Setting limits on quark Contact Interaction scales from ATLAS inclusive jet cross–section measurements

Oscar Tovey Garcia

Supervisor: Tancredi Carli

September 2021

Abstract

Inclusive jet cross–sections are one of the basic measurements from proton–proton (pp) collider experiments such as the LHC, and can be used to test the Standard Model and search for new physics. We focus on the contact interaction (CI), a parametrisation of a potential beyond Standard Model process involving the flavor-symmetric $q\bar{q} \rightarrow q\bar{q}$ interaction via an unknown intermediate state. Using APPLGRID and CIJET, next–to–leading order (NLO) inclusive jet cross–sections were computed for QCD and for several contact interaction models at different scales Λ_{CI}, and exclusion limits were investigated by hypothesis testing on individual bins with the CL_s method. Further investigation, accounting for correlated systematic uncertainties across multiple bins, was carried out using the `pyhf` Python package to evaluate asymptotic formulas. Lower exclusion limits at the 95% confidence level for identical chirality color singlet $q\bar{q}$ pairs were found to be $\Lambda_{\mathrm{CI}} = 22.5$ TeV (single-bin CL_s) and $\Lambda_{\mathrm{CI}} = 26.9$ TeV (`pyhf`). Weaker limits were found for mixed chirality and color octet states.

1 Introduction

The interaction of two quarks to form another two quarks while obeying conservation laws is one of the simplest problems in quantum chromodynamics (QCD), and yet it offers a wealth of insight into the nature of matter in our Universe. One way to study such processes is by measuring the inclusive jet cross–sections from proton–proton collider experiments, such as the Large Hadron Collider (LHC). Jets produced in such collisions are classified into transverse momentum (p_T) and rapidity (y) bins, and the counts obtained for each bin are divided by the collider beam luminosity and the bin widths to obtain the inclusive jet cross–section measurements. This can also be modelled theoretically using Monte Carlo algorithms, and the predictions compared to the data to test the Standard Model, as in Figure 1 [1].

Inclusive jet cross–sections are also important measurements for probing processes beyond the Standard Model. One such process is the four quark contact interaction (CI), an effective field theory where two quarks interact by some unknown process parametrised by a scale Λ_{CI}. This is analogous to Fermi's model of β decay, which preceded electroweak theory. Contact interactions can take many forms to account for different chiral and color structures.

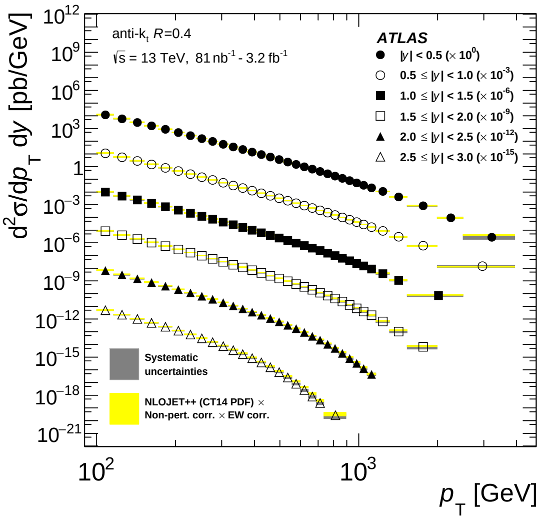

Figure 1: **QCD inclusive jet cross–sections.** A comparison of measurements and QCD calculations of inclusive jet cross–sections by the ATLAS experiment [1] over a range of p_T and rapidity bins. Very good agreement is found, despite the cross–sections varying over many orders of magnitude.

The aim of this project is to improve our understanding of such processes and set exclusion limits on the scale Λ_{CI}. The next section (2) contains a brief introduction to the relevant theoretical background. Section 3 describes the methodology of the study, with the results presented in section 4. Avenues for further investigation are suggested in section 5, and finally section 6 presents a summary of the project and concluding remarks.

2 Theoretical background

We wish to study the general flavor-symmetric $q\bar{q} \to q\bar{q}$ interaction with leading order (LO) Feynman diagram shown in Figure 2. The LO diagram contains two vertices. The next–to–leading order (NLO) Feynman diagrams have three vertices, NNLO diagrams have four, and so on. There are infinitely many possible diagrams which all sum together to give the total QCD cross–section. The size of each term is proportional to α_S^n, where n is the number of vertices and α_S is the strong coupling constant. Calculating the contribution of each diagram quickly becomes extremely difficult as the number of vertices (and loops) increases.

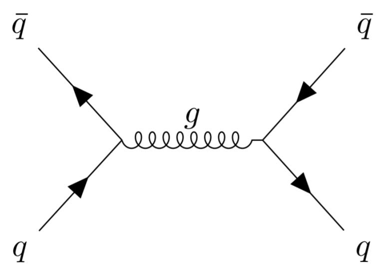

Figure 2: **Quark interaction leading order Feynman diagram.** This diagram represents a calculation, with a value proportional to α_S^2, where α_S is the strong coupling constant.

One of the most important features of QCD is asymptotic freedom: the strong coupling constant α_S runs to smaller values as the centre of mass energy of the system increases, or, equivalently, as the length scale of the interaction decreases. This is in contrast to quantum electrodynamics (QED), where the coupling constant increases with energy. An important consequence of asymptotic freedom in QCD is that at high energies α_S becomes much smaller than unity, hence the higher–order Feynman diagrams become negligible and the total cross–section can be calculated to high accuracy as a truncated perturbation series [6]. In practice, it is often sufficient to just calculate LO or LO and NLO terms for fruitful analysis.

Jets produced in high-energy pp collisions at the LHC are sufficiently energetic to be treated in the perturbative regime. However, since we are colliding protons and not individual quarks, the calculation of the cross–section must take into account the momentum distributions of the quarks within a proton. These are called parton density functions (PDFs) and can be obtained by experiments such as deep inelastic scattering. Many PDFs and their uncertainties have been tabulated and are readily available online [2]. The factorisation theorem in QCD then allows the total cross–section to be calculated by taking the convolution of the truncated perturbation series with a PDF.

The four quark contact interaction (CI) is a beyond–Standard Model effective field theory for the flavor-symmetric $q\bar{q} \to q\bar{q}$ interaction. The calculation of the CI contribution to an inclusive jet cross–section can be constructed in a similar way as for QCD: by convolving a perturbation series with a PDF. The CI contribution is negligible compared to QCD at low p_T but becomes significant at p_T greater than 1 TeV. Its strength is characterised by the scale Λ_{CI}; a smaller scale results in a larger deviation from QCD. These features are illustrated in Figure 3. CI models can be calculated for different quark pair chirality and color states.

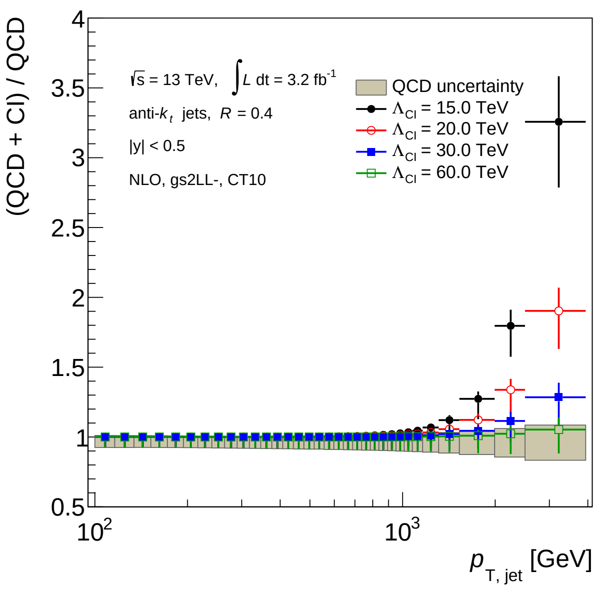

Figure 3: **Ratio of QCD + CI to QCD inclusive jet cross–section calculations**. The ratio approaches unity at low p_T, indicating that the CI contribution is negligible compared to the QCD contribution in this region of phase space. However, for $p_T > 1$ TeV, the CI contribution becomes significant. As Λ_{CI} becomes larger, the deviation of CI from QCD becomes smaller.

3 Methods

3.1 Software for calculations

We used CIJET1.1 [5] to calculate CI contributions to inclusive jet cross–sections for different color and chirality states. The program was first used to carry out CPU–intensive grid calculations for all the p_T and y bins for which we had data. For the highest p_T bins, the calculation with 3 million Monte Carlo points took up to one week on Condor. CIJET can then be used to convolve the grids with any PDF and compute the LO and NLO CI cross–section contributions for different models and scales Λ_{CI}. These calculations use a rapid grid interpolation algorithm and are completed in a few seconds. The QCD inclusive jet cross–sections were calculated using APPLGRID [3], which operates in a similar way.

A software framework comprising over 2000 lines of Python code was developed during this internship. The code defines TheoryQCD, TheoryCI and Data classes which, upon object initialisation, search for files containing theory calculations or data tables and store the nominal cross–sections in numpy arrays and the systematic and statistical uncertainty components in Python dicts containing numpy arrays. If the files for the desired theory calculations are not found, then convolution scripts from APPLGRID or CIJET are called to produce the files, and they are moved to the correct directory for subsequent access. These classes are used by a set of analysis and plotting modules (with pyROOT) for quick and convenient investigation of the results.

3.2 Treatment of uncertainties

The data carries over 100 systematic uncertainty components, which are correlated across all the p_T and y bins but are independent of each other. There are also some uncorrelated statistical uncertainties. Correlated statistical uncertainties can also be included, but such correlations are typically small so have been neglected in our analysis.

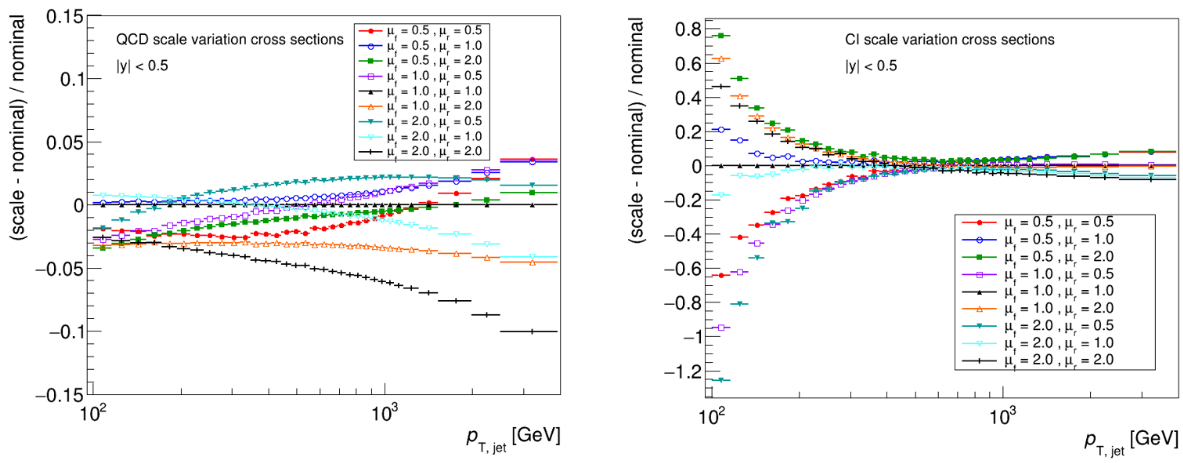

Figure 4: **QCD and CI scale variations.** These correlated systematic uncertainties lead to asymmetric uncertainty bands in both QCD and CI predictions.

The QCD and CI have systematic uncertainties that arise from the missing higher–order terms. These are evaluated by varying the renormalisation and factorisation scales, μ_r and

μ_f. The nominal cross–sections are calculated using $\mu_r = \mu_f = 1$, while the "scale variation" systematic uncertainties are calculated by computing cross–sections with the scales varied by factors of two and subtracting the result from the nominal cross–section. These are plotted in Figure 4. In addition to scale variations, the QCD also has systematic uncertainties that come from the uncertainty in the value of α_S, since this value affects the perturbation series used in the cross–section calculation (see section 2). The uncertainties due to α_S were investigated and found to be small compared to the scale variations, so are neglected in subsequent analysis. There are also uncertainties in the PDFs used to generate the cross–sections; these were also neglected, but should be included in subsequent studies.

3.3 Methods for setting limits on $\Lambda_{\mathbf{CI}}$

Our first approach to setting limits on Λ_{CI} involved performing pull calculations. The theory–data pull is defined as (Theory $-$ Data)/σ, where σ is the combined uncertainty calculated, depending on the sign of the pull, by adding in quadrature either all the upper or all the lower uncertainty components from the theory and data. Assuming the pull follows a Gaussian distribution with zero mean and unit standard deviation, a pull greater than 1.96 corresponds to a p–value less than 0.05 and hence leads to exclusion at the 5% significance level. We calculated the pull of each individual p_T bin for CI hypotheses over a fine Λ_{CI} scan and rejected all hypotheses where the pull exceeded 1.96 in any of the bins with $p_T > 500$ GeV. This calculation was repeated over the different rapidity bins and chirality and color models.

One major limitation of this method is that it is not well adapted for bins where the QCD prediction itself significantly overestimates the data, since then the QCD-only pull already approaches (or in some cases even exceeds) the 1.96 limit. Such bins must be excluded to set sensible limits. Another issue is that the systematic uncertainty components are added in quadrature and thus are treated as if they were uncorrelated, but in reality they are correlated across all the p_T and y bins.

To overcome these issues, the convention in high energy physics is to use the CL_s method [7] to set limits on parameters that are strictly non-negative. The χ^2-statistic for the data is calculated both under the assumption that only background processes are occurring (the null hypothesis), and under the assumption that both signal and background processes are present (the alternative hypothesis). The corresponding p–values are then computed by integrating the χ^2 distribution between the obtained χ^2 value and infinity for both the null and alternative hypotheses. These p–values are called CL_b for the background only calculation and CL_{s+b} for the computation with both signal and background. Finally, the CL_s is computed by taking the ratio of these two quantities, $CL_s = CL_{s+b}/CL_b$.

The advantage of this method over computing a single χ^2 statistic for combined signal and background is that it realistically accounts for the sensitivity to the signal even when the data deviates from the background alone. This is why this method was used in the search and eventual discovery of the Higgs boson at LEP and the LHC [4].

We calculated the CL_{s+b} and CL_b by calculating the χ^2 for each bin and using the `chi2.sf()` function from the `scipy.stats.distributions` module to convert the χ^2 to a p–value. Then CL_s is simply the ratio of these two quantities.

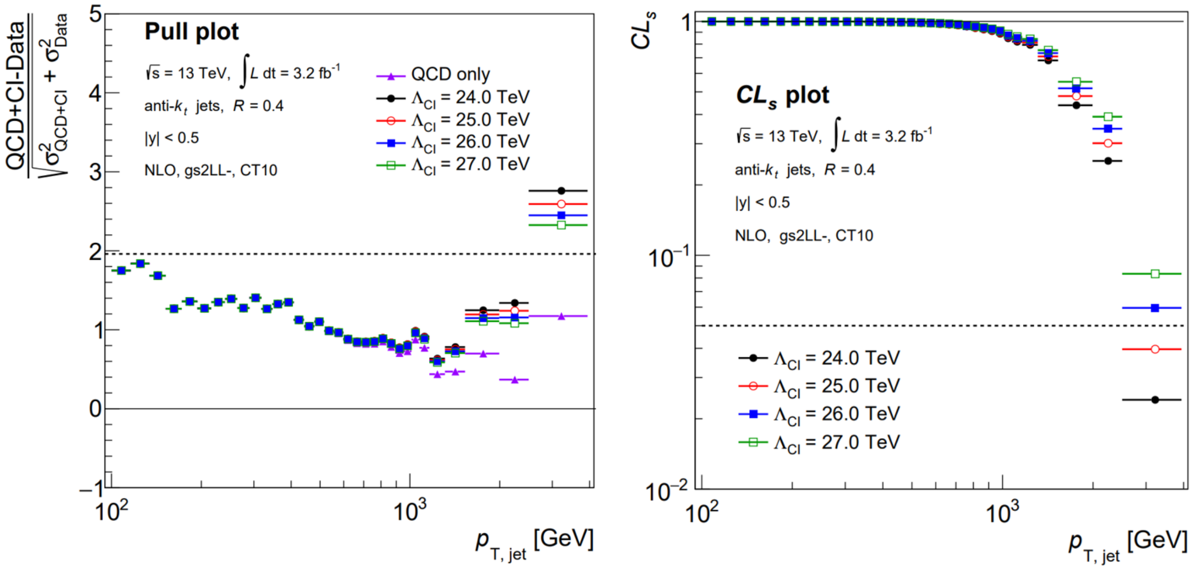

Figure 5: **A comparison of pull (left) and CL_s (right) for each p_T bin of the LL singlet model in the $|y| < 0.5$ rapidity bin.** In the pull plot, the critical region at 5% significance is above the dashed line (pull > 1.96) while for the CL_s plot it is below the dashed line ($CL_s < 0.05$). The QCD is observed to fluctuate quite significantly from the data. This causes the pull to exclude all the models shown, leading to an overestimation of the limit. By contrast, the CL_s accounts for the QCD fluctuations and, from the models shown, only excludes $\Lambda_{\text{CI}} \leq 25$ TeV.

4 Results

4.1 Limits

We computed exclusion limits on Λ_{CI} using a 2015 data set at $\sqrt{s} = 13$ TeV with integrated luminosity of 3.2 fb^{-1} from the ATLAS experiment [1].

A side-by-side comparison between pull and CL_s is shown in Figure 5. Since our QCD predictions systematically overestimated the data, any exclusion limits set by the pull method would be artificially high, and so those results are not shown in this document. A better method is to calculate the CL_s for each individual p_T bin and scan over Λ_{CI}, rejecting any values where the CL_s drops below 0.05 (to test at the 5% significance level). The results from a scan in steps of 0.5 TeV of this analysis are shown in Table 1.

The above analysis does not account for the correlations of the systematic uncertainties between the bins. To remedy this, we used a Python package called `pyhf`, which can calculate the observed CL_s and a band of expected CL_s by using asymptotic formulas or by throwing toys. For our preliminary analysis, we set up `pyhf` to run using asymptotic formulas and processed a fine scan over Λ_{CI} in steps of 0.1 TeV. Large fluctuations were observed in the results when all the p_T bins were included, so the computation was repeated with just the three highest p_T bins where sensitivity is largest. The results were improved, as shown in Figure 6, and an exclusion limit of $\Lambda_{\text{CI}} \leq 26.9$ TeV was found for the identical chirality (LL or RR) color singlet models.

$\boldsymbol{ATLAS}$ 2015 data $\sqrt{s} = 13$ TeV 3.2 fb^{-1} CT10						
Rapidity	**Exclusion Limits for Λ_{CI} (TeV)**					
	LL singlet	LR singlet	RR singlet	LL octet	LR octet	RR octet
$\|y\| < 0.5$	25.5	12.5	25.5	18.0	12.0	18.0
$0.5 \leq \|y\| < 1.0$	17.5	9.5	17.5	12.0	9.0	12.0
$1.0 \leq \|y\| < 1.5$	15.0	8.5	15.0	11.0	9.0	11.0
$1.5 \leq \|y\| < 2.0$	11.5	6.5	11.5	8.0	7.5	8.0
$2.0 \leq \|y\| < 2.5$	7.0	4.5	7.0	5.0	5.0	5.0

Table 1: **Preliminary results for exclusion limits on Λ_{CI} using single-bin CL_s at the 5% significance level.** The limits are obtained by calculating the CL_s for each p_T bin separately across a scan of different Λ_{CI} in intervals of 0.5 TeV and excluding any models where the $CL_s < 0.05$ for at least one p_T bin. The highest limit is an exclusion of $\Lambda_{\text{CI}} \leq 25.5$ TeV for LL and RR color singlet states in the $|y| < 0.5$ rapidity bin. Mixing chirality or color states reduces the limit. Our sensitivity to contact interaction decreases with increasing rapidity.

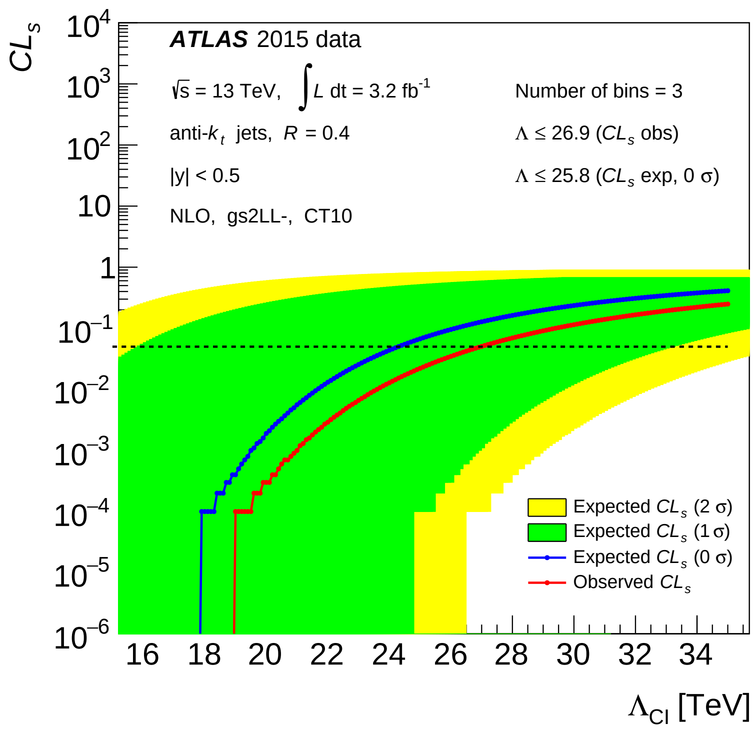

Figure 6: **CL_s exclusion bands.** Using `pyhf`, the expected CL_s bands at 1 and 2 σ, and the observed CL_s were computed over a fine Λ_{CI} scan from 15.0 to 35.0 TeV in steps of 0.1 TeV, for the LL color singlet model in the $|y| < 0.5$ rapidity bin. Only the three highest p_T bins were included in this calculation, where CI sensitivity is greatest. The exclusion limits at 95% confidence level are $\Lambda_{\text{CI}} \leq 26.9$ TeV from the observed CL_s, or $\Lambda_{\text{CI}} \leq 25.8$ TeV from the expected CL_s.

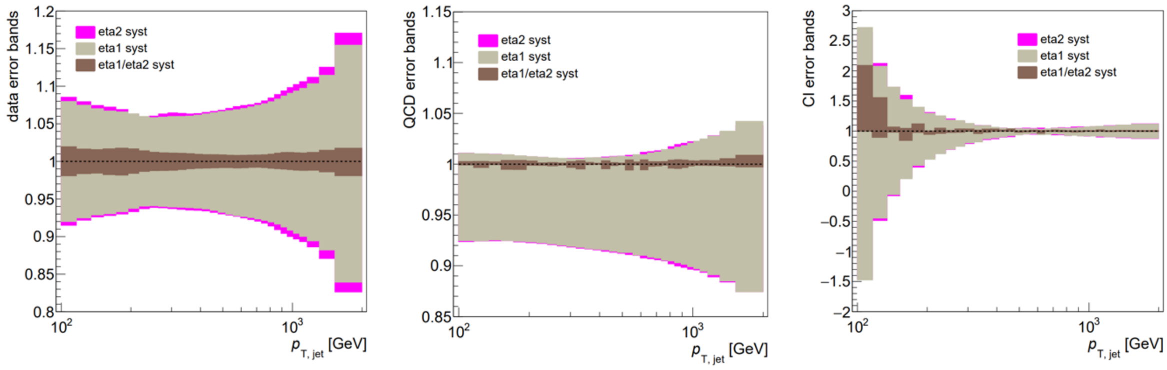

Figure 7: **Data, QCD and CI systematic uncertainty bands for two rapidity bins and their ratio.** The ratio is observed to have significantly smaller total systematic uncertainties, as the correlated upwards or downwards fluctuations due to the individual systematic uncertainty components "cancel" out. This is particularly prominent in the QCD calculation (central plot). The rapidity bin labelled eta1 corresponds to $|y| < 0.5$, and eta2 corresponds to $0.5 \le |y| < 1.0$.

4.2 Ratios

We also investigated the effect of taking ratios between different rapidity bins on the total systematic uncertainty of the data and theory. For each uncertainty component, the 'varied' cross–section is calculated by adding the uncertainty to the nominal cross–section. Then the ratio was computed between the two varied cross–sections corresponding to the same uncertainty component at two different rapidity bins. Since the systematic uncertainty components are correlated across rapidity bins, an upwards fluctuation in one bin is likely to correspond to a similar upwards fluctuation in the other bin, so the ratios of varied cross–sections approach the ratios of nominal cross–sections.

The new systematic uncertainty components are then calculated by subtracting the ratio of nominal cross–sections from the ratio of varied cross–sections. The resulting total systematic uncertainties are significantly reduced compared to the original uncertainties, as demonstrated in Figure 7 for a ratio between the two lowest rapidity bins. The upper and lower bands are computed separately, by adding all the upwards or all the downwards fluctuating uncertainty components, respectively, in quadrature.

5 Future studies

Further investigation should be carried out to examine the effect of increasing number of bins on the resulting CL_s computed by pyhf in order to ensure that bin choice does not significantly affect the results. The PDF uncertainties should also be included in the theory calculations.

Sensitivity could be increased by taking ratios of the cross–sections across different rapidity bins or centre of mass energies to reduce the systematic uncertainties, as discussed in subsection 4.2. This is expected to result in stronger limits.

Analysis of data from the LHC full Run 2 or the high luminosity LHC in the future could lead to increased sensitivity to CIs by measuring jets at higher p_T. Speculatively, the Future Circular Collider, with its unparalleled centre of mass energy $\sqrt{s} = 100$ TeV, might allow researchers to uncover any processes that underlie contact interactions, if they exist.

6 Conclusions

We investigated the exclusion limits on the CI scale Λ_{CI} by studying inclusive jet cross–sections.

To achieve this, inclusive jet cross–sections were calculated to NLO using APPLGRID for the QCD contribution and CIJET for the CI contribution. Statistical methods were used to compare these results with data from the ATLAS experiment and set exclusion limits. A Python framework was developed during the internship to quickly and conveniently call the relevant computations or read tables, and to carry out analysis and make plots.

Our preliminary results were obtained by calculating the theory–data pull, ignoring uncertainty correlations, and performing hypothesis tests at the 5% significance level. These led to artificially high limits on account of the large fluctuations of the QCD compared to the data. Calculating the CL_s for each bin led to much more sensible limits, shown in Table 1. Stronger limits were obtained for states where both quarks have the same chirality, and for color singlet states. Furthemore, it was found that our sensitivity to contact interactions decreases as rapidity increases. This was expected because the CI/QCD cross–section ratio is largest at high p_T, and higher p_T scales are probed in the central region of the detector.

Using the `pyhf` package to deal with correlated systematic uncertainties through the use of asymptotic formulas, our preliminary analysis for the model involving equal chirality (LL or RR) color singlet $q\bar{q}$ pairs excludes contact interactions with $\Lambda_{\text{CI}} \leq 26.9$ TeV. This is comparable to the $\Lambda_{\text{CI}} \leq 22.5$ TeV exclusion limit obtained for the same model from the earlier single-bin CL_s analysis.

Acknowledgements

I would like to thank Tancredi Carli for supervising this project over 11 weeks, providing excellent guidance and direction to the work. I am immensely grateful for our regular Zoom meetings where we had interesting discussions that helped me understand the aims of the project and many aspects of high energy physics analysis. I would also like to thank Bogdan Malaescu for helping me understand the relevant statistical methods and for providing advice on writing this report, and Mark Sutton for helping me install the APPLGRID software.

References

[1] M. Aaboud et al. Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 05:195, 2018.

[2] A. Buckley et al. LHAPDF6: parton density access in the LHC precision era. *Eur.Phys.J. C75*, 3:132, 2015.

[3] T. Carli, D. Clements, A. Cooper-Sarkar, et al. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project. *Eur. Phys. J. C*, 66:503–524, 2010.

[4] Atlas Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[5] Jun Gao. CIJET: A program for computation of jet cross sections induced by quark contact interactions at hadron colliders. *Computer Physics Communications*, 184(10):2362–2366, 2013.

[6] David J Griffiths. *Introduction to elementary particles; 2nd rev. version.* Physics textbook. Wiley, New York, NY, 2008.

[7] A L Read. Presentation of search results: the CLs technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, sep 2002.