Modeling Quark Compositeness at the Compact Muon Solenoid

Amanda Farah - Supervisor: Dr. Lenny Spiegel

Fermi National Accelerator Laboratory

Abstract

Compositeness is the idea that quarks and leptons may have structure. It was proposed in order to solve some inconsistencies with the standard model. Many compositeness models predict an interaction that can be approximated as a contact interaction that produces two leptons. This study uses PYTHIA simulations to develop analysis methods to distinguish these contact interactions from their main backgrounds. An approximation of a limit on the energy scale associated with contact interactions was found to be 19.01 TeV in the case that contact interactions are not found.

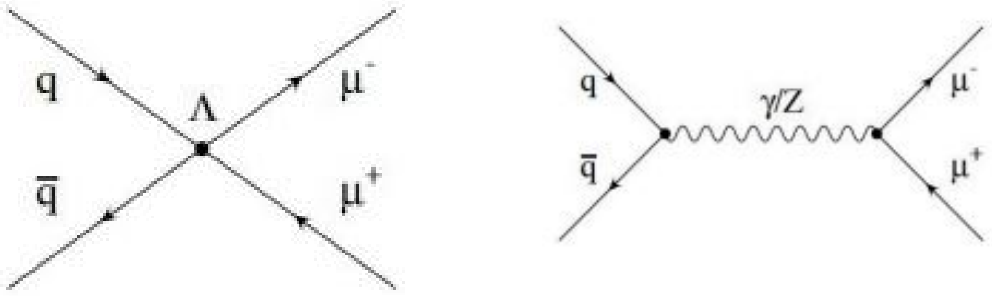
Keywords: Compositeness, Contact interaction, Dilepton, CMS

1. Introduction

Currently, the standard model states that quarks and leptons are point particles that they take up zero space and have no internal structure. There are, however several characteristics of the standard model that could be explained if this were not true. That is, if quarks and leptons were made up of smaller particles, called preons.

A notable pattern that preons could explain is the existence of generations of matter. All matter that we readily observe in our universe is made up of first generation matter (up and down quarks and electrons), so there is no apparent reason for the other two generations. If preons existed, the heavier generations could be explained as excited states of the first one. This is one

Email address: afarah@sas.upenn.edu (Amanda Farah - Supervisor: Dr. Lenny Spiegel)



(a) Contact Interaction

(b) Drell-Yan

Figure 1: Contact Interaction and Drell-Yan Feynman diagrams. The contact interaction (right) is predicted by compositeness models. It has a significant standard model background, however. The Drell-Yan process (left) also produces two leptons (muons, in this case) with a similar, but not identical, mass spectrum. In the case where the final muon helicity states of DY and CI are the same, there is a quantum mechanical interference between the Drell-Yan and Contact Interaction processes. Helicity is explained below

of many ideas motivating the search for compositeness the idea that quarks and leptons are made of preons.

Some compositeness models predict a particle interaction that CMS is capable of detecting. This can be approximated as a contact interaction (CI) in which two preons from a quark-antiquark pair make contact and produce one positively charged and one negatively charged lepton (this is also called a dilepton pair). As shown in Figure 1, this process has a significant background in the standard model Drell-Yan (DY) process.

Different compositeness models have different signatures than others. Firstly, CI could interfere constructively or destructively with DY, leading to more or less dilepton events, respectively. Secondly, for both constructive and destructive interference, the muons produced could have different helicities. Helicity is the projection of spin onto the direction of momentum, and is usually called left-handed (L) or right-handed (R). Thus, there are three different possible muon helicity states: LL, RR, and LR. LL and RR create nearly identical signatures for all of the discussed probes of CI, but LR differs in some ways.

As of now, there is no evidence for quark compositeness. This may be because preons are bound together so tightly in quarks that no collider has yet been able to break them apart to produce contact interactions. This idea

of a preon binding energy is very important to compositeness models. We define an energy Λ that is closely related to the binding energy. If it is not possible to find evidence of compositeness, the goal then becomes to set a lower limit on Λ.

This study focuses on using simulations to develop methods to differentiate contact interactions from Drell-Yan in the case where the final leptons are muons. It aims to use simulated data to develop analysis methods that will be applied to real data when available. The simulation software used is PYTHIA 8, and the analysis software used is ROOT.

This paper will first outline the methods used to find CI and how these methods were developed. It will then show the results of these methods being applied to simulated data. Lastly, conclusions and further study will be discussed.

2. Methods

Contact interactions can be found and differentiated from Drell-Yan in two different ways. One of these is to look at the dimuon invariant mass spectrum. This is the distribution of the mass of the two muon system after collision. An example is shown in Figure 2. If contact interactions exist, more high mass events will occur, and the upper tail of this distribution will be higher than that for just DY. Due to this phenomenon, it should be more possible to distinguish CI from DY at higher masses. For this reason, it is typical to make lower mass cuts in simulation (i.e. to only generate events that create dimuons with an invariant mass above a certain point) in order to cut out areas where CI is not differentiable from DY and obtain high statistics for high mass events.

The other method to distinguish contact interactions from DY is through the Collins-Soper frame angle distribution. This will be explained in Subsection 2.2.

2.1. Verification of Physics in Simulation

The simulation software used in this study was PYTHIA 8.215. It is an event generator used for high energy physics events. In order to simulate the entire detector using PYTHIA 8, which takes multiple CPU days on a large cluster, it is first necessary to verify that PYTHIA accurately approximates contact interactions.

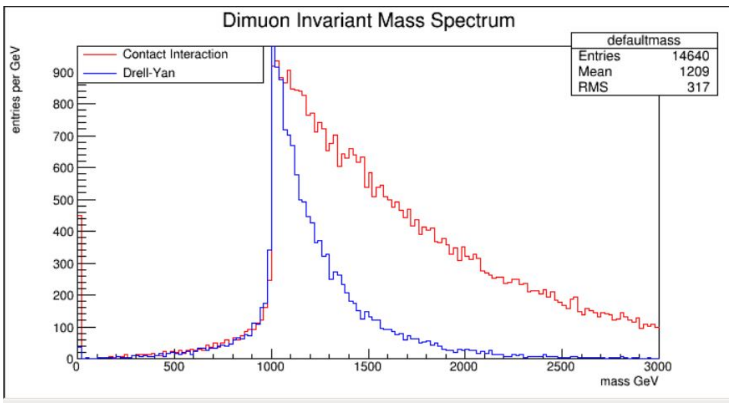


Figure 2: Example of invariant mass spectra. These are the distributions of invariant masses for several thousand events. This is a drastic example of the difference between the contact interaction (red) and Drell-Yan invariant mass spectra. Note how the right tail of the CI distribution is far thicker than that of the DY. In this simulation, Λ is set to be low, so CI happen very frequently at the energies achieved by the LHC, and many high-mass events occur. This makes the stark contrast between DY and CI distributions seen in this example. In reality, Λ values lower than 20 TeV in standard models for dilepton final states have been ruled out experimentally, and one would not expect to see a difference between DY and still possible CI Λ values until the integrated luminosity of the LHC experiments has increased significantly, leading to many more very high mass entries in the dilepton mass spectra.

One important check of this approximation is to ensure that infinite Λ contact interaction models give the same results as the standard model after analysis. This should be the case because if Λ is infinite, quarks would never dissociate into preons and contact interactions would never occur, so the only interaction detected would be the standard model processes. Thus, it is possible to check the accuracy of PYTHIA in modeling contact interactions by verifying that infinite Λ PYTHIA simulations look almost entirely like Drell-Yan simulations. This was done previously for PYTHIA 6 by Sowjanya Gollapinni in 2011. The variable being compared between DY and CI is the cross section of interaction, σ.

The physics verification process was accomplished in three stages. First, the DY events from the PYTHIA 6 study were recreated. Then, events were generated in PYTHIA 8 and the PYTHIA 8 code was modified until the results matched the earlier PYTHIA 6 study's results. Then, PYTHIA 8 Drell-Yan results were compared to PYTHIA 8 high-Λ contact interaction results and scripts were modified until there was an agreement at the 99% level.

2.1.1. Recreation of PYTHIA 6 Study

[1] used PYTHIA 6.4 to generate 25,000 events at four minimum mass cuts for both Drell-Yan only as well as infinite Λ contact interaction events. The study simulated a 7 TeV proton center of mass energy. These parameters were recreated using PYTHIA 6.426, though only 10,000 events were run per mass cut in order to save time. The resulting cross sections are shown in Table 1. The percent differences between the two sets of PYTHIA 6 results were too large to confirm that the studies were identical. It was determined that more statistics was needed to establish that the studies were different, however. This was a favorable approach because when the code that produced the earlier study was compared to the code written for this study, they were found to be identical in all functional ways. There was no known way to produce more than 25,000 events using PYTHIA 6, as the network running the jobs (`cmslpc`) canceled them after more than a few hours of running, and it is not possible to submit PYTHIA 6 jobs to the batch system at the LPC (`CONDOR`). Thus, this study was deemed inconclusive and the focus was turned to matching PYHTIA 8 results to the earlier PYTHIA 6 study.

Low Mass Cut (GeV)	My σ (fb)	Sowjanya's σ (fb)	% Difference
45	691300	779900	12.0%
120	8151	7883	3.34%
500	27.39	26.89	13.7%
1000	1.115	0.9653	14.4%

Table 1: Comparison of standard Drell-Yan cross sections between two different PYTHIA 6 studies. The left-middle column has 25k events per data point while the right-middle column has 100k events per data point.

PYTHIA 8 σ (nb)	2011 PYTHIA 6 σ (nb)	Percent Difference
7.725	7.799	0.953%

Table 2: Comparison between Drell-Yan events in PYTHIA 6 and PYTHIA 8.

2.1.2. PYTHIA 6 and 8 Continuity

It was necessary to ensure that PYTHIA 6 and PYTHIA 8 modeled the standard model background similarly, so a PYTHIA 8 script was created to emulate the study done in [1]. Both the PYTHIA 8 script and 2011 PYTHIA 6 studies ran 100,000 events at a 7 TeV center of mass energy. The point of comparison was the cross section of Drell-Yan events with a lower mass cut of 45 GeV. The results are shown in Table 2. The two studies agreed at the 99% level, so the processes in PYTHIA 6 and 8 for DY were deemed to be nearly identical. PYTHIA 8 then became the main focus of this study.

2.1.3. Drell-Yan vs. Contact Interactions at "Infinite" Λ in PYTHIA 8

Unlike PYTHIA 6, PYTHIA 8 has no option for specifying infinite Λ. Thus, a very high value of Λ is chosen to approximate infinity. In this case, it was determined by overlaying mass spectra and comparing cross sections that Λ = 1,000 TeV was high enough. Two PYTHIA 8 scripts were run for this verification. These scripts only differed in the specification of contact interaction events or just Drell-Yan events. Both scripts ran 50,000 events at each mass cut with a center of mass energy of 13 TeV to emulate 2016 LHC energies. The resulting cross sections for different mass cuts are shown in Table 3. All of the percent differences in this table are less than 1%, so it was concluded that PYTHIA 8 handles contact interactions correctly. To ensure that these results were not a statistical phenomenon, the 500 GeV minimum mass cut was retested with 100,000 events. The percent difference for this test was 0.067%, which is low enough to conclude that CI and DY

are essentially the same for this mass cut. To test that the assumption that Λ=1000 TeV was high enough to model infinite Λ, a higher Λ value was tested. This was for a minimum mass of 500 GeV and $\Lambda = 10^5$ TeV. The percent difference for this test was 0.650%, not significantly different than Λ=1,000 TeV. Thus, the earlier assumption was a good one.

Low mass cut (GeV)	**Drell Yan σ (fb)**	**Contact Interaction σ (fb)**	**Percent Difference**
120	17800	17820	0.11%
200	2379	2383	0.17%
500	88.93	89.23	0.34%
1000	5.510	5.558	0.87%
1800	0.3176	0.3160	0.505%

Table 3: Cross sections of contact interactions and Drell-Yan interactions at different minimum mass cuts. Run in PYTHIA 8 with 50,000 events and Λ=1000TeV.

After these tests, it was concluded that PYTHIA 8 handled CI and DY events well enough to model the products of these events propagate through the whole CMS detector. A full-detector simulation of contact interactions was requested for various values of Λ, multiple mass cuts and different CI models. An example of one of the submitted PYTHIA 8 fragments can be found in Appendix B.

2.2. Collins-Soper Frame Angle

Once the full-detector simulation is completed and once run data is available, there need to be multiple methods of searching for CI in these data. Invariant mass spectra are a big part of one of these methods. Another analysis is helped by the Collins-Soper Frame Angle. The Collins-Soper (CS) frame is the center of mass frame of the dimuon system. This is not to be confused with the collision center of mass frame. The CS angle (θ) is the angle that the negative muon makes with the z axis in the CS frame. A more thorough explanation of θ is given in Appendix A, including a derivation of the equation for θ in terms of variables that CMS can detect.

Generally, the cosine of θ is the quantity that is most frequently discussed in these analyses.The distribution of $\cos\theta$ is asymmetric about zero. There are generally more events with a positive value for $\cos\theta$ than with a negative one. As is visible in Figure 5, this asymmetry is more pronounced for

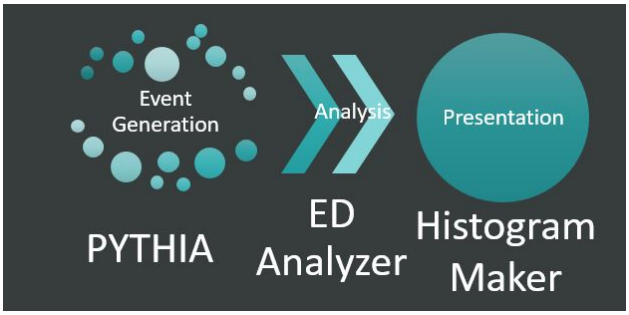


Figure 3: Outline of steps necessary to create and analyze events.

Drell-Yan events than for LR contact interaction events. Determining the asymmetry of $\cos\theta$ for dimuon events can aid in differentiating between CI and DY in the case where the LR model is true.

Thus, it is necessary to define a quantity called the forward-backward asymmetry, A_{FB}, which is a measure of the asymmetry of the CS angle distribution[4]. This quantity is given by $A_{FB} = \frac{N_F - N_B}{N_F + N_B}$ where N_F is the number of events with $\cos\theta > 0$ and N_B is the number of events with $\cos\theta < 0$. CI and DY are most differentiable at high mass events. We therefore plot $A_F B$ as a function of mass. As is clear in Figure 6, this allows for a clear distinction between DY and CI if enough high mass events are detected.

2.3. Workflow

In order to simulate and analyze events, several steps were needed. The basic outline of this is depicted in Figure 3. To test and modify PYTHIA scripts and functions, a PYTHIA script was written and used to generate events. This script outputs an nTuple in the form of a ROOT file, which will now be referred to as `file1.root`. Then, `file1.root` is input into the `EDAnalyzer`. This parses the TTree in `file1.root`, extracts relevant data, and analyzes it. The `EDAnalyzer` generates another nTuple, called `file2.root` with an invariant mass plot, Collins-Soper frame angle, and other useful values and plots for analysis. Then, `file2.root` is run though a histogram maker to compare, overlay, or polish the plots it contains. The user can create various histogram maker scripts depending on their needs and the variable being analyzed.

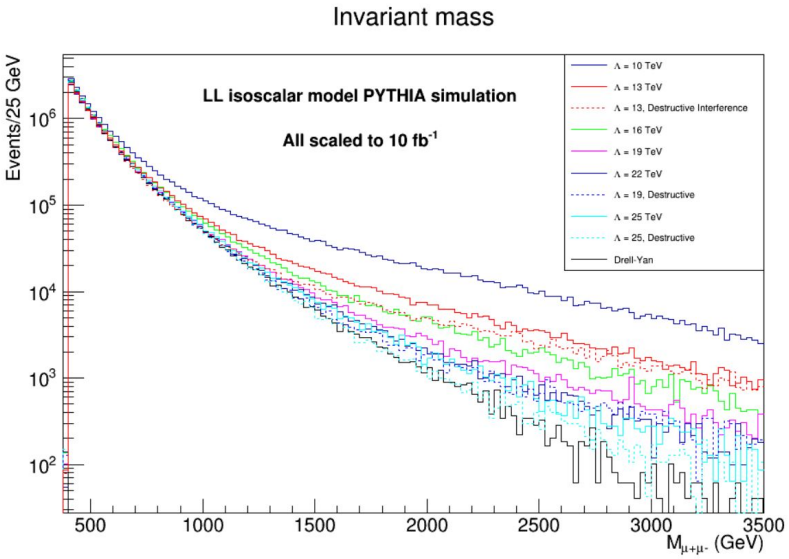


Figure 4: . Simulated dimuon invariant mass spectrum for varying values of Λ and standard model Drell-Yan background. Solid lines denote the constructive interference model while dashed lines are destructive interference. Each of these spectra is scaled to 10 fb^{-1} and were originally created with 1 million simulated events with a minimum mass cut of 400 GeV and using the left-left isoscalar model.

3. Results

By the end of this study, the validity of physics processes relating to compositeness in PYTHIA 8 was confirmed. This is exemplified in Figure 4 where it is possible to see that the invariant mass spectra of CI converge to that of DY as Λ increases. Because of this confirmation, it was acceptable to submit a full-detector simulation request for various CI models. Analysis methods were developed to process the results of this simulation as well as data from CMS. One of these methods employs the asymmetry of the Collins-Soper frame angle, as explained in Section 2.2. The result of applying this analysis to DY, the LL, and LR models generated in PYTHIA with 25,000 events, a minimum mass cut of 400 GeV, and Λ set to 14 TeV is shown in Figures 5 and 6.

The development of a tool to use the invariant mass spectra to determine a lower limit on Λ was also initiated. There preliminary results were obtained using this tool, though much work is needed before it is ready to process data.

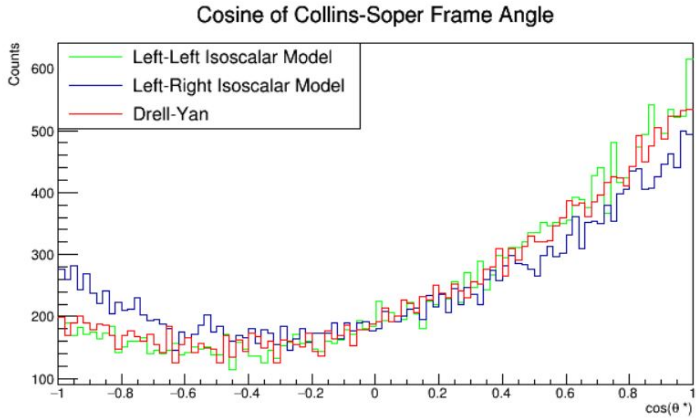


Figure 5: Collins-Soper angle distribution for DY and two different CI models. The RR model overlaps exactly with the LL model so is not shown. Note the left-skewedness of this distribution. It is this asymmetry that differentiates the LR model from the DY background. This plot contains 25,000 simulated events per model shown with a minimum mass cut of 400 GeV. For the CI events, Λ=14 TeV and the interference is constructive.

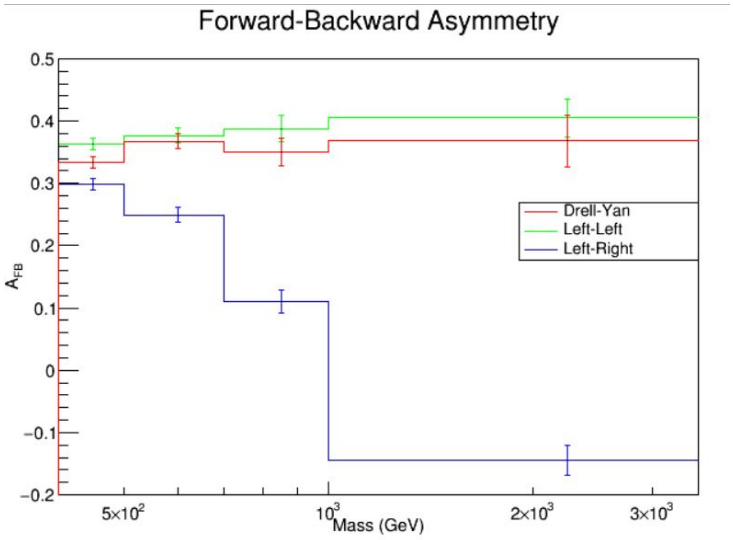


Figure 6: Forward-Backward Asymmetry as a function of mass. Note that the LL and DY values are within each others errors and also do not change drastically as a function of mass. LR, on the other hand, deviates greatly from the other two at high masses - even having asymmetry in the other direction at masses above 1 TeV. RR is not shown as it is identical to LL. This plot was generated using the same simulated events as in Figure 5.

Λ (TeV)	16	17	18	19	20	21	22	23	24	25
$\beta(TeV^{-2})$	0.00391	0.00346	0.00309	0.00277	0.00250	0.00227	0.00207	0.00189	0.00174	0.00160
CI Yield (events)	92.7	80.2	75.1	70.2	68.2	65.7	62.493	62.2	60.9	64.157
DY Yield (events)	52.095	52.095	52.095	52.095	52.095	52.095	52.095	52.095	52.095	52.095
Difference in Yield (events)	40.6	28.1	23.0	18.1	16.1	13.6	10.4	10.1	8.8	12.1
r (unitless)	0.440	0.635	0.777	0.988	1.113	1.316	1.719	1.773	2.039	1.481

Table 4: Analysis of background to signal ratio for integer values of Λ. Yields are scaled by cross section to correspond to $10fb^{-1}$ of data.

3.1. Lower Limit on Λ

In the case where contact interactions are not found, it is then possible to at least determine the lowest energy scale Λ at which they can still occur. This is done by performing a counting experiment on the number of events above a certain mass cut as a function of Λ and using the calculated yield for DY as the background. The Higgs Combine Tool was used for this statistical analysis with the number of observed events approximated by the DY yield. First, yield (number of events) was plotted as a function of β where $\beta = 1/\Lambda^2$. β is used because the cross section of interaction for CI is defined in terms of multiples of $1/\Lambda^2$. This plot is done for various different mass groups and is shown in Figure 7. This is then fitted to a degree-2 polynomial to obtain yield as a continuous function of β. Yields are then extracted from this function and used as rates in the Higgs Combine Tool to obtain values of the background to signal ratio (r) at integer values of Λ. The results of this are shown in Table 4 and the datacard used for the Higgs Combine Tool is shown in Appendix C. When the number of events from the background equals the number of predicted events from the signal within 95% confidence, r=1. To find the exact value of Λ that corresponded to r=1, r was then plotted as a function of Λ, as shown in Figure 8. This value of Λ was found to equal 19.01 TeV. This is an approximation of the lower limit on Λ if contact interactions are not found.

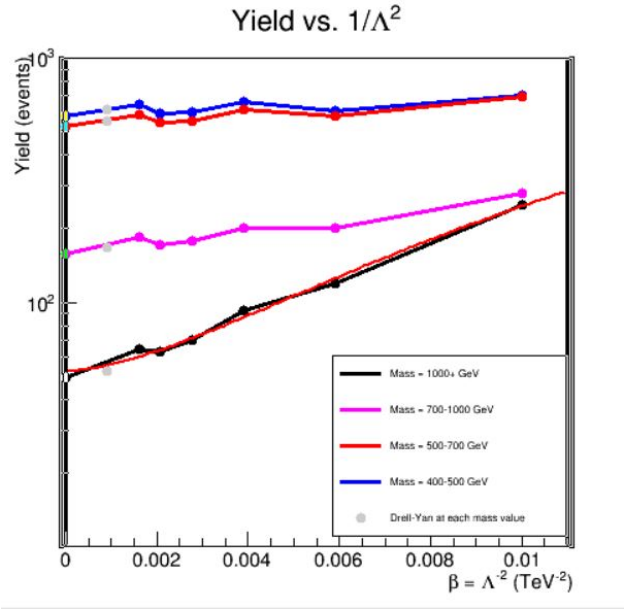


Figure 7: Yield (number of events scaled by cross section) vs. β for different mass groups. Each value of Λ corresponds to 1 million events being run with a minimum mass cut of 400 GeV.

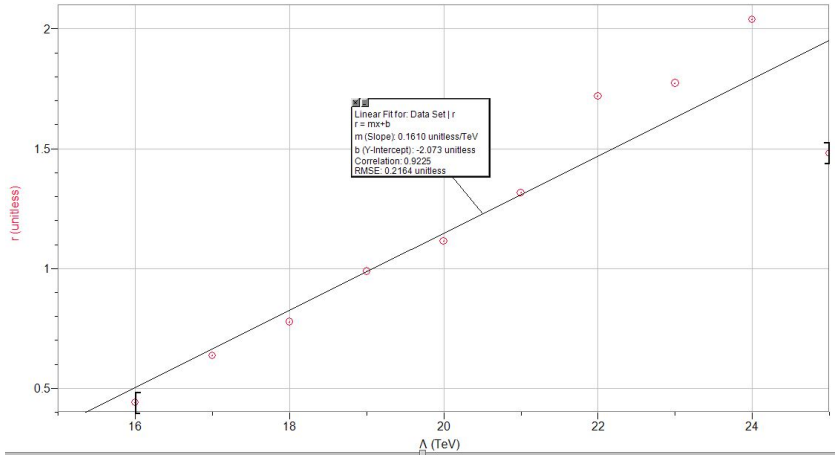


Figure 8: Background to signal ratio (r) as a function of Λ. The best fit line intersects r=1 at Λ = 19.01 TeV.

4. Conclusion

More development needs to be done on the Higgs Combine Tool in order to perform a shape analysis on multiple mass groups with a maximum likelihood fit, rather than a simplified counting experiments on just one mass group. In addition, more systematic uncertainties need to be taken into account while running this tool. The analysis methods developed during this study need to be applied to real data when it is available in a few months. This will play a part in understanding or constraining the hypothesis of compositeness.

5. Acknowledgments

First, I would like to thank my supervisors, Dr. Lenny Spiegel and Dr. Pushpa Bhat, for their daily support and help throughout this project. I would also like to thank Dr. Steve Mrenna, an author of PYTHIA, who explained a lot of the software to me. Shawn Zaleski also played a huge role in the development of the analysis code used for this project and was an excellent teacher of all topics covered in this project. I am grateful to Judy Nunez, Sandra Charles, Dr. Elliot McCrory, Mayling Wong-Squires, Charles Orozco, and the rest of the SIST program for allowing me to do research at Fermilab.

6. References

[1] Gollapinni, Sowjanya, 2011. `https://twiki.cern.ch/twiki/pub/CMS/CI2dimuonsEXO11009Presentations/DYAccuracy_ZandMSTP.pdf`

[2] Kottachchi, Chamath. "SEARCH FOR CONTACT INTERACTIONS USING THE DIMUON MASS SPECTRUM IN P-P COLLISIONS AT S = 8 TeV AT CMS." Diss. Wayne State U, 2014.

[3] Nagashima, Yorikiyo. 2013. Elementary Particle Physics Vol.2: Foundations of the Standard Model. Weinheim, Germany: WILEY-VCH.

[4] The ATLAS Collaboration. 2014. Search for contact interactions and large extra dimensions in the dilepton channel using proton-proton collisions at s = 8 TeV with the ATLAS detector. Eur. Phys. J. C. (2014) 74:3134

Appendix A. Collins-Soper Frame Angle Derivation

This appendix was created while referencing [3] heavily.

Before deriving the CS frame angle, it is first necessary to review some basics of Lorentz transformations and to define a new variable.

Appendix A.1. Lorentz Boost of Momentum 4-vector in the z direction

Call $\vec{P}$ a momentum four-vector such that $\vec{P} = \begin{bmatrix} E \\ p_x c \\ p_y c \\ p_z c \end{bmatrix}$. And let $\vec{P}'$ be the momentum four vector Lorentz boosted along the z-axis. $\vec{P}'$ is then given by $\vec{P}' = \vec{L}(\beta\hat{z})\vec{P}$ where $\vec{L}(\beta\hat{z})$ is the Minkowski matrix for a transformation in the z direction and is given by

$$\vec{L}(\beta\hat{z}) = \begin{bmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{bmatrix} \tag{A.1}$$

where $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$. Then,

$$\vec{P}' = \begin{bmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} E \\ p_x c \\ p_y c \\ p_z c \end{bmatrix} = \begin{bmatrix} \gamma E - \beta\gamma p_z c \\ p_x c \\ p_y c \\ -\gamma\beta E + \gamma p_z c \end{bmatrix}$$

$$\Rightarrow E' = \gamma E - \beta\gamma p_z c \tag{A.2a}$$

$$p_z' c = \gamma p_z c - \beta\gamma E \tag{A.2b}$$

Appendix A.2. Rapidity

Let us define an angle y such that $y \equiv \frac{1}{2}\ln(\frac{E+p_zc}{E-p_zc})$. We will call this *rapidity*. It then follows that

$$\begin{aligned}
y =& \frac{1}{2}\ln(\frac{E+p_zc}{E-p_zc}) \\
=& \frac{1}{2}\ln(\frac{(1+v_z/c)}{(1-v_zc)}\frac{\gamma mc^2}{\gamma mc^2}) \\
2y =& \ln(\frac{1+\beta}{1-\beta}) \\
e^{2y} =& \frac{1+\beta}{1-\beta} \\
(1-\beta)e^{2y} =& \beta+1 \\
e^{2y}-1 =& e^{2y}\beta+\beta \\
\frac{e^{2y}-1}{e^{2y}+1} =& \beta \\
\frac{\sinh(y)}{\cosh(y)} =& \beta
\end{aligned}$$

$$\tanh(y) = \beta \tag{A.3a}$$
$$\cosh(y) \equiv \gamma \tag{A.3b}$$
$$\Rightarrow \sinh(y) = \beta\gamma \tag{A.3c}$$

Equations A.2 then become

$$E' = \cosh(y)E - \sinh(y)p_zc \tag{A.4a}$$
$$p'_zc = \cosh(y)p_zc - \sinh(y)E \tag{A.4b}$$

Appendix A.3. Collins-Soper frame angle θ in terms of lab frame variables

Let us call k_1 and k_2 the four momenta of leptons one and two. These are just like the four momenta described in Appendix A.1. Note that here we are calling momentum k, not p. Then, $Q = \sqrt{(k_1+k_2)(k_1+k_2)}$ is the dilepton momentum in the lab frame. From these, we have Q_T is the transverse momentum of the lepton pair relative to the beam (z) axis. We then define four variables, $k^{\pm}_{1,2}$ as $k^{\pm} = (E \pm k_z)/\sqrt{2}$.

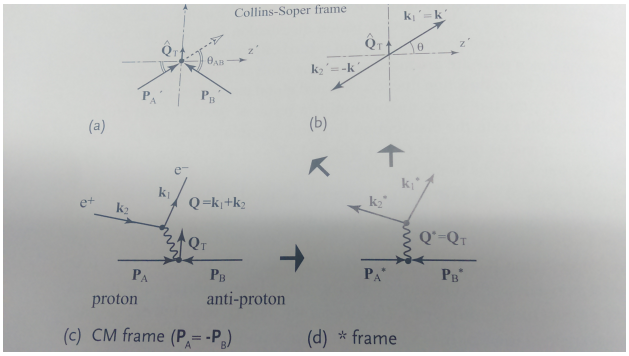


Figure A.9: Kinematics in the (c) center of mass frame, (d) intermediate (*) frame, and (a,b) dimuon rest frame (CS frame). Figure taken from [3]

We wish to determine the rest frame of the dimuons. We do this in two steps, as shown in Figure A.9. We start in the proton-proton center of mass (CM) frame and Lorentz boost along the beam axis (z axis) to a frame in which the dimuon momentum in the z direction is zero. We call this the * frame. In other words, $Q^*_z = 0$ or $Q^* = Q_T$. Since Q_T is perpindicular to the beam axis, it does not change under this transformation, and $Q^*_T = Q_T$. This Lorentz transform is shown in Figure A.9c to A.9d. Now, we boost in the $-Q_T$ direction. To get into the dimuon rest frame, we set $\gamma = \frac{\sqrt{Q^2+Q_T^2}}{2}$ and $\beta\gamma = \frac{|Q_T|}{Q}$. Boosting using these parameters puts us in the dimuon rest frame, or Collins-Soper (CS) frame. Variables specific to the CS frame are denoted with a '.

k_1, k_2, and Q are all manifestly covariant under rotation about the z' axis. The magnitude of each lepton's momentum in the CS frame is $k = |\mathbf{k}| = Q/2$. Then, The four momenta in the CS frame are:

$$k_1' = \begin{bmatrix} k \\ k\sin(\theta) \\ 0 \\ k\cos(\theta) \end{bmatrix}, k_2' = \begin{bmatrix} k \\ -ksin(\theta) \\ 0 \\ -kcos(\theta) \end{bmatrix}$$

The x'-z' plane is then defined by $\boldsymbol{P_A, P_B, Q}$. We can describe the four momenta in the * frame by a Lorentz boost back in the positive $\boldsymbol{Q_T}$ direction. Following equations A.2, we get

$$k_1^* = \begin{bmatrix} \gamma k(1+\beta\sin\theta) \\ \gamma k(\sin\theta+\beta) \\ 0 \\ k\cos\theta \end{bmatrix}, k_2^* = \begin{bmatrix} \gamma k(1-\beta\sin\theta) \\ \gamma k(-\sin\theta+\beta) \\ 0 \\ -k\cos\theta \end{bmatrix}$$

Now, we write out the variables $k^{\pm}$ (defined above) in this frame.

$$\begin{aligned} k_1^{+*} &= (E_1^* + k_{1z}^*)/\sqrt{2} = k[\gamma(1+\beta\sin\theta)+\cos\theta]/\sqrt{2} \\ k_1^{-*} &= (E_1^* - k_{1z}^*)/\sqrt{2} = k[\gamma(1+\beta\sin\theta)-\cos\theta]/\sqrt{2} \\ k_2^{+*} &= (E_2^* + k_{2z}^*)/\sqrt{2} = k[\gamma(1-\beta\sin\theta)-\cos\theta]/\sqrt{2} \\ k_2^{-*} &= (E_2^* - k_{2z}^*)/\sqrt{2} = k[\gamma(1-\beta\sin\theta)+\cos\theta]/\sqrt{2} \end{aligned} \tag{A.5}$$

Now for some algebra.

$$\begin{aligned} k_1^{+*}k_2^{-*} - k_1^{-*}k_2^{+*} = & \\ & (k^2/2)[((\cos\theta+\gamma)^2 - \beta^2\gamma^2\sin^2\theta) - ((-\cos\theta+\gamma)^2 - \beta^2\gamma^2\sin^2\theta)] \\ & = 2k^2\gamma\cos\theta \\ & = 2(\frac{Q}{2})^2\frac{\sqrt{Q^2+Q_T^2}}{Q}\cos\theta \\ \therefore \cos\theta = & \frac{2}{Q\sqrt{Q^2+Q_T^2}}(k_1^+k_2^- - k_1^-k_2^+)^* \end{aligned} \tag{A.6}$$

θ is the Collins-Soper frame angle. The first factor is made up of variables that are measurable in the lab frame:the dimuon invariant mass and transverse momentum. The second facor, however, is not. We can prove that this factor is invariant under z boosts, and since the * frame is only boosted along the z axis from the lab frame, show this expression in terms of lab frame variables. The equation for the Collins-Soper frame angle θ is then given by

$$\cos(\theta) = \frac{2}{Q\sqrt{Q^2+Q_T^2}}(k_1^+k_2^- - k_1^-k_2^+) \tag{A.7}$$

The proof that $(k_1^+k_2^- - k_1^-k_2^+)$ is a relativistic invariant along z axis boosts follows.

From A.4, we know that

$$\begin{aligned} E^* &= E\cosh y - k_z\sinh y \\ k_{1z}^* &= -E\sinh y + k_z\cosh y \end{aligned} \tag{A.8}$$

Then,

$$\begin{aligned} k_1^{+*} &= (E_1^* + k_{1z}^*)/\sqrt{2} \\ &= (E_1 + k_{1z})\cosh y - (E_1 + k_{1z})\sinh y \\ k_2^{*-} &= (E_2^* - k_{2z}^*)/\sqrt{2} \\ &= (E_2 - k_{2z})\cosh y + (E_2 - k_{2z})\sinh y \\ k_1^{+*}k_2^{-*} &= (E_2 - k_{2z})(E_1 + k_{1z})(\cosh^2 y - \sinh^2 y) \end{aligned} \tag{A.9}$$

It is a trigonometric identity that $\cosh^2\psi - \sinh^2\psi = 1$, so it follows that $k_1^{+*}k_2^{-*} = (E_2 - k_{2z})(E_1 + k_{1z})$. Similarly, $k_1^{-*}k_2^{+*} = (E_1 - k_{1z})(E_2 + k_{2z})$. Thus, it follows that

$$\begin{aligned} (k_1^+k_2^- - k_1^-k_2^+)^* &= (E_2 - k_{2z})(E_1 + k_{1z}) - (E_1 - k_{1z})(E_2 + k_{2z}) \\ &= k_1^+k_2^- - k_1^-k_2^+ \end{aligned} \tag{A.10}$$

Therefore, the quantity $k_1^+k_2^- - k_1^-k_2^+$ is invariant under z boosts and equation A.7 is written entirely in lab frame variables.

Appendix B. Full Detector Simulation Card Example

```
import FWCore.ParameterSet.Config as cms

from Configuration.Generator.Pythia8CommonSettings_cfi import *
from Configuration.Generator.Pythia8CUEP8M1Settings_cfi import *

generator = cms.EDFilter("Pythia8GeneratorFilter",
    maxEventsToPrint = cms.untracked.int32(1),
    pythiaPylistVerbosity = cms.untracked.int32(1),
    filterEfficiency = cms.untracked.double(1.0),
    pythiaHepMCVerbosity = cms.untracked.bool(False),
    comEnergy = cms.double(13000.),
    PythiaParameters = cms.PSet(
        pythia8CommonSettingsBlock,
        pythia8CUEP8M1SettingsBlock,
        processParameters = cms.vstring(
            'SoftQCD:nonDiffractive = on',
            'SoftQCD:singleDiffractive = on',
            'SoftQCD:doubleDiffractive = on',
```

```
            'ParticleDecays:limitTau0 = on',
            'ParticleDecays:tauMax = 10',
            'Tune:pp 5',
            'Tune:ee 3',
            'ContactInteractions:QCffbar2mumubar = on',

            'ContactInteractions:Lambda = 22000',
            'ContactInteractions:etaLL = 0',
            'ContactInteractions:etaRR = 1',
            'ContactInteractions:etaLR = 0',

            '23:onMode = off',
            '23:onIfAny = 13',
            'PhaseSpace:mHatMin = 300',


        ),
        parameterSets = cms.vstring('pythia8CommonSettings',
                                    'pythia8CUEP8M1Settings',
                                    'processParameters',
                                    )
    )
)
```

Appendix C. Sample Datacard for Higgs Combine tool

```
# Simple counting experiment, with one signal and a background process
imax 1  number of channels
jmax 1  number of backgrounds
kmax 1  number of nuisance parameters (sources of systematical uncertainties)
------------
# we have just one channel, in which we observe 52 events
bin 1
observation 52
------------
# now we list the expected events for signal and all backgrounds in that bin
# the second 'process' line must have a positive number for backgrounds, and 0 for signal
# then we list the independent sources of uncertainties, and give their effect (syst. error)
# on each process and bin
bin              1     1
process         CI     DY
process          0     1
rate           12.062   52.095
------------
lumi   lnN       -     1.10     arbitrarily assigned
~
```