

A more precise measurement of D meson mass difference through $D_{(s)}^+ \rightarrow K^+K^-\pi^+$ decay

John Bodenschatz

Physics Capstone

University of Cincinnati

April 26, 2021

Abstract

Using the three body decay of $D_{(s)}^+ \rightarrow K^+K^-\pi^+$, we improve the statistical error of the mass difference measurement $m_{D_s^+} - m_{D^+}$. The most recent high precision measurement (2013) is $98.68 \pm 0.03(\text{stat}) \pm 0.04(\text{syst})$ MeV/ c^2 [2]. From 2017 LHCb data, we yield $4,970,691 \pm 11,548$ D^+ signal events and $8,524,400 \pm 17,505$ D_s^+ signal events. We measure $98.900 \pm 0.0036(\text{stat}) \pm x.xxx(\text{syst})$ MeV/ c^2 . Through future work we will be able to establish a systematic uncertainty on the same order as statistical.



Social Responsibility Statement

Throughout the past year I have had the pleasure of studying high energy physics and learning all about CERN, LHCb, and more thanks to Dr. Michael Sokoloff. I have been able to work on CERN's global computing system to acquire vast amounts of data to make a real physics measurement. I have also learned a great deal along the way about the research process and how I work best when faced with problems. As for the project itself, these sort of studies are conducted fairly frequently and for a good reason. Making more and more high precision measurements is a pinnacle of experimental physics. Every significant digit matters and can have an impact on how we understand the results, especially on the scale of mesons and other such particles. I have vastly developed my skills as an individual and collaborative researcher during this time. At the beginning of my time with Dr. Sokoloff, there were three things he said he wanted us to learn: how to communicate effectively, how to work with others, and how to think critically about certain problems. I have worked a great deal at developing my ability to accomplish these three tasks, for both research and life. The opportunities provided to me through this research will forever have an impact on my future endeavors. Following graduation, I will be attending Marquette University where I will pursue a Ph.D. in Computational Mathematical and Statistical Sciences. I am hoping to develop a deeper understanding of the application of computational techniques and how they can solve real-world problems. My education in physics has outfitted me with an incredible set of analytical tools that will certainly go far in life.

1 Introduction

Mesons are a family of colorless subatomic particles composed of quark-antiquark pairs. They are particularly sensitive to the strong force, the force responsible for holding together atomic nuclei and governing interactions between particles containing quarks. By studying the mass difference of mesons, we can determine constants such as m_c/m_b which are fundamental parameters of Quantum Chromodynamics (QCD) [1]. QCD, analogous to Quantum Electrodynamics (QED), involves the theory describing the action of the strong force [10]. For this study in particular, the mesons are composed of the following constituent quarks:

$$\begin{aligned} |D^+\rangle &= |c\bar{d}\rangle \\ |D_s^+\rangle &= |c\bar{s}\rangle \end{aligned}$$

These mesons will be used to measure the mass difference $m_{D_s^+} - m_{D^+}$. The most recent high precision measurement of this mass difference was done on Run 1 data from the LHCb experiment [2]. This study yielded $68,787 \pm 321$ D^+ events and $284,694 \pm 540$ D_s^+ events and resulted in a measurement of $m_{D_s^+} - m_{D^+} = 98.68 \pm 0.03 \pm 0.04$ MeV. The Particle Data Group (PDG) average for this measurement is currently listed as 98.69 ± 0.05 MeV [4].

By an early estimate of the amount of data in Run 2 compared to Run 1 at the LHC, we estimate that it may be possible to reduce the statistical error of this measurement by a factor of 10. In 2017 alone, 40 PB of data produced from LHC experiments were recorded at the CERN data center [5].

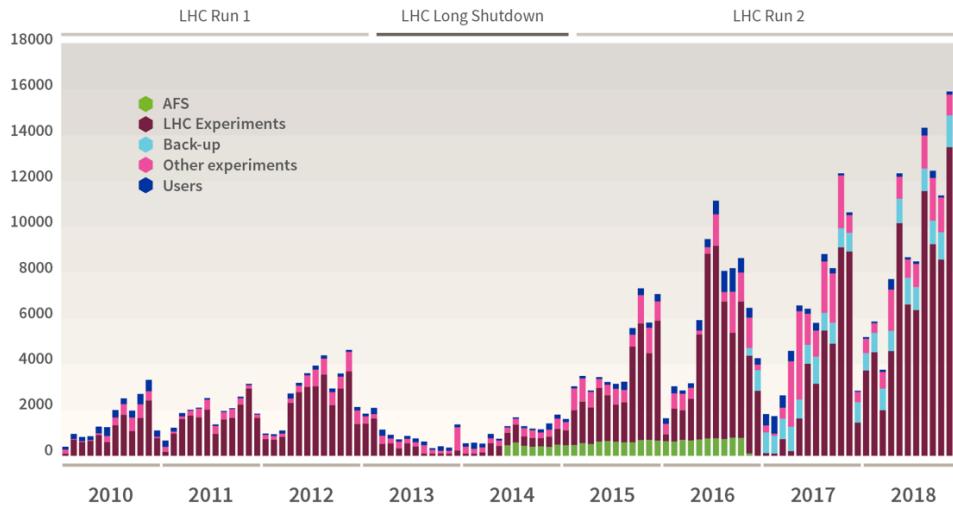


Figure 1: Image from CERN Computing showing the data (TB) recorded at CERN month-by-month.

2 Detector and Dataset

The LHCb detector is part of one of four main experiments at CERN’s Large Hadron Collider (LHC) in Geneva, Switzerland. This experiment’s focus is to study the slight differences between matter and anti-matter by studying particles containing b quarks. As depicted in Figure 2, the 5600-ton detector is composed of many sub-detectors that span 10 meters tall, 13 meters wide, and 20 meters along the beam pipe [7]. Unlike some other detectors at the LHC, the LHCb experiment is built to measure forward particles- or particles that are projected in one direction roughly along the beam pipe.

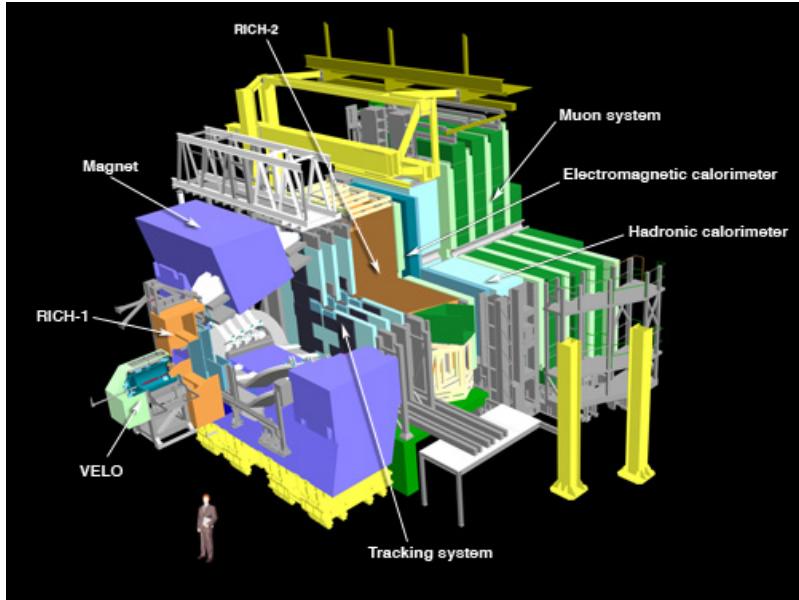


Figure 2: Diagram of the LHCb detector pointing out several useful sub-detectors.

A system of ring-imaging Cherenkov detectors (RICH1 and RICH2) are used to identify charged hadrons that pass through the detector. The two mesons that compose the daughters of the studied decay, the kaon and the pion, are both hadronic. The magnet on the detector is used to bend charged particles slightly away from the beam and into other parts of the sub-detectors, namely the tracking system. The tracking system is composed of three stations of silicon-strip detectors and straw drift tubes (sometimes referred to as T-Stations). The T-Stations are used to measure the momentum of particles passing through.

When operational, the LHCb records approximately 10 million collisions per second [8]. After filtering, only about 1 million events are saved to storage; this translates to 35 GB of data per second. The detector observes and records decay channels from many particles. For this analysis we are only interested in the $D_{(s)}^+ \rightarrow K^+ K^- \pi^+$ decay mode. Data for this analysis was taken from all 2017 (Run 2) proton-proton collisions, totaling to 1.3 TB of data before cuts. The flow of data as taken from the detector to analysis can be seen in Figure 3, outlined in red.

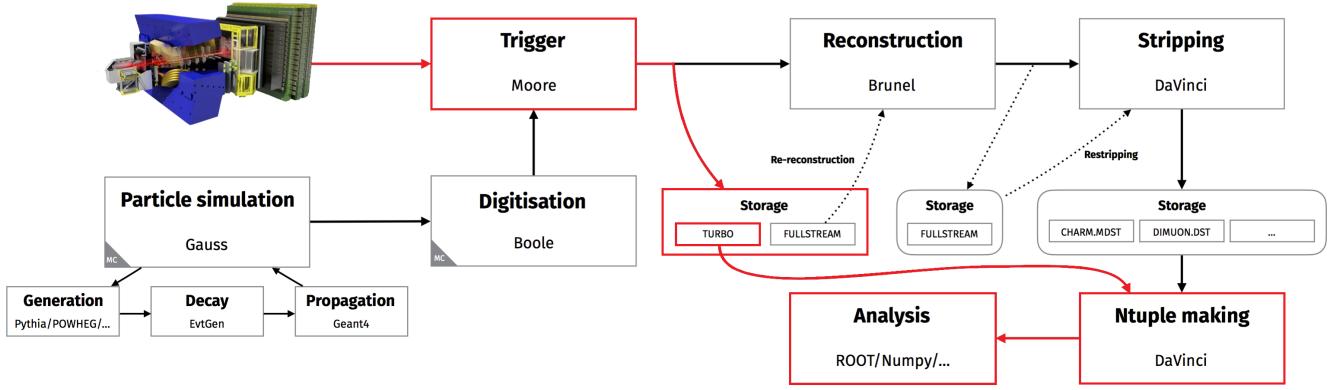
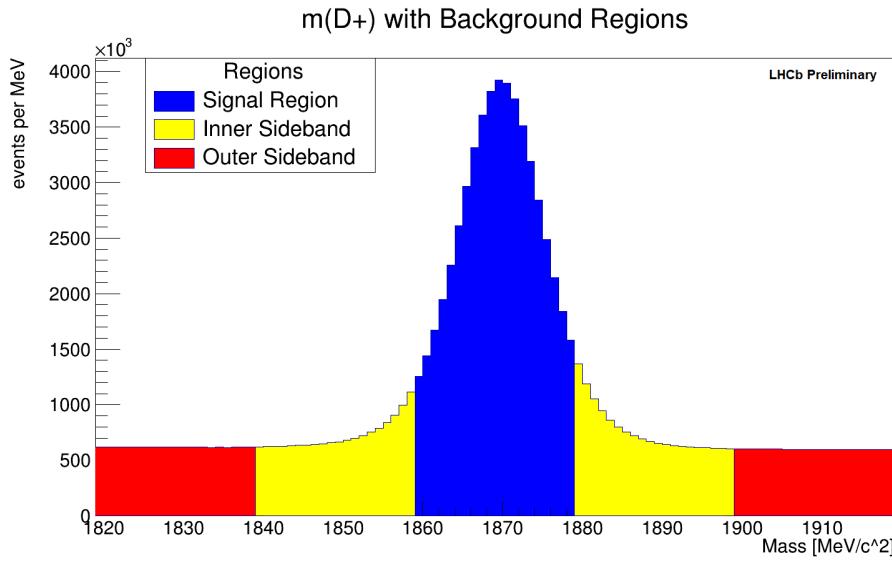


Figure 3: Red path highlighting the flow of data used for this study.

3 Selection

Not all raw data that is collected is used for the analysis. Each candidate event must pass a set of selection criteria, often referred to as "cuts", to be accepted for fitting. The goal in cutting on the data is to reduce the background while retaining a large amount of signal events, i.e., increasing the signal-to-background ratio. However, cuts must also remain physically valid and justifiable. We cannot cut on a variable simply because it gives us a better result- we must be able to say what purpose the cut serves with regards to eliminating "bad" events.

The raw data provided us with $3.5 \cdot 10^8 D^+$ candidate events and $5.0 \cdot 10^8 D_s^+$ candidate events. Figure 4 shows a sample of the raw D^+ data before any cuts are applied. Here, the signal-to-background ratio is about 8.

Figure 4: Sample of the D^+ data before cuts are applied.

To select well-identified kaons (pions) the difference in the logarithms of the global likelihood of the kaon (pion) hypothesis relative to the pion (kaon) hypothesis provided by the ring-imaging Cherenkov detectors is required to be greater than five (zero) [2]. In addition to this, the product of the neural net likelihoods of the kaon hypothesis for each kaon and the pion hypothesis for the pion is constrained to be greater than 0.7. These cuts, in effect, ensure that the kaons and pions as identified by the detector truly *are* kaons and pions, rather than some other misidentified charged hadron.

To eliminate kinematic reflections due to misidentified pions, the invariant mass of at least one kaon pair is required to be within $10 \text{ MeV}/c^2$ of the nominal value of the ϕ meson mass [2, 9]. This means that the D meson decay for our study will be dominated by those including an intermediate ϕ meson. This cut can be seen illustrated in Figure 5, a Dalitz plot showing the phase-space of the D^+ mother particle. The vertical band in the right figure is the so-called ϕ region. By selecting only these decays, we are able to reduce a significant amount of background from the mass-fitting; almost all of the events in the background of the Dalitz plot are cut.

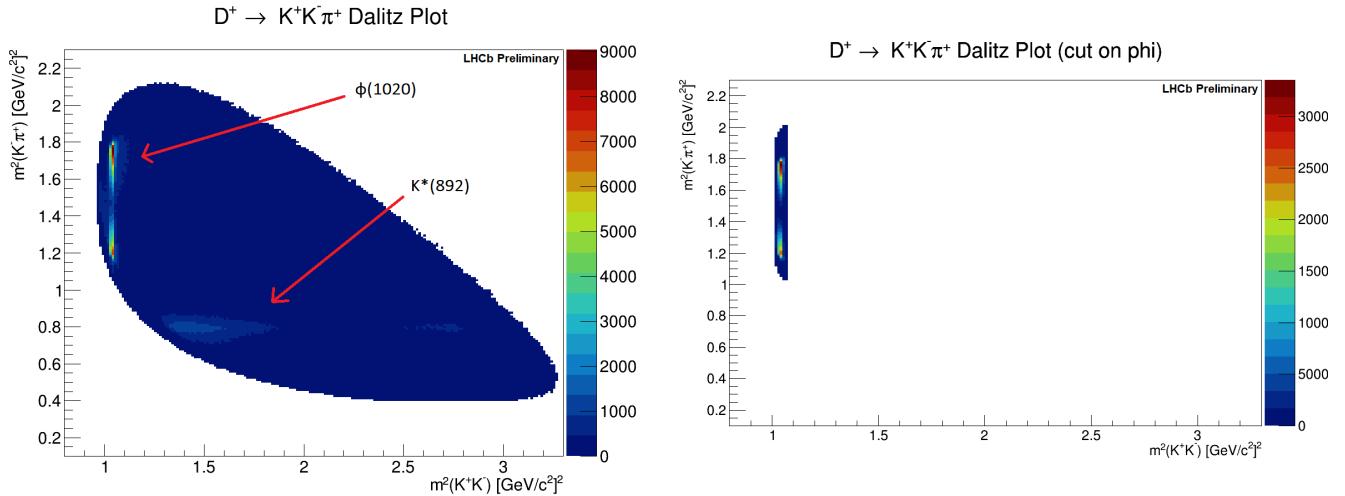


Figure 5: Dalitz plot of $D^+ \rightarrow K^+ K^- \pi^+$ decay before and after the ϕ cut.

There are a handful of cuts on the reconstructed D_s^+ mother particle. The impact parameter χ^2 , which is an indicator of the particle's distance of closest approach to the pp interaction vertex is restricted to be less than 4. The flight distance χ^2 is restricted to be greater than 225. This is to ensure a good fit for flight distance to the primary vertex of the decay. The end vertex χ^2 , a measure of the agreement between there constructed tracks of final decay particles at their decay vertex, is restricted to be less than 9. Finally, the cosine of the direction angle (DIRA) is restricted to be greater than 0.99999, to ensure that the particle is following closely to the beam.

There is one final part of the selection process that is unique to this analysis and comes from the fact that the D^+ data and D_s^+ data are stored separately. Ideally, we will have one spectrum of data spanning both the central value of the D^+ mass peak and the D_s^+ mass peak such that we can have one fit for both peaks simultaneously. However, the raw data for D^+ decays spans a reconstructed mass range of 1790 – 1950 MeV, while the raw data for D_s^+ decays spans a range of 1890 – 2050 MeV. This leaves us with an "overlap" region in the range of 1890 – 1950 MeV.

For reasons that are not entirely clear, the counts in these bins are *not* equal as one would expect. Figure 6 illuminates the issue we face. After cuts the central overlap region bin differences are on the order of 10^0 whereas the bins themselves are on the order of 10^3 . Since this is relatively small, we choose the center, 1920 MeV, as the point at which we stop using D^+ data and start using D_s^+ data to construct a complete spectrum.

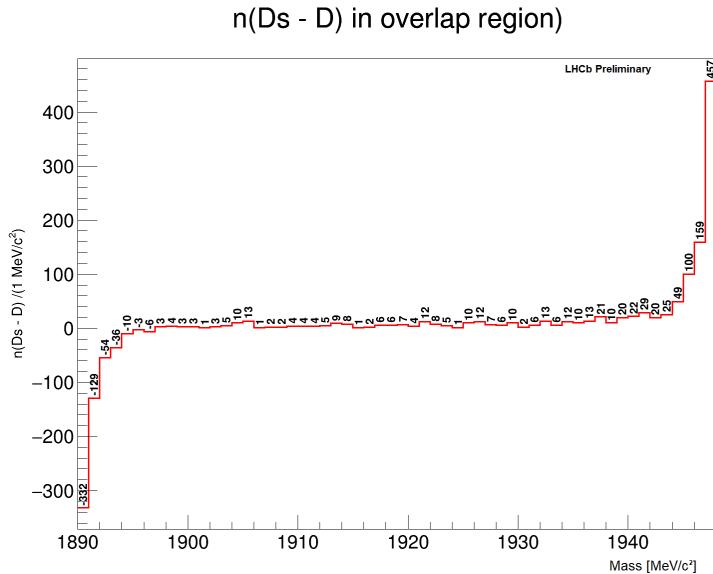


Figure 6: Difference of events in each data set in the overlap region.

4 Fit Results

After all cuts are made and extraneous variables are removed from the data set, we are left with 320 MB of data. This is approximately 0.025% of the original, raw data. The fit is built from six components. The D^+ signal is modeled by a Gaussian and a Crystal Ball function with common mean. The D_s^+ signal is modeled by a double Gaussian and a Crystal Ball function, all with common mean. The second Gaussian in the D_s^+ fit allows for some added flexibility. Crystal Ball functions are composed of a Gaussian core with a power law tail on the lower end. Though they take various forms, an example can be seen in Equation 1 where N is a normalization constant. The two Crystal Ball functions for this fit are fixed to have the same parameters, α and n .

$$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) & \frac{x-\bar{x}}{\sigma} > -\alpha \\ \left(\frac{N}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \cdot \left(\frac{N}{|\alpha|} - |\alpha| - \frac{x-\bar{x}}{\sigma}\right)^{-n} & \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \quad (1)$$

The "pull plot" is a useful tool when measuring how good a fit is. It is defined in the following way:

$$\text{pull} = \frac{\text{binVal} - \text{fitVal}}{\sqrt{\text{fitVal}}} \quad (2)$$

where $binVal$ is the real value of a bin and $fitVal$ is the fitted value. When looking at a pull plot, we generally aim for all pulls to be within 5σ of the origin. We also hope to see the pulls relatively randomly distributed, indicating that there are no underlying patterns that may have been missed or that the minimization of the fit was poor).

The final fit can be seen in Figure 7. Overall, the fit is pretty good. The pull values are slightly higher than desired in the overlap region, but this is okay considering the bins in this region are on the order of 10^3 events, whereas bins in the signal regions are on the order of 10^5 events. There is some periodic motion of the pull throughout the D_s^+ signal peak which is likely a result of poor fitting parameters or minimization. The second Gaussian in this peak was implemented to help resolve these issues, and it does a fair amount, but the fit could certainly be better in this area. The uncertainties on all parameters in the fit are small compared to the parameters themselves, which is another indicator of a good fit. The signal-to-background ration on the fit is about 400.

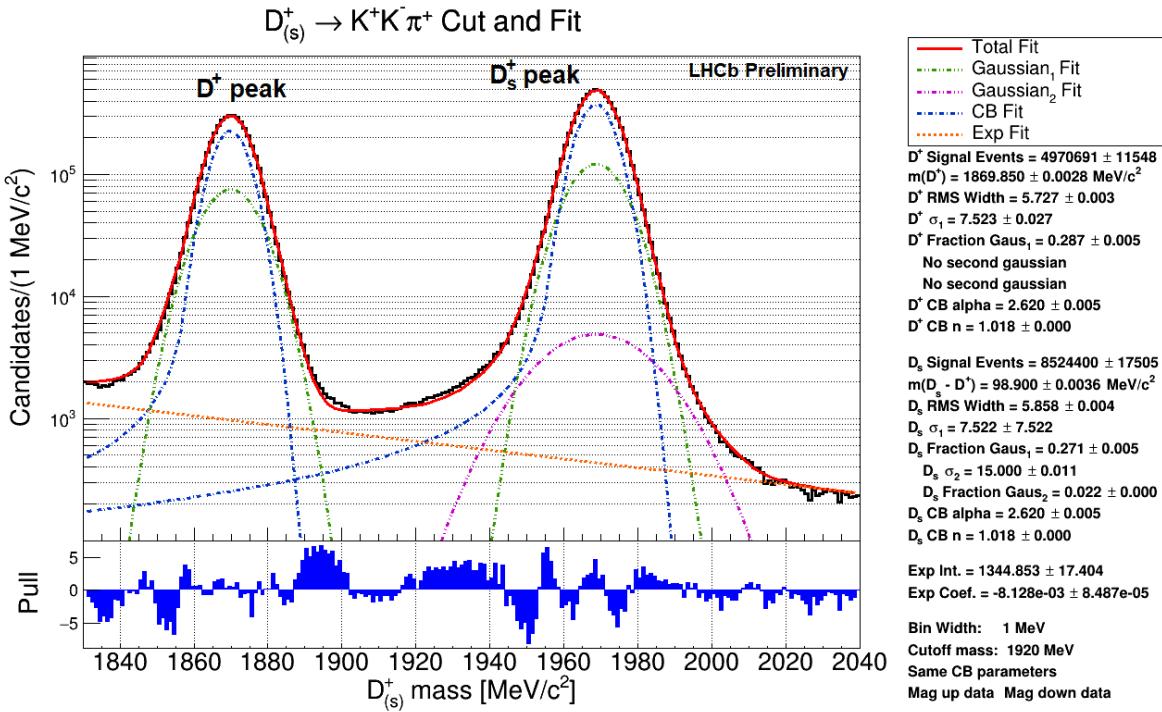


Figure 7: D^+ and D_s^+ signal and background fit after all applied cuts.

The fit is repeated twice: once where the mass of the D^+ and D_s^+ are fit, and a second time where the mass of the D^+ is fit and the mass difference $m(D_s^+) - m(D^+)$ is fit. We fit $4,970,691 \pm 11,548$ D^+ signal events and $8,524,400 \pm 17,505$ D_s^+ signal events. We arrive at the following mass and mass difference measurements:

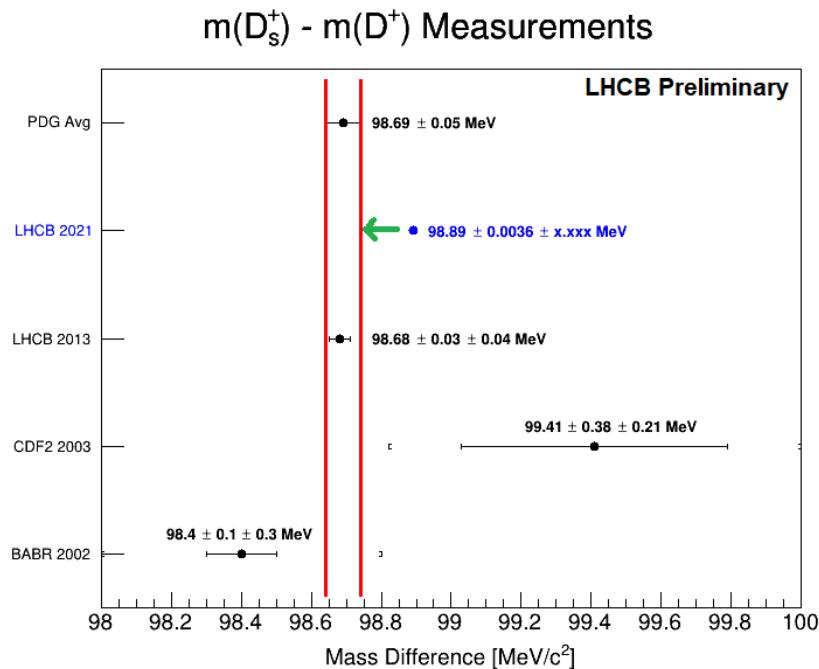
$$\begin{aligned} m(D^+) &= 1869.850 \pm 0.0029(\text{stat}) \pm x.xxx(\text{syst}) \text{ MeV}/c^2 \\ m(D_s^+) &= 1968.750 \pm 0.0023(\text{stat}) \pm x.xxx(\text{syst}) \text{ MeV}/c^2 \\ m(D_s^+) - m(D^+) &= 98.900 \pm 0.0036(\text{stat}) \pm x.xxx(\text{syst}) \text{ MeV}/c^2 \end{aligned}$$

Organization	Year	$m_{D_s^+} - m_{D^+}$ [MeV]	D^+ Signal Events	D_s^+ Signal Events
BaBar[3]	2002	98.4 ± 0.3 (stat)	—	$47,794 \pm 311$
LHCb[2]	2013	98.68 ± 0.03 (stat)	$68,787 \pm 321$	$248,694 \pm 540$
LHCb - UC	2020	98.97 ± 0.1 (stat)	$73,644 \pm 285$	$12,529 \pm 1602$
LHCb - UC	2021	98.89 ± 0.0036 (stat)	$4,970,691 \pm 11,548$	$8,524,400 \pm 17,505$

Table 1: Table of previous $m_{D_s^+} - m_{D^+}$ measurements.

5 Summary

As stated before, this is not the first time this measurement has been made via pp collision data- it is simply an attempt to reach a higher precision than it's predecessors. Table 1 and Figure 8 lay out the development of this measurement over the past 20 years. Table 1 also includes the measurement made at the end of my summer 2020 research term. The improvement in results from then to now is realized by reducing the statistical error by about a factor of 30. From the LHCb measurement in 2013, we have reduced the statistical error by a factor of about 8. As it stands, this measurement is 4σ away from the PDG average. This is, for now, an admissible deviation. There is still some work do be done on the measurement, especially with regards to the systematics. We believe that such considerations will reduce our measurement, as indicated by the green arrow in Figure 8.

Figure 8: Comparison of $m_{D_s^+} - m_{D^+}$ measurements.

5.1 Future Work

As has been alluded to, this measurement has some work to be done before it can be complete. The main focus moving forward will be to establish the systematics of the decay and how to appropriately account for them in the analysis. We believe that the most dominating systematic uncertainty will come from the calibration of the momentum scale of the detector. This issue arises from uncertainties involving the strength of the magnetic field produced by the detector. There are also other systematics to consider, but as for now it is estimated that they will pale in comparison to the momentum scale uncertainty. We estimate the systematic error to be on the same order as the statistical error, or about 3 keV.

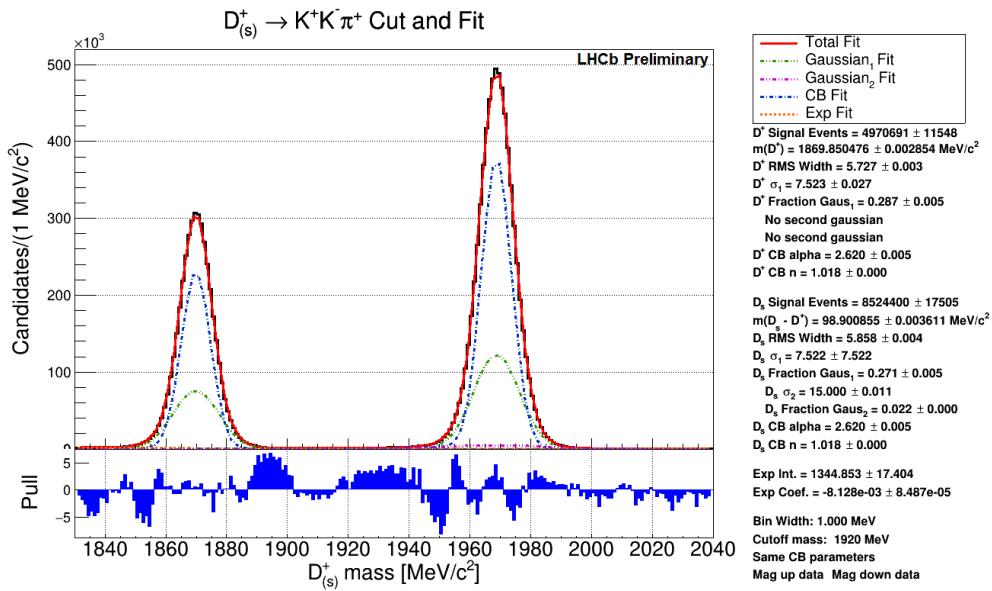


Figure 9: D^+ and D_s^+ signal and background fit after all applied cuts (linear y-axis).

Acknowledgments

I would like to thank Dr. Michael Sokoloff for helping me through this research. Over the past year I have learned so much, not only about high energy physics, but about the research process as a whole. His guidance and wisdom have proved invaluable to my advancement as a researcher. I would also like to thank the other students researching under Dr. Sokoloff. To those who had already been a part of the group when I joined and were able to answer so many questions and to those who joined along with myself and helped with the learning process.

This research was made possible thanks to grants from the National Science Foundation.

References

- [1] Jose L. Goity, Chandana P. Jayalath, *Strong and Electromagnetic Mass Splittings in Heavy Mesons*, arXiv:hep-ph/0701245
- [2] LHCb Collaboration, R. Aaij *et al.*, *Precision measurement of D meson mass differences*, arXiv:hep-ex/1304.6865
- [3] BaBar Collaboration, B. Aubert *et al.*, *Measurement of D_s^+ and D_s^{*+} production in B meson decays and from continuum e^+e^- annihilation at $\sqrt{s} = 10.6$ GeV*, Phys. Rev. **D65** (2002) 091104, arXiv:hep-ex/0201041
- [4] 2020 Review of Particle Physics. P.A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020)
- [5] CERN. (2018, June). Key Facts and Figures – CERN Data Centre.
- [6] CERN. (2008b). LHCb - Large Hadron Collider beauty experiment. LHCb Public. <https://lhcb-public.web.cern.ch/en/detector/detector-en.html>
- [7] CERN. (2021). LHCb. <https://home.cern/science/experiments/lhcbl#:~:text=It%20is%202021%20metres%20long,of%20Ferney%2DVoltaire%2C%20France.>
- [8] CERN. (2008). LHCb - Large Hadron Collider beauty experiment. LHCb Public. <https://lhcb-public.web.cern.ch/en/Data%20Collection/Triggers-en.html#:~:text=When%20LHCb%20is%20up%20and,million%20proton%20collisions%20every%20second.>
- [9] J. Beringer *et al.* (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).
- [10] Fritzsch, H. (2012, September 27). The history of QCD. CERN Courier. <https://cern-courier.com/a/the-history-of-qcd/>