UNIVERSITI MALAYA

RESEARCH PROPOSAL FOR

TOP QUARK RECONSTRUCTION WITH THE CMS DETECTOR

by

MOHAMAD ZUHAILY MAHMUD

Supervised by

DR. NURFIKRI NORJOHARUDDEEN

National Centre for Particle Physics

Department of Physics

Faculty of Science

Universiti Malaya

November 2022

Contents

1 Introduction

1.1 Problem Statement

What is the most efficient methodology to reconstruct semileptonic and hadronically decaying top quark.

1.2 Objectives

Reconstruct semileptonic and hadronically decaying top quark using various method such as chi-square fit and machine learning.

1.3 CERN, LHC & CMS

Conseil Européen pour la Recherche Nucléaire or CERN is an European laboratory for particle physics located near Geneva, Switzerland. They consist of thousands of scientists and engineers aiming to study the basic constituent of matter[1].

CERN make use the world largest and most complex scientific equipment to achieve this which is the Large Hadron Collider (LHC). Using collision of sub-atomic particle accelerated near the speed of light inside the LHC, they study how these particles interact with each other.

The LHC first ran is on 10 September 2008. From its first run until today, the operation process has been stopped three times. Once due to technical problem and the other two are for upgrading the LHC equipment. It was first stopped due to an electrical problem in the cooling system, causing a rise in the equipment temperature at about 100C on 18 September 2008 and later resume operation on 20 November 2009. Later in March 2010, CERN announce that there is a prob-

lem with the superconducting wire, requiring it to be operated only at half-energy (7TeV). It is shutdown in February 2013 to fix the problem. Year 2010 until 2013 is considered the first run of LHC. It then resumes operation in April 2015. In December 2018, there was another long shutdown in order to upgrade the LHC equipment which ended recently in July 2022. [2]. We can summarise these into three runs which is Run 1 in 2010-2013, Run 2 in 2015-2018 and Run 3 is currently running.

LHC is a powerful particle accelerator that consists of 27 kilometres ring of superconducting magnets and is located around 50-175 metres underground. Two particle beam, which are usually proton, is accelerated in opposite direction near the speed of light in the ring to be collided and is guided using these superconducting magnets at ultra-high vacuum space. This require the magnets to be chilled to -271.3C using a distribution system of liquid helium. Plus, there is another set of special magnets near the collision area to "squeeze" the particle close together to increase the probability of collision [3].

LHC consists of multiple detectors. One of them is the Compact Muon Solenoid (CMS). It is located at one of the collisions point in the LHC. CMS act as a giant, high speed camera, which capture particle collision from all direction up to 40 million times each second [4]. The transverse slice of CMS is shown in figure 1.1 display the components of the CMS detector alongside typical particle tracks that goes through inside it .
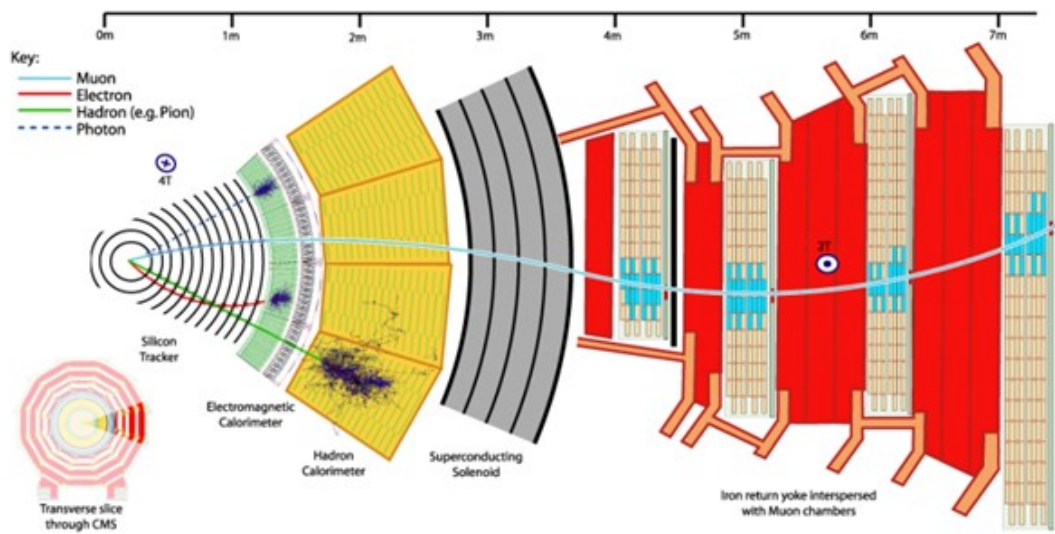
Figure 1.1: Transverse slice of CMS, showing the different sub-detectors and how different particles interact [5]

2 Literature Review

2.1 Top Quarks

Top quarks are one of the fundamental particles in the SM with a mass of 173 ±0.27 GeV/c2 with +2/3 charge and is located in the third generation particle. The invariant mass of top quarks is an interesting quality as it contributes a large portion of the electroweak radiative correction for the particles Z , Higgs and photon[6].

Top quarks that decay semileptonically and hadronically produce a set jets which a b-jet. When we talk about jet, the question "what is the importance of jets". Defining the uses of jets in physics is not simple as it is. One way of classifying their uses is according to the different possible origins for the partons that give rise to the jets. Jets can be used to determine the invariant mass of a heavy particle it originate from. Sum of the momenta of the jet should have an invariant mass that is close to their origin, which is the heavy particle [7]. Other than that, Rate of production of jets that originate radiatively such as from the emmission gluon could provides information on the value of strong coupling [8–11].

By studying the top quarks particle, we can also understand and explain the nature of the Higgs particles in more details. Higgs is relatively new particles that had just been discovered back in 2012 by the AT-LAS and CMS experiments at the LHC, CERN near Geneva, Switzerland. The Higgs boson is a special particle that gives mass to fundamental particles and itself. The Higgs decays predominantly to $b\bar{b}$, yielding signals that is quite like $t\bar{t}$. This entails that we can integrate the $t\bar{t}$ identification into the Higgs boson analysis. This is done by kinematic mass reconstruction and cross section measurement. If both values differ significantly, this would hint at a new physics[12].

2.2 Top Quark Production

So far, $t\bar{t}$ production that have been recorded are from Tevatron and LHC. From the CDF and D0 experiments at Tevatron, they are able to conclude a combined value of cross-section for $t\bar{t}$ production which is $\sigma_{t\bar{t}} = (7.60 \pm 0.20(stat) \pm 0.36(syst))pb$ assuming a top-quark mass of $m_t = 172.5 GeV$ with precision of 5.4%. Meanwhile, in 2015 at the LHC, a $p\bar{p}$ collisions at a center-of-mass energy of $5TeV$ results in an inclusive $t\bar{t}$ production. The inclusive $t\bar{t}$ cross section was measured using a data sample corresponding to an integrated luminosity of $26pb^-1$ [13].
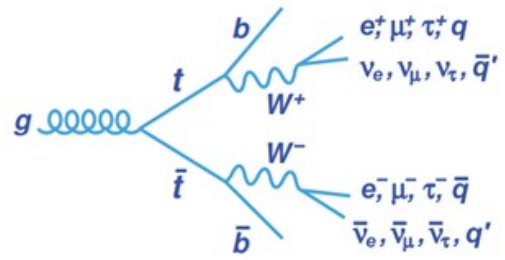
2.3 Top Quark Decay

Figure 2.1: Leading-order Feynman diagram for $t\bar{t}$ decay [14]

The decay of $t\bar{t} \rightarrow W^+bW^-\bar{b}$ has a branching ratio (BR) close to unity described by the Cabibo-Kobayashi-Maskawa (CKM) matrix, $V_{tb} \approx 0.9992$, meaning that this decay is exclusively used for top-

3

quark measurement. There is also a very small chances that it would decay into strange quark and down quark, $V_{tb} \approx 0.0387$ and $4V_{tb} \approx 0.0084$ respectively. The W can decay hadronically or leptonically. This leads to the final states of $t\bar{t}$ events to be either "lepton+jets" ($t\bar{t} \rightarrow \ell^+\nu b q\bar{q}'\bar{b}$ and $qq'b\ell^-\bar{\nu}\bar{b}$), "all jets" ($t\bar{t} \rightarrow q\bar{q}'b\bar{q}q'\bar{b}$) and "dileptons" ($t\bar{t} \rightarrow \ell^+\nu b\ell^-\bar{\nu}\bar{b}$).

The measurement performed at LHC of the top quark decay process results in a solid agreement with the SM prediction [14].

Figure 2.2 and figure 2.3 show the statistics of $t\bar{t}$ cross section from the ATLAS and CMS obtained from [13].
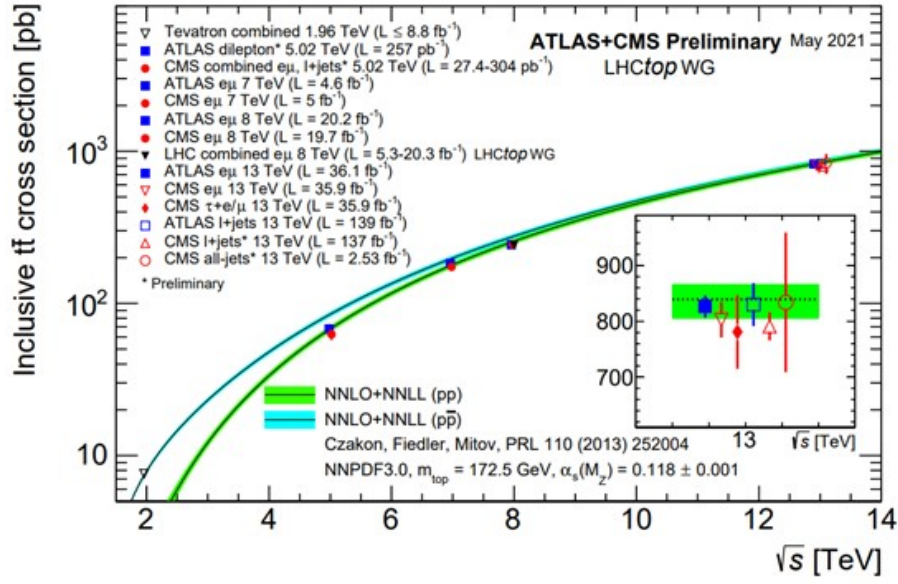
Figure 2.2: Measured and predicted $t\bar{t}$ production cross sections from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp [15]

Measurement	$\sigma_{t\bar{t}}$ (pb)	stat (%)	exp (%)	th (%)	lumi (%)
ATLAS $e\mu$ 7 TeV	182.9	1.7	2.3	2.0	1.8
CMS $e\mu$ 7 TeV	173.6	1.2	$^{+2.6}_{-2.3}$		2.2
ATLAS $e\mu$ 8 TeV	242.4	0.7	2.3	3.1	1.7
CMS $e\mu$ 8 TeV	244.9	0.6	$^{+2.6}_{-2.2}$		2.6
ATLAS $e\mu$ 13 TeV	818	1.0	3.3		2.7
CMS e/μ+jets 13 TeV	834.6	0.3	2.7		2.7

Figure 2.3: Summary of most precise inclusive $t\bar{t}$ cross section measurements from the ATLAS and CMS [15]

4

2.4 Current Studies on Top Quarks

Today, most top quarks investigations are performance studies and precision measurement. What technique is the best to reconstruct the top quarks with the most precision efficiently. By efficiency, we are talking about how fast the analysis can be, how precise it is, how many memory storage consumption can we reduce and many more. Slight difference in measurement could indicate a hint to new particle or new physics that describe fundamental particle interaction beyond the Standard Model (SM). However, in the search of something new beyond the SM, there are a much more complicated works that need to be done. One of the challenges in reconstructing top quark decay is that decay leptonically produce neutrino that is hard to detect. This led to reconstruction based only on conservation of momentum. Plus, assigning jets to their origin is a challenge by itself as one collision produced multiple jets output. In all-hadronic state, exist huge amount of background jets due to QCD multijet production [16].

Various studies have been done prior to this date in term of reconstructing top quark production efficiently. Most of them turned to deep machine learning or neural network in reducing the complexity of reconstructing top quark [6, 17–20]. [17] use kinematic fitting to identify which leptons and jets that is produced via $t\bar{t}$ decay. [6] make use of the δR method which select two closest jet to reconstruct the decay. Meanwhile [18] make use of Lorentz's four momentum concept as an inspiration for their Lorentz Neural Network (LNN). [21] and [22] also studies on methods in resolving top quark reconstruction, focusing on semileptonic $t\bar{t}$ decay reconstruction without using machine learning. It algorithm could be used as an inspiration for the much more simpler semileptonic top decay reconstruction. Nevertheless, from the reading, we can conclude that the most efficient way to reconstruct top quark decay is by using machine learning techniques.

3 Brief Project Overview

3.1 Analytical Techniques

Based on literature review, the most effective way to reconstruct top quarks are by using machine learning techniques. Due to this, the project aims to use machine learning to reconstruct hadronically decaying top quarks. Technique from [6] and [19], the χ^2 techniques are a good starting point to start this project. Other than that, we can also use top quark analysis object for jets such as the jet algorithm, jet energy scale and resolution, b-tagging algorithm, missing transvers momentum and the flow particle approach [15] are also few of the option approaches to reconstruct the hadronically decaying top quark.

3.2 Computational Studies

Since most of the previous studies concluded that machine learning is the best way to approach the problem, the need to understand and utilize machine learning is an absolute. Using python language, two suitable libraries are the Scikit and TensorFlow [23]. Both are a free machine learning libraries for Python. The reconstruction of top quark also will make use of the ROOT library designed by CERN specifically for high energy physics research. This will be run on Windows subsystem Linux (WSL). Secure Shell Protocol (SSH) to CERN machine would also help in running the program since thousands to millions of data would be used in this project to train the machine learning.

3.3 Monte-Carlo samples

This project will use Monte-Carlo generated samples [24] to train the machine learning program to rec-

ognize the jets that belongs to the top quark. Performance studies will then be used to rate how a certain analysing technique performed.

3.4 Resources

This project is under computational physics category as it studies the computational method to reconstruct the top quark decay. Thus, this project will be solely being done on the computer whether it is locally or using CERN server as the hardware to run the programs. Fortunately, Dr Nurfikri Norjoharuddeen has provide the CMS CERN membership to all his undergraduate final year project students. Being one of the official CMS members, we are now able to utilize some of the utility provided by CMS. The programs that are specifically installed to run this project are in table 2.

No	Program
1	Windows Subsystem Linux (WSL)
2	Visual Studio Code (VS Code)
3	Miniconda (including ROOT library)
5	WinSCP

Table 3.1: Computer programs that are going to be use in this project

All the programs that are listed in table 1 are free to use. Another important equipment that is required is a good internet connection as this project will use a lot on CERN online utilities applications. The most important one is the CERN Box.

3.5 Gantt Chart

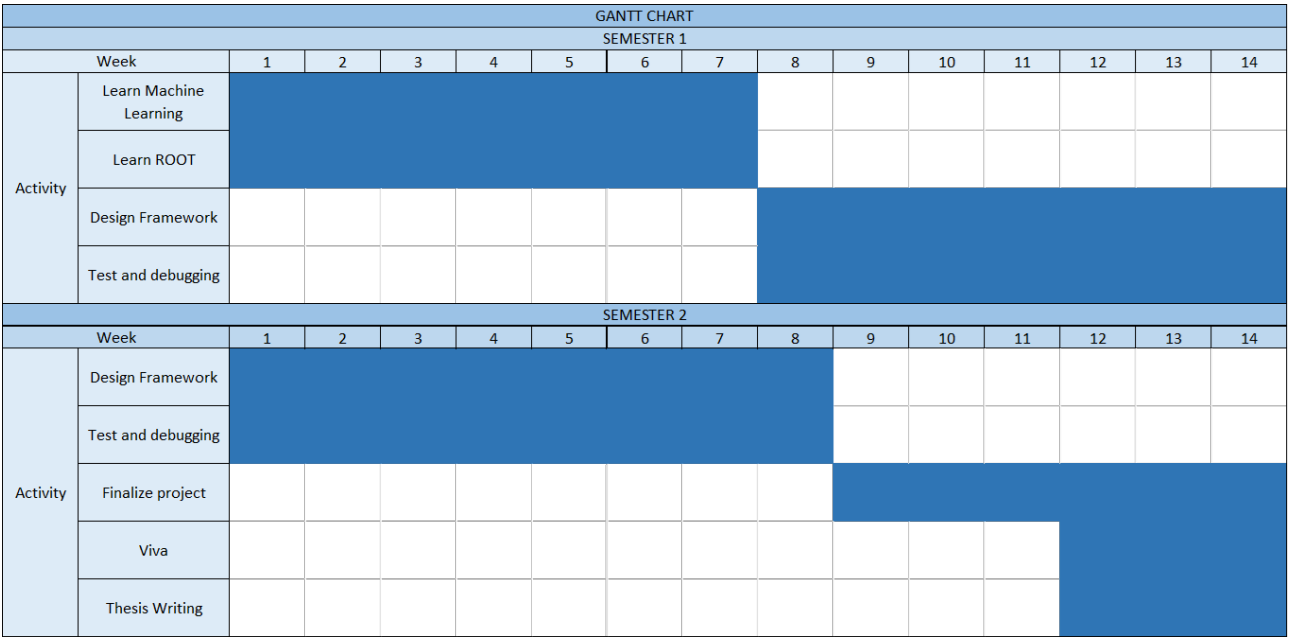| GANTT CHART															
SEMESTER 1															
Week		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Activity	Learn Machine Learning	██	██	██	██	██	██	██							
	Learn ROOT	██	██	██	██	██	██	██							
	Design Framework								██	██	██	██	██	██	██
	Test and debugging								██	██	██	██	██	██	██
SEMESTER 2															
Week		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Activity	Design Framework	██	██	██	██	██	██	██	██						
	Test and debugging	██	██	██	██	██	██	██	██						
	Finalize project									██	██	██	██	██	██
	Viva												██	██	██
	Thesis Writing												██	██	██

Figure 3.1: Gantt chart of the project

4   Summary

The final goal of this project is to create or at least recreate a particle reconstruction for semileptonic and hadronic top quark decay using machine learning techniques. Being able to create a few of the analysis techniques and compare their performance efficiency is already a good enough result. From the literature review, we can hypothesis that the χ^2 techniques will probably be the most efficient one since most of the papers conclude so.

References

[1]   Christine Sutton. *CERN*. Sept. 2022. URL: https://www.britannica.comhttps//www.britannica.com/topic/CERN.

[2]   David G.C. Jones. *Large Hadron Collider*. Aug. 2022. URL: https://www.britannica.comhttps//www.britannica.com/technology/Large-Hadron-Collider.

[3]   CERN. *Large Hadron Collider*. Oct. 2022. URL: https://www.home.cern/science/accelerators/large-hadron-collider.

[4]   CERN. *CMS*. Oct. 2022. URL: https://home.cern/science/experiments/cms.

[5]   Dave Barney. *An overview of the CMS experiment for CERN guides*. 2003. URL: https://cds.cern.ch/record/2629323/files/CMSdocumentforGuides.pdf.

[6]   Kiarash Naderi. "Deep Learning Techniques for Top-Quark Reconstruction". In: (2017). URL: https://cds.cern.ch/record/2280195.

[7]   Gavin P. Salam. "Towards jetography". In: *The European Physical Journal C* 67.3-4 (May 2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-010-1314-6.

[8]   G. Abbiendi et al. "Measurement of the strong coupling α_s from four-jet observables in e^+e^- annihilation". In: *The European Physical Journal C* 47.2 (June 2006). DOI: 10.1140/epjc/s2006-02581-y. URL: https://doi.org/10.1140%2Fepjc%2Fs2006-02581-y.

[9]   F. D. Aaron et al. "Jet production in ep collisions at high Q 2 and determination of α_s". In: *The European Physical Journal C* 65.3-4 (Nov. 2009), pp. 363–383. DOI: 10.1140/epjc/s10052-009-1208-7. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-009-1208-7.

[10]   S. Chekanov et al. "Inclusive jet cross sections in the Breit frame in neutral current deep inelastic scattering at HERA and determination of α_s". In: *Physics Letters B* 547.3-4 (Nov. 2002), pp. 164–180. DOI: 10.1016/s0370-2693(02)02763-6. URL: https://doi.org/10.1016%5C%2Fs0370-2693%5C%2802%5C%2902763-6.

[11]   "Measurement of the energy dependence of hadronic jet rates and the strong coupling α_s from the four-jet rate with the DELPHI detector at LEP". In: *The European Physical Journal C* 38.4 (Jan. 2005), pp. 413–426. DOI: 10.1140/epjc/s2004-02060-7. URL: https://doi.org/10.1140%2Fepjc%2Fs2004-02060-7.

[12]   Peter Bärnreuther. "Top quark pair production at the LHC". PhD thesis. Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University, 2012. URL: https://d-nb.info/1025921933/34.

[13] R.L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097. URL: https://cds.cern.ch/record/2836514.

[14] T. Aaltonen et al. "Combination of the top-quark mass measurements from the Tevatron collider". In: *Physical Review D* 86.9 (Nov. 2012). DOI: 10.1103/physrevd.86.092003. URL: https://doi.org/10.1103%2Fphysrevd.86.092003.

[15] Ulrich Husemann. "Top-quark physics: Status and prospects". In: *Progress in Particle and Nuclear Physics* 95 (July 2017), pp. 48–97. DOI: 10.1016/j.ppnp.2017.03.002. URL: https://doi.org/10.1016%2Fj.ppnp.2017.03.002.

[16] Regina Demina and Evelyn J. Thomson. "Top Quark Properties and Interactions". In: *Annual Review of Nuclear and Particle Science* 58.1 (Nov. 2008), pp. 125–146. DOI: 10.1146/annurev.nucl.58.110707.171224.

[17] J. Erdmann et al. "From the bottom to the top—reconstruction of $t\bar{t}$ events with deep learning". In: *Journal of Instrumentation* 14.11 (Nov. 2019), P11015. DOI: [10.1088/1748-0221/14/11/P11015](10.1088/1748-0221/14/11/P11015). URL: https://dx.doi.org/10.1088/1748-0221/14/11/P11015.

[18] Greif, Kevin and Lannon, Kevin. "Physics Inspired Deep Neural Networks for Top Quark Reconstruction". In: *EPJ Web Conf.* 245 (2020), p. 06029. DOI: 10.1051/epjconf/202024506029. URL: https://doi.org/10.1051/epjconf/202024506029.

[19] G. Aad et al. "Measurement of the top quark mass with the template method in the $t\bar{t} \to$ lepton + jets channel using ATLAS data". In: *The European Physical Journal C* 72.6 (2012), p. 2046. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-2046-6. URL: https://doi.org/10.1140/epjc/s10052-012-2046-6.

[20] R. Santos et al. "Machine learning techniques in searches for $t\bar{t}h$ in the $h \to b\bar{b}$ decay channel". In: *Journal of Instrumentation* 12.04 (Apr. 2017), P04014–P04014. DOI: [10.1088/1748-0221/12/04/p04014](10.1088/1748-0221/12/04/p04014). URL: https://doi.org/10.1088%2F1748-0221%2F12%2F04%2Fp04014.

[21] Jiří Kvita. "Study of methods of resolved top quark reconstruction in semileptonic $t\bar{t}$ decay". In: *arXiv.org* (2018). DOI: 10.1016/j.nima.2018.05.059,10.1016/j.nima.2022.167172. URL: https://arxiv.org/abs/1806.05463.

[22] Regina Demina, Amnon Harel, and Douglas Orbaker. "Reconstructing $t\bar{t}$ events with one lost jet". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors*

and Associated Equipment 788 (July 2015), pp. 128–136. DOI: [10.1016/j.nima.2015.03.069](). URL: [https://doi.org/10.1016%2Fj.nima.2015.03.069]().

[23] Aurelien Geron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly, 2017.

[24] Aurelien Geron et al. *Mathematical Method for Physics and Engineering Third Edition*. Cambridge University Press, 2006.