The impact this overall issue had on the finalized Δ values is not readily understandable, however, all Δ values all fall within 1-5% of . Therefore, they do not substantially affect the expected invariant mass. There is also not a reason to believe arbitrary binning affected because the bin size is still fine enough to easily differentiate the minimums and maximums within each histogram. Hence the finalized statistical significance should not stray too heavily from the most optimal values.

Regardless, could the existence of a new elementary particle be asserted from similar results? This is difficult to determine since the investigation calculated significance using a non-standard formula. Furthermore, the predetermined histogram weights and arbitrary bin size limited certain calculations. Yet if Z` were to hypothetically exist, then the results demonstrate that Z` most likely has an invariant mass of (201.3 ± 9.2) GeV. Despite these measurements and limitations, I argue there is enough information within these types of datasets for the LHC to detect a new elementary particle. However, any such detection must also be supported by an impeccable physical theory before the scientific community should consider these proclamations statistically significant.