

Ph.D. Dissertation

Top quark mass measurement using
charmed meson within b-jet

by

Geonmo Ryu

Department of Physics

The Graduate School of the University of Seoul

February 2017

Top quark mass measurement using charmed meson within b-jet

by

Geonmo Ryu

A Ph.D. Dissertation submitted to the Department
of Physics at the Graduate School of the University
of Seoul in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

December 2016

Approved by

Inkyu Park

Advisor

**This certifies that Ph.D. Dissertation
of Geonmo Ryu is approved.**

Thesis Committee Chair: Hyunsoo Min

Thesis Committee Member: Inkyu Park

Thesis Committee Member: Jason Sanghun Lee

Thesis Committee Member: Intae Yu

Thesis Committee Member: Tae Jeong Kim

The Graduate School of the University of Seoul

December 2016

Abstract

Top quark mass measurement using charmed meson within b-jet

Abstract here

Keyword: LHC, CMS, top quark, J/ψ , D0, D*, charmed meson

Ph.D. Dissertation

Top quark mass measurement using
charmed meson within b-jet

by

Geonmo Ryu

Department of Physics

The Graduate School of the University of Seoul

February 2017

Contents

1. Introduction	1
2. Theoretical overview	2
2.1 Standard Model	2
3. Experimental Setup	4
4. Computing	5
4.1 GRID Computing for LHC experiment	5
4.1.1 LHC Computing Grid(LCG)	5
4.1.2 Middleware	6
4.1.3 SRM, GridFTP and ROOT protocol	8
4.1.4 Dataset and Data Aggregation System	10
4.1.5 PhEDEx : Large Size File Transfer System	11
4.1.6 CMSSW	12
4.2 System Setup at University of Seoul	13

5. Physics object reconstruction	16
6. Reconstruction of the charmed mesons	17
7. Conclusion	18
References	19
국문초록	20
감사의 글	21

List of Figures

2.1	Elementary particles of the standard model. There are 6 flavor quarks, 6 flavor leptons, 4 gauge bosons and higgs boson,[1] . . .	3
4.1	Tier3 center at University of Seoul has a storage(left top), computing(right top), network(left bottom) and user interface(right bottom) for CMS users	14
4.2	StoRM server provides a abstract layer to support various protol. For data transfer, it runs GridFTP server and mapped SRM URL and Transfer URL(TURL). In contrast, StoRM can provide direct access point of parallel network file system like GPFS and Lustre for CE to avoid bottle neck effect. If worknodes(WNs) use indirect access though like I/O server, its bandwidth has a limitation by I/O server.	14
4.3	StoRM system consist of "StoRM Frontend and Backend". A frontend server provides a access point to SRM client	15

List of Tables

2.1	The magnitude of CKM matrix	3
-----	---------------------------------------	---

Chapter 1

Introduction

Introduction is here.

Chapter 2

Theoretical overview

2.1 Standard Model

The Standard model is a gauge theory which compiles the three of four fundamental interactions like as electromagnetic, weak and strong except gravity. As Fig 2.1, this theory describes 12 quarks and lepton include antiparticles, 4 kinds of interaction mediate bosons and higgs boson.[2] In theoretically, the model is consist of electroweak theory, QCD and higgs mechanism. The QCD is a theory about string interactions of between quarks and gluons. The electroweak theory explains about electromagnetic interaction for charged particles and weak interaction for left-handed fermion, W and Z gauge bosons. Then, the higgs mechanism explains how the fermions and gauge bosons obtained the masses.

CHAPTER 2. THEORETICAL OVERVIEW

Figure 2.1: Elementary particles of the standard model. There are 6 flavor quarks, 6 flavor leptons, 4 gauge bosons and higgs boson,[1]

$$V_{CKM} = \begin{vmatrix} V_{ud} = 0.975 & V_{ub} = 0.211 & V_{ub} = 0.005 \\ V_{cd} = 0.211 & V_{cb} = 0.974 & V_{cb} = 0.04 \\ V_{td} = 0.005 & V_{tb} = 0.041 & V_{tb} = 0.999 \end{vmatrix}$$

Table 2.1: The magnitude of CKM matrix

Chapter 3

Experimental Setup

Chapter 4

Computing

4.1 GRID Computing for LHC experiment

4.1.1 LHC Computing Grid(LCG)

The Collision data which are produced from CMS detector are too huge to handle at small size cluster. In other words, many computing resource are required to analyze. However, to gather resources to specific site is not efficiency due to maintenance cost problem. A GRID computing is a solution to solve the problem. The GRID computing is a concept and protocol to share the computing resource for valid users. For LHC computing, CERN created "LHC Computing GRID" middleware by reusing the existing GRID softwares like as EGEE's European Middleware Initiative(EMI) and Open Science GRID(OSG) middleware. On this thesis, only EMI middleware is

CHAPTER 4. COMPUTING

deal with as contents.

Overview

The middleware is a software and configuration system used in the Grid computing federation that is used to collect resources scattered around the world and conduct CMS research. This middleware basically provides services to properly install and coordinate external programs to perform 1) resource management, 2) resource availability reporting, and 3) user authentication. In a properly configured tier center, the middleware can be used to transfer resources in a large variety of ways, such as transferring large data sets or providing services for external access.

4.1.2 Middleware

This thesis basically deals only with EMI middleware. EMI middleware is a program that helps you to set the existing software that is used in accordance with the existing purpose to suit the characteristics of the hardware. Basically, the resources shared in grid computing are Computing Element (CE) and Storage Element (SE).

Computing Element

It is divided into programs such as PBS / torque and HTCondor that are used to manage jobs in a cluster computing environment to manage com-

CHAPTER 4. COMPUTING

puting resources, and programs that store the status of job managers. By default, EMI uses a package named CREAM-CE to organize the results into a program called BLAH. However, there is a package called ARC that can help you manage better on smaller tier centers. In this case, we provide separate software provided by ARC, but the final report provides information for use by senior agencies using a protocol called LDAP.

Storage Element

There are two main ways to manage storage resources. (Dcache / DPM), and 2) a method of providing resources after consolidating them (StoRM / ARC). There are different packages that need to be installed depending on the method, so you need to make good choices to make it more effective for the site administrator. Basically, in an environment where storage resources are provided by utilizing a partition system such as a SAN, a 1) method of managing files by software will be preferred. However, in the case of a method (NAS) that manages storage resources at the operating system level such as Luster or GPFS, it is necessary to select 2) method to reduce the inefficiency of storage resource access of each computational resource There is. At the user level, SEs should basically provide an externally accessible file transfer system using the SRM protocol. In addition, the ROOT protocol was created by CERN can be used to provides interactive access.

CHAPTER 4. COMPUTING

Providing information

The status of each resource should be managed by a higher authority for the purpose of grid computing. This machine is called the BDII server in EMI middleware. Each site collects the information to be provided to the user through the site-BDII service, such as the status of the resources, the status of the resource, whether or not a specific service is provided, and distributes it as the LDAP protocol. Finally, the information are provided to the user is called top-BDII, and it is not necessary to manage it at the site. and users can use it even if they provide the content to CERN top-BDII or the Taiwan who handles Asian-pacific sites.

4.1.3 SRM, GridFTP and ROOT protocol

SRM

”Storage Resource Manager” is a protocol that can be used to manage storage resources as its name suggests. In fact, this protocol is not meant to be used to send files directly, but rather to prepare for file transfers or to free up space, and does not take care of direct transmission. However, among the SE packages, the packages that provide the SRM protocol are configured to include the transport protocol, so they are not specifically distinguished at the administrator level. All grid computing sites must provide this SRM protocol by default and must provide a one-to-one mapping with the actual

CHAPTER 4. COMPUTING

transferable URL. (SRM \rightarrow GridFTP / SRM \rightarrow ROOT). With the help of this protocol, it is possible to prevent the problems of transmission failure due to the external environment that may occur when the transmission service is provided. In addition, it is also possible to configure it to be out of the dependency between the actually stored site and the URL of the file, just like the environment of cloud computing. However, the Global URL system through this SRM protocol is not used in the LCG environment.

GridFTP

In a real grid computing environment, most sites send data using a protocol called GridFTP. GridFTP is a reconfigured protocol based on the FTP protocol used in general file transfer. It has been proposed to overcome the limitations of FTP. It is built with optimized protocols such as high performance through parallel transmission, security, and minimal transmission errors. GridFTP provides access to files in grid computing environment in the form of Transfer URL (TURL). Therefore, the actual transmission must use this TURL, and if the service environment does not provide it, the transmission will fail.

CHAPTER 4. COMPUTING

Network Protocol

The ROOT framework, developed at CERN Labs, is renowned for its integrated software for high energy physics research. This program provides the I / O protocol over the network, which is utilized by many sites because of the convenience of this protocol. This protocol manages the transferable files as a ROOT IO file. If the file is a ROOT I / O file, it is possible to have interactive access in real time on the network. Although there is a performance limitation in transferring the entire file, it is very convenient to use only one file for testing, or if the file of the local cluster is stored, direct access using this protocol is very convenient.

4.1.4 Dataset and Data Aggregation System

The particle collision results measured by the CMS detector and the information produced through the Monte Carlo simulation usually have very little capacity. However, due to the theoretical limitations of the Monte Carlo simulation that we will be using, we need to deal with a great many collision events to reduce statistical errors. Therefore, for research purposes, it is necessary to use a very large set of data for the purpose of the crash event. This is called a dataset. These datasets will have their own labels according to their purpose and content. In the case of detected data, the label includes the year and the trigger information used. In the case of Monte

CHAPTER 4. COMPUTING

Carlo simulation, various information such as the type of generator used, the type of Hadronizer used, the trigger used, and the type of physical cut are included in the corresponding label. Therefore, it is easy to distinguish what contents are included even if only those labels are included. To manage these datasets, CMS manages a homepage called Data Aggregation System (DAS) using Oracle Database program. This homepage provides services for users who are not familiar with computer usage to easily retrieve and use data. It also provides information on the name of the dataset as well as the types of files included and the type of site storing the dataset. It is built to take advantage of all relevant information. Ordinary users can use the homepage directly or use a simple script called `das_client` to query the oracle DB server.

4.1.5 PhEDEx : Large Size File Transfer System

PhEDEx is a high-capacity transmission management service for data-set transfer. Unlike the commonly known concept of cloud computing, in an LCG environment, when a user runs a grid job, they wait until they can go to the sites where the datasets are stored and start working. This is because the size of the datasets is so large that it is better to move to a site that is already stored rather than importing and working with the data, but if the dataset used is not popular When demand for a dataset suddenly increases,

CHAPTER 4. COMPUTING

it can be very inefficient because it waits for grid jobs to start at the site. To solve this problem, you need to either: 1) store precious data sets in multiple sites in advance; or 2) increase the computational resources of sites that have many datasets. 2) method is not realistic because it actually requires upgrading the site, but if you store frequently used datasets in several sites in advance, you can get results faster only if you use the dataset. There is. PhEDEx is a service that allows users to request that certain datasets be saved to the desired sites in order to make these tasks easier. PhEDEx does not send the dataset immediately upon request. This is because it is necessary to coordinate dataset transfer requests so that bottlenecks do not occur in the dataset's transmit and receive sites.

4.1.6 CMSSW

CMSSW is framework software for CMS researchers. This framework contains all the algorithms and analysis codes for conducting the study. The most important work done in CMSSW is to resolve version dependencies between external programs and self-developed modules. If you use a certain version of CMSSW, the version between the external program and the internal code will be fixed, so you can minimize the bugs even if you use several programs together. On the user side, you do not have to spend time installing and maintaining these programs, so you can do your research with-

CHAPTER 4. COMPUTING

out wasting time. To maximize performance and usability, CMSSW supports the combination of `c++` code and python scripts to complete the program. Many configuration files and internal development modules are already built in a specific version. On the user side, you can create your own modules so that you can easily add new modules, overwrite existing modules, or modify configuration files. It has been taken into consideration from the design stage of CMSSW. Researchers use these programs to conduct their own research.

4.2 System Setup at University of Seoul

The University of Seoul Tier3 Center connects to the KREONET scientific research network with a bandwidth of 10Gbps, making it possible to fast transmit to overseas. Currently, the EGEE association is registered as a "KR-UOS-SSCC" grid site, and the CMS is registered as a "T3_KR_UOS" site. The "Scientific Linux 6" operating system is installed which is supplied by CERN to install packages of EMI middleware for CMS research on each machine and mount the virtual directory via the internet called CVMFS (CERN Virtual Machine File System) to keep latest update. The computational resources are constructed and reported using PBS / torque due to the lack of support about HTCondor by EMI middleware, but it has been built so that it can work on HTCondor for local users. The storage device of the Tier3 center is built with a package called StoRM. This package allows

CHAPTER 4. COMPUTING

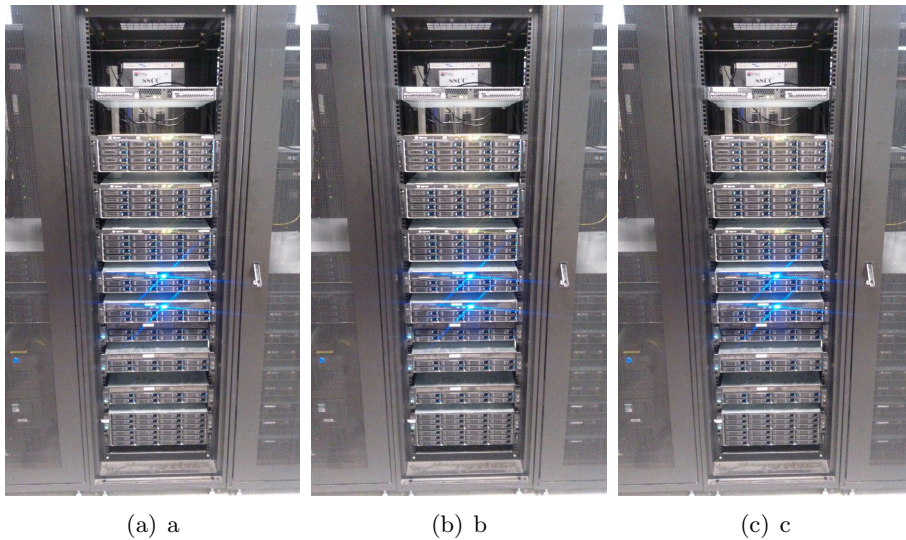


Figure 4.1: Tier3 center at University of Seoul has a storage(left top), computing(right top), network(left bottom) and user interface(right bottom) for CMS users

Figure 4.2: StoRM server provides a abstract layer to support various protocol. For data transfer, it runs GridFTP server and mapped SRM URL and Transfer URL(TURL). In contrast, StoRM can provide direct access point of parallel network file system like GPFS and Lustre for CE to avoid bottleneck effect. If worknodes(WNs) use indirect access though like I/O server, its bandwidth has a limitation by I/O server.

you to distribute data when the file system is already provided in the kernel stage, so it is a program that helps minimize the bottleneck of work in the WN where the actual work should be done. Because, the file system uses a parallel network file system called Luster, our center provides better performance when it is configured to access direct files like StoRM rather than the packages which manages files through a separate database like dCache / DPM.

CHAPTER 4. COMPUTING

Figure 4.3: StoRM system consist of "StoRM Frontend and Backend". A frontend server provides a access point to SRM client

Chapter 5

Physics object reconstruction

Chapter 6

Reconstruction of the charmed mesons

Chapter 7

Conclusion

References

- [1] Nasfarley88. Standard model of elementary particles. Wikipedia.
file:Standard Model of Elementary Particles.svg, October 2014.

- [2] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*,
C40(10):100001, 2016.

국문초록

류 건 모

서울시립대학교

대학원 물리학과

(지도교수 : 박 인 규)

[국문초록을 여기 넣어주세요.]

주요어휘: LHC, CMS, 탑 쿼크, J/ψ 중간자, D0, D*, 참 중간자

감사의 글

이후 작성 예정입니다.