BIG DATA IN ECOLOGY

By

MORTEZA SHAHRIARI NIA

A DISSERTATION PROPOSAL PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHYLOSOPHY

UNIVERSITY OF FLORIDA

2015

I dedicate this to my parents who selflessly devoted the best of their lives to raise their kids to be hard working, humble, selfless and to take steps towards the greater human wisdom

ACKNOWLEDGMENTS

I would like to thank Dr. Daisy Zhe Wang for believing in me and providing the great opportunity of tackling the whole real of big data. I would also like to thank Dr. Yuguang Fang for his great support and commitment. Dr. Paul Gader and Dr. Stephanie Bohlman were great mentors, without contributions of whom this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Proposal Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Phylosophy

BIG DATA IN ECOLOGY

By

Morteza Shahriari Nia

April 2015

Chair: Dr. Daisy Zhe Wang
Major: Electronics and Computer Engineering

Ecological sciences benefit from the huge diversity of plant species which play an important role in large scale ecological aspects such as global warming, land cover change, CO^2 emission, invasive species, fire hazard, and etc. State-of-the-art species classification techniques utilize remote sensing data such as hyperspectral and LiDAR, however this task involves plenty of field data collection which is both highly time consuming, costly and can only be accomplished by ecological experts. Among thousands of the most commomly found plant species there is huge similarities between them from a remote sensing point of view which makes the task of species classification very daunting; therefore we see a whole body of literature specifically dedicated to this issue which is yet far from real world scenarios with thousands of possible species. While this is an indicator of the importance and complexity of the issue, little has been done to tackle the problem from a computational point of view harnessing the power of "big data". Periodic airborne campaigns can generate terrabytes of data on vast swaths of land. To tackle these problems we propose to use probabilistic knowledge bases and deep learning both of which work best when there is lots and lots of data. Probabilistic knowledge base captures ecological expert knowledge in terms of probabilistic rules, which will be maped to remote sensing data and used to infer new facts and therefore enhance species classification accuracy. Deep learning on the other hand as a semi-supervised algorithm will benefit

from the vast amounts of data available and capture intrinsic features of data through its layered architecture and thus help in reducing the amount of labeled data required.

CHAPTER 1
INTRODUCTION

Extracting knowledge at web-scale and reasoning based on the aggregate constructed massive knowledge base is turning into reality with the availability of massive computational and storage capabilities. Knowledge is any kind of information that is readily available for end user consumption without limitations of the traditional search and lookup approach. Once a user creates a query, only the best known answer to that specific query will be returned and this alleviates the user from looking up traditional search results and eyeballing contents until the desired information is found deep in documents. As an example, if the user queries What plant communities does longleaf pine usually belong to in Florida? a knowledge based system would return Sandhills, upland pine forests, mesic flatwoods, scrubby flatwoods, and seepage slope with probabilities assigned for each. You can further expand on these and get more structured detail. Whereas traditionally you would do a keyword search, face a long list of documents containing related words to the query and you would have to look into and find the answer for yourself. This concept can be extended to domain specific areas such as ecology, forestry, geology and etc.

In this project we plan to construct a knowledge base for ecological domains. We will parse and analyze thousands of resources, whether textual or non-textual (hyperspectral, LiDAR, etc), to be able to represent similar answers as above for scientific queries. Resources include but are not limited to plain text of scientific publications, images, tabular content, hyperspectral images, LiDAR point clouds, already available data repositories and so on. There are several steps being taken in this regard in other domains with shared storage and/or computation facilities such as iDigBio, iPant, iAnimal and etc. The closest to our perspective of constructing a knowledge base is the DeepDive series from Stanford University which have constructed different knowledge bases for Geology and Paleobiology.

I. Data Integration Constructing a biological knowledge base requires acquiring and combining structured knowledge from various plant databases such as TRY database (121 datasets covering a wide variety of biomes and geographic areas), Trait databases such as TraitNet , Land Cover (planetary diversity, trends, and aquatics) along with other major data portals with hundreds of contributed expert datasets such as ecologicaldata.org and datadryad.org. DataOne (dataone) and EarthCube (earthcube) are other recent projects funded by NSF that provide scientific data archiving resources for ecological, environmental and geosciences data produced by scientists worldwide. The NSF-sponsored National Ecological Observation Network (NEON) project is huge national investment to systematically collect continental-scale ecological data, including field data on all biological levels of organization from soil microbes to mammals, energy and CO2 flux data, and high resolution remote sensing images, at over 100 observation sites in 20 eco-climatic domains for the next 30 years. NEON will be collecting 128 high level datasets, which includes 18 bioclimate, 31 bio diversity, 30 biogeochemistry, 10 ecohydrology, 7 disease, and 22 land-use data sets (keller2010neon). These are apart from hundreds of low level data types being collected (keller2010neonLevel). Besides well-structured data, there is a great wealth of data buried deep in textual content. We plan to investigate the use of Natural Language Processing techniques to extract additional facts on individual plants from plain text data (unstructured) available in scientific publications, encyclopedias or crowd sourced data sources such as Wikipedia or Citizen Science. Challenges in constructing biological knowledge base from unstructured text involve large-scale recognition and co-reference resolution of known entities (e.g. plants, geospatial names), determining relations between entities, automatic geo-location and temporal content identification. Facts can be extracted on plant distributions, color description, habitat localization, and temporal variations on seasonal changes as well as relationships between different plants such as correlation and co-existence patterns, among others. U.S. Department of Agriculture (USDA) Plants Database for example, provides

a long list of features for Longleaf pine specie, each of which has certain significance in distinguishing this species from others in close families that might be hard to detect only relying on hyperspectral/LiDAR data. These are all readily available knowledge ready to be integrated in automated processes.

Projects such as iDigBio(idigbio) and BudBurst(BudBurst) put a step further and utilize crowd sourcing techniques to collect and annotate a variety of biodiversity data which includes images of species, transcription of museum bio-collections, sound and video samples, genetic information, geo-location, taxonomic, and provenance information. There are other parallel projects such as iPlant/iAnimal (iplant) (a gene level data portal) that besides providing data storage and sharing facilities also provide some computation capabilities and code repositories. Such systems remain passive and let the scientists do all the labor if they may.

Expert knowledge is only available through years of experience and active study of relevant scientific papers. This information is not available and we plan on utilizing Natural Language Processing and Data Mining techniques to extract location-aware probabilistic co-associative species list. All the information buried deep in tables (e.g. in the USDA Plants Database), actual text, charts, figures in papers. Such information co-occurred species rival and invasive species per area are all documented but the data is not available up-to-date and ready to be used. Accommodation of big data knowledge base techniques for ecological applications can be very beneficial but demands dedicated time and collaborated expertise of both computer scientists and ecological scientists. Apart from field data which is of obvious benefits we are going to need structured textual knowledge which would help us train our models and extract additional information from text and needless to say that these need to be built collaboratively. We can use techniques to extract and unify this information to a format such that potential future applications would be possible. Similar projects are being funded one of which is the GeoDeepDive from Hazy group at Stanford which are geared more towards geological

concerns. GeoDeepDive collects information from tens of thousands of scientific papers to collect information about geological entities and tries to aggregate this information geared towards various user queries.

The diversity in collected data can help in addressing research questions that were never possible before. For instance, this can enable correlating geographic location of plant specimen occurrences to animals, insects and climate information over time at ecological scale. This can lead to better understanding of the mutual impact of climate change and bio-diversity, precision invasive species monitoring at continental scale, fire potential and spread route prediction among other applications. By understanding which species would be higher carbon reserves we can have educated plans to control carbon emissions and maintain climate consistency and fight global warming. This is all possible after utilizing remote sensing techniques at ecological scale that surpass local measurements and field works which are not scalable in nature. At this level we can study the impact of biotic and abiotic factors on species distributions and invasive species impacts. This can potentially lead to identifying the Achilles' heel of invasive and exotic species. All this can have huge economic impacts whether directly through their affects in farming-related industries or indirectly by facilitating a healthier and more sustainable environmental studies. With plans on periodic flights of NEON AOP facilities we can study species through time with high spectral and spatial resolutions. This can be combined with other sources of knowledge such as MODIS and HyspIRI that have been collected for a long time which are of lower resolution.

II. Knowledge Base Construction Inference There is a growing body of work on automatic knowledge base construction (Weikum and Theobald). This literature can be clustered into 4 main groups: (1) approaches such as YAGO (Suchanek et. al), YAGO2 (Hoffart et. al), DBpedia (Auer et. al), and Freebase (Bollacker et. al), which are built on Wikipedia infoboxes and other structured data sources; (2) approaches such as Reverb (Fader et. al), OLLIE (Mausam et. al), and PRISMATIC (Fan et. al), which use open

information (schema-less) extraction techniques applied to the entire web; (3) approaches such as NELL/ ReadTheWeb (Carlson et. al), PROSPERA (Nakashole), and DeepDive/ Elementary (Niu et. al), which extract information from the entire web, but use a fixed ontology/ schema; and (4) approaches such as Probase (Wu et. al), which construct taxonomies (is-a hierarchies), as opposed to general KBs with multiple types of predicates. The recent publication by Google called The Knowledge Vault (Dong et. al) is most similar to methods of the third kind.

Other school of thought is on integrating and fusion of knowledge bases such as (Dong et. al) where the goal is that given we have various knowledge bases available to us how to merge these together and have metrics on reducing errors and resolving conflicts. Knowledge fusion identifies true subject-predicate-object triples extracted by multiple information extractors from multiple information sources. These extractors perform the tasks of entity linkage and schema alignment, thus introducing an additional source of noise that is quite different from that traditionally considered in the data fusion literature, which only focuses on factual errors in the original sources.

We are working on encoding such knowledge as probabilistic rules over the temporal cross-continental image data and use Markov Logic Networks (MLN) (Richardson and Domingos 2006) to perform probabilistic reasoning. There has been several works (Dennison and Roberts 2003, Powell et at. 2007, Roth et al. 2012) on utilizing a regionally specific spectral library on extracting endmember (plant species) abundances from a hyperspectral image. By combining such approaches in a distributed computing infrastructure with MLN inference engine, we can eliminate the cost and time required for field plant identification and reach quality plant classification at ecological scale. Due to the complexity of the knowledge base construction algorithms, including extraction, reasoning, and the size of the open domain, big data analysis techniques need to be applied.

Authors have made extensive contributions to large-scale information extraction using statistical machine learning on massively parallel frameworks (wang2010probabilistic, wang2010querying, wang2011hybrid). This work is included in popular open source analytics framework MADLib (hellerstein2012madlib, grant2012madden). The authors have expertise in constructing knowledge base from open domains using NLP methods (wang2012automatic). Needles to say that this does not come for free, as building and maintaining knowledge bases is stat-of-the-art research, we several challenges: scientific text annotation, fact extraction, rule construction, noisy facts/rule identification, weight learning, scalable inference, and so on. These are issues that are being actively investigated through major academic institutions as well as top-notch industry research centers. Accommodating existing approaches to the field of ecology would need some expert knowledge to verify current systems status and extensive tests and experiments to ensure that the whole pipeline works as expected. This also requires a cluster computing environment to be able to reach solutions in reasonable time frames.

1.1 Proposed Work

This office is a division of the UF Graduate School and the Office of the Provost and is in place to ensure consistent formatting for all documents published by the University of Florida. This office is comprised of the editors, who will review the document after submission to the Graduate School. [5]

1.1.1 224-B HUB (352) 392-1282

The Graduate School Editorial Office provides definitive answers for unique formatting issues not addressed by the formatting guidelines or the consultants in the Application Support Center (ASC). We work hand-in-hand with the ASC Lab in our mission to help students navigate smoothly through the thesis and dissertation submission process.

1.1.2 gradedit@aa.ufl.edu

Questions regarding the following items should be addressed to the Graduate School Editorial Office: [9, 10, 12]

1. First and final submission requirements
2. Final clearance requirements
3. Required forms
4. Clearing prior
5. Required journal article

In addition, the Graduate School Editorial Office can provide assistance with any questions regarding the online Electronic Document Management (EDM) system. For more information from the Graduate School, consult the following: If you are a doctoral student, please visit our checklist:

http://www.graduateschool.ufl.edu/files/checklist-dissertation.pdf

If you are a master?s thesis student, please visit our checklist:

http://www.graduateschool.ufl.edu/files/checklist-thesis.pdf

All posted deadlines are available at

http://www.graduateschool.ufl.edu/files/editorial-deadlines.pdf

Graduate Catalog:

http://gradcatalog.ufl.edu/

The Editorial Office is comprised of

- Lisa De LaCure, Editor
- Anna Pardo, Editor
- Stacy Wallace, Coordinator
- Cara Mannion, Office Assistant

1.2 Application Support Center (ASC)

The Application Support Center office is a division of Academic Technology as a section of the Help Desk. It is in place to provide assistance to students formatting their thesis or dissertation for publication by the University of Florida. This office is not a part of the Graduate School.

1.2.1 224 HUB (352) 392-4357 (Choose Option 5)

However, the office does offer technical assistance to graduate students. The office is comprised of the technical consultants, who report to the UF Help Desk. In addition to their other duties, they will help guide thesis and dissertation students through the submission process and will troubleshoot the preliminary electronic document BEFORE submission to the Graduate School.

1.2.2 asc-hd@ufl.edu

The ASC Lab employs consultants for the purpose of assisting students with formatting issues such as:

- using the template
- formatting figures or tables
- compiling the document

Typically, students bring their files on a portable media storage device, or laptop, for troubleshooting. Allowing consultants to view the pdf of your file in order to make suggestions regarding format. Consultants are not responsible for fixing any formatting issues but will point out formatting errors for you to fix. If there are any issues which you do not know how to correct we will make every effort to find a solution to your problem and help you correct the issue. Antonio and Chaouche [4]

The ASC is comprised of

- J. K. Booth, ASC Manager
- The ASC Support Staff Team

There have been no major formatting changes since August, 2006. However, we have made several improvements to the LaTeX Template since then so it's always advisable to work with latest version possible. Download the .zip file that comprises the LaTeX Template and extract the folder containing the template files. Change the name of the folder to avoid any possibility of overwriting your files should you need to re-extract the folder again later in the process. [5]

1.3 Basic Tools Needed

You must have access to an installation of LaTeX. We use MiKTeX (current version is 2.9) and suggest you download and install the complete (approx 500 MB) version rather than the minimal installation. Our template uses several packages in addition to the basic build that are absolutely necessary for it to work. [9, 10, 12]

1.4 MiKTeX

This template assumes that you have MiKTeX 2.5 or later installed on your system. This setup is supposed to auto-install any packages that are called but not already installed as long as you are connected to the Internet. We recommend that if you need to install MiKTeX that you select the complete installation if possible.

1.4.1 Required Files and Programs

To correctly implement the UF ETD LaTeX 2_εtemplate in accordance with the UF Gradschool Editorial Office Guildlines. The following files and/or packages are required:

1. MiKTex
2. Some text editor
3. Hanging Package
4. Caption Package
5. Hyperref Package

This is an example of a "short" list. Not because there's only 5 items on the list but because each item is less than one line in length. Since short lists are relatively rare the default spacing for the itemize and enumerate environments is for the "long" list where at least one item on the list wraps to a second line. In order to generate a correct short list you need to insert a `\vspace{-10pt}` command after all but the last item on the short list.

The enumerate and itemize environments have been modified to meet Editorial Office guidelines (A special thank you to Antonio Paiva for both the suggestion and the code) but require the ufenumerate.sty file to be in the same folder as your main file. It is loaded

via the usepackage{ufenumerate} command. The itemize environment is modified by a set of commands in the usersetcommands file.

1.4.2 Optional Files and Programs

If you need to add a package please remember to place it before the hyperref package in the packages file. Hyperref needs to have several modifications in place to work properly and is essential for the required links. To ensure that it works correctly it must be the last package loaded - and even then it's a delicate operation.

The following programs are not needed but may be very useful when editing documents in LaTeX:

- WinEDT: This text editor is recommended for use editing TeX-files as it has many useful built in macros and is easy to use
- This program can be found and downloaded here: http://www.winedt.com/
- The GIMP (GNU Image Manipulation Program)
  - A freeware graphics editing program for picture editing and file conversions
  - Comparable to Adobe Photoshop
  - Can be downloaded here: http://www.gimp.org/
- A good reference of LaTeX2ϵ commands
  - This should be included on the ETD website here: http://etd.helpdesk.ufl.edu/tex.php

This is an example of "nested" lists. In the itemize environment you can choose an alternative symbol for the "sub-list." The method of specifying this symbol is `\item[-]` where the optional symbol is inserted into the square brackets. Unless you are referring to an item by number, itemized lists are generally preferable to enumerated ones. The difference between itemize and enumerate environments is illustrated by repeating this list below:

1. WinEDT: This text editor is recommended for use editing TeX-files as it has many useful built in macros and is easy to use

2. This program can be found and downloaded here: http://www.winedt.com/

3. The GIMP (GNU Image Manipulation Program)

   (a) A freeware graphics editing program for picture editing and file conversions
   (b) Comparable to Adobe Photoshop

   (c) Can be downloaded here: http://www.gimp.org/

4. A good reference of LaTeX2ϵ commands

   (a) This should be included on the ETD website here: http://etd.helpdesk.ufl.edu/tex.html

1.5 Test Compile Before You Start

The easiest way to compile the template is to double-click the "make.bat" file included in the template folder. Unfortunately this file only works in a Windows environment. Make sure you can compile the standard template BEFORE you start putting your content into the files (just to be on the safe side). Since there is little resemblance between the standard TEX Report Class and the ufthesis.cls styles when you latex the ufsampleETD.tex file it will generate several warnings possibly even an error or two. If you latex the file and it stops compiling because of an error press "r" then "enter" and LaTeX will ignore the rest of the errors and warnings. Latex the file again, also pressing "r" and "enter". When all of the warnings are done you can then dvipdfm the file. (Using WinEDT, I just click the dvi - pdf button on the toolbar).

This should generate the sampleETD PDF file. As long as you get a PDF in the correct format the errors and warnings generated by LaTeX up to this point are irrelevant. However, if the output at this stage is garbled or non-existent we need to do some troubleshooting: If you've simply unzipped the template and it fails to compile without having made any changes make sure you have the FULL MikTeX installation. There are many LaTeX editors available and almost as many ways to compile a TEX file.

The method we use is to latex filename twice, then dvipdfm filename to create the pdf. Some Editors are more sensitive to errors than others and are unable to bypass the

errors. Some others use a different method of compiling the file and can't be re-configured. The template works well on our set-up. If you can't get the template to compile on your machine using the ufthesis.cls file, change the style of the document to the standard report class. Hopefully, you will then at least be able to see the content. Once your content is ready you can come in to the ETD Lab to compile your document in the correct style.

CHAPTER 2
KNOWN ISSUES

2.1 Common Problems

2.1.1 Can't Find dvipdfm

If you're using a MacIntosh and typesetting your file results in an error stating it can't find the file dvipdfm. That option is listed in three places and will need to be changed. The preamble of the main file, and two places in the packages.tex file (the graphix package and the hyperref package). Replace dvipdfm with dvipdfmx and typeset using XeLaTeX and that should take care of the problem.

2.1.2 Prime Notation

If you've placed your own content into the template and it fails to compile the most common reason is the use of prime notation in math mode. If you say anything like A' while in the math mode it generates a conflict that will cause the rest of the document to stop compiling. The solution is to use an alternative method of producing the same notation. Replace `A'` with `A^{\prime}` and you will avoid this issue. Since prime notation is quite common we defined a shortcut for this command which allows you to replace `A'` with `A\p`. [7]

2.1.3 Long (and/or Wide) Tables

Another problem in LaTeX is the inability to handle long tables. While there are some packages that address this problem none of them quite fit the Editorial Office guidelines. The caption is not repeated but we do need "Table x-y. Continued" on each subsequent page and a repeat of the column headings on each page as well. The following table is the best example of the correct format I can produce. The disadvantage of this method is that much of it is manually set up and changes in the text will cause changes in the table. [13]

For example, at one time the following table was perfectly positioned at the beginning of a page - right after a full page of text. Now the footnote appears on the page with

text that is two full pages before the footnote mark! For best results avoid the use of footnotemark and footnotetext commands inside of tables and try to keep your footnotes outside of floats whenever possible.

2.1.4 Page Size

When installing MiKTeX one of the set up questions asks for the default paper size. Unless you select "Letter" it will default to A4. Although the Template specifies letter size in several locations an installation with an A4 default will create a document using A4 paper size. The Editors will notice that your margins are incorrect and comment on this during first submission (it probably won't cause rejection on first submission). However, it will cause a problem for final approval. The only solution I know of is to re-install MiKTex and give the "Letter" size as the default paper size. The alternative would be to compile your document on an AT Lab computer which should have all been set to letter size when the Hard Drive Image was created.

2.1.5 Single Appendix

Since LaTeX numbers everything automatically there is an interesting problem that occurs whenever an author wants to create a document with a single appendix. When this happens, the Editorial Office states that instead of "numbering" the appendix as "A" the word "APPENDIX" should appear in the Table of Contents on the same line as the appendix title without the letter "A." This can be done by suppressing the chapter numbers but then anything that needs to be numbered (Tables, Figures, and/or equations is numbered as a continuation of the previous chapter. This is fine if there's nothing to number, but most of the time this is not the case.

The file appendix.tex is used to control the number of appendices through the counter "noa." Set the value to 1 if there is a single appendix any larger value will work for multiple appendices. If the value is 1 input only appendixA, if the value is two or more allow the other appendices to be input (adding input statements if needed). The beginning

Table 2-1. Feasible triples for highly variable Grid, MLMMH.

Time (s)	Triple chosen	Other feasible triples
0	(1, 11, 13725)	(1, 12, 10980), (1, 13, 8235), (2, 2, 0), (3, 1, 0)
2745	(1, 12, 10980)	(1, 13, 8235), (2, 2, 0), (2, 3, 0), (3, 1, 0)
5490	(1, 12, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
8235	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
10980	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
13725	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
16470	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
19215	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
21960	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
24705	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
27450	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
30195	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
32940	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
35685	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
38430	(1, 13, 10980)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
41175	(1, 12, 13725)	(1, 13, 10980), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
43920	(1, 13, 10980)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
46665	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
49410	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
52155	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
54900	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
57645	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
60390	(1, 12, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
63135	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
65880	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
68625	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
71370	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
74115	(1, 12, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
76860	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
79605	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
82350	(1, 12, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
85095	(1, 12, 13725)	(1, 13, 10980), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
87840	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
90585	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
93330	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
96075	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
98820	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
101565	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
104310	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
107055	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
109800	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
112545	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)

Table 2-1. Continued

Time (s)	Triple chosen	Other feasible triples
115290	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
118035	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
120780	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
123525	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
126270	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
129015	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
131760	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
134505	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
137250	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
139995	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
142740	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
145485	(1, 12, 16470)	(1, 13, 13725), (2, 2, 2745), (2, 3, 0), (3, 1, 0)
148230	(2, 2, 2745)	(2, 3, 0), (3, 1, 0)
150975	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
153720	(1, 12, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
156465	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
159210	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
161955	(1, 13, 16470)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)
164700	(1, 13, 13725)	(2, 2, 2745), (2, 3, 0), (3, 1, 0)

of appendixA has another ifthenelse statement regarding the number of appendices and then demonstrates how a landscaped page is inserted as the first page of an appendix. [14]

2.1.6 Decimal Alignment

One of the strangest facts about LaTeX is that it doesn't have a simple method of aligning numbers in a table on the decimal point. The workaround is to create two separate columns align the first to the right and the second to the left and set the separator to a decimal point. This will give the illusion of a decimally aligned column. [8]

Table 2-2. How to align decimals in a numerical column

Category	Result
first	3.14159
second	16.2
third	123.456

As you can see in Table 2-2 the result is an illusion of a decimally aligned column exactly as preferred by the Editorial Office. If done carefully nobody will ever know your dirty secret!

2.2 Images That Do Not Show

If you're trying to use the package psfrag you must change the method of compilation for it to render your images correctly. In the main file, change dvipdfm to dvips. Make this same change in the packages file with both the graphix and hyperref package options. You then must compile in the folowing manner:

1. latex filename
2. latex filename
3. bibtex filename (if needed)
4. latex filename
5. latex filename (these latex commands are only needed if the bibtex command was used)
6. dvips filename
7. ps2pdf filename

This can still have a negative effect on the hyperref package and result in broken links and/or incorrect margins in the TOC, LOT and LOF.

CHAPTER 3
REFERENCES

3.1 Manually

You can do your citations and references in LaTeX manually. However, if you do, you lose one of the top reasons for using LaTeX in the first place. Create a bibliography.tex file using the ufsampleETD.bbl file as a guide and include it using the `\include{bibliography}` command. NOTE: This method will not allow you to adjust the format of your bibliography by changing the .bst file called in the main document.

3.2 BibTeX

We use the package natbib in the template. [11] It is used by a large number of journals and offers the widest variety of citation and reference listing options with the least amount of overhead and/or complexity. Determine whether you prefer numbered or un-numbered reference listing. Go to the packages.tex file and make sure your preference is uncommented. Comment (or delete) the other option and you're ready to select a bibliography style.[1] This is done with the `\bibliographystyle{bibstyle}` command. We have included several basic reference styles(the .bst file types). The differences are noted in the following list. Note that the plain style does not change in the either the numbered or author-yearnumbers. [3]

- Numbered References (using the numbered, sort and compress natbib option)
  - plain: numbered citations-brackets, numbered reference list in alphabetical order, full first name last name.
  - ufinit: numbered citations-brackets, numbered reference list in citation order, initials and last name.
  - plainnat: numbered citations-brackets, numbered reference list in alphabetical order, full first name last name.
  - abbrvnat: numbered citations-brackets, numbered reference list in alphabetical order, initials and last name.
  - unsrtnat: numbered citations-brackets, numbered reference list in citation order, full first name last name.

  - chicagoReedWeb: numbered citations-brackets, numbered reference list in alphabetical order, last name, full first name.
  - apa-good: numbered citations-brackets, numbered reference list in alphabetical order, last name, first and middle initials.

- Un-numbered References (using the authoryear natbib option)
  - plain: numbered citations-brackets, numbered reference list in alphabetical order, full first name last name.
  - ufinit: numbered citations-parentheses, numbered reference list in citation order, initials and last name.
  - plainnat: author-year citations-brackets, un-numbered reference list in alphabetical order, full first name, last name.
  - abbrvnat: author-year citations-brackets, un-numbered reference list in alphabetical order, initial first name, last name.
  - unsrtnat: author-year citations-brackets, un-numbered reference list in citation order, initial first name, last name.
  - chicagoReedWeb: author-year citations-parentheses, un-numbered reference list in alphabetical order, last name, first name, line replaces repeated author.
  - apa-good: author-year citations-parentheses, un-numbered reference list in alphabetical order, last name, first initials.

We have included some additional .bst files that may or may not be useful. Please note! We use several citation commands to illustrate the different results and the bibliography in this document cannot be used as an example of any specific reference system. It is YOUR responsibility to determine the reference style of the journal you want to emulate and obtain the necessary .bst files to emulate that style.

3.3 Footnotes

To create footnotes you simply use the `\footnote{text}` command.[1] This works well except when it is used inside any environment that produces (or can produce) a

[1] See, it really does work

box.[2] In that case the manual suggests you use `\footnotemark` inside the environment and just outside the environment use `\foottext{text}` to define the text to accompany the footnote mark. [6] Frankly, I have had some trouble making this option work, particularly in the longtable package and recommend that you avoid footnotes inside any "environment" other than the normal text mode if at all possible.[3] I'm repeating our example of the decimal alignment work-around to illustrate the footnotemark - footnotetext example.

Table 3-1. You shouldn't have two tables (or figures) with the same caption. Even if there's a small difference well into the caption it is best to put the difference first to uniquely identify the table or figure.

Category	Result
first	3.14159[4]
second	16.2
third	123.456

[2] I've seen examples of documents that failed to compile simply because of the placement of a footnote command.

[3] Footnotes should be single-spaced with a space between each note. They should also start re-numbering at one each chapter!

[4] This is an example of the footnotemark process for box creating environments - I sure hope this works!

CHAPTER 4
QUANTUM CHEMISTRY

I've borrowed a few examples of equations I've run across to illustrate how the equations should look in the thesis or dissertation. [2]

4.1 The Electronic Problem

The starting point of any discussion of quantum mechanics is the non-relativistic, time-dependent Shrödinger equation [**?**],

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \hat{H}\Psi(r,t) \ , \tag{4–1}$$

where $\hbar$ is Planck's constant, $\Psi(r,t)$ is the wave function of the quantum system and $\hat{H}$ is the Hamiltonian, which is an operator that contains a kinetic energy term and a potential energy term.

When we restrict the wave function to be a product of a function of time and a function of space, as, for example, is the case when the potential energy term of the Hamiltonian is independent of time, the time-independent Scrhödinger equation can be expressed as:

$$\hat{H}\Psi = E\Psi \ . \tag{4–2}$$

The specific form of the Hamiltonian for molecules is:

$$\begin{aligned}\hat{H} = -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \\ + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \ ,\end{aligned} \tag{4–3}$$

where A and B, etc., label nuclei, and i, j, etc., label electrons, Z is the atomic number, and Hartree atomic units ($\hbar = e = m_e = 1$) have been used.

The Born-Oppenheimer approximation, which is a useful and central approximation in quantum chemistry, separates electronic and nuclear motions. Assuming that the nuclei are fixed (since nuclei are much heavier than the electrons), the nuclear kinetic energy term, which is the first term in equation (4–3), can be neglected, and the repulsion between nuclei, the third term of equation (4–3), is a constant. This approximation leads to an electronic Hamiltonian,

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \ , \tag{4–4}$$

and the Schrödinger equation becomes:

$$\hat{H}_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \ . \tag{4–5}$$

4.2 Hartree-Fock Approximation

Except for the simple case of H_2^+, molecules are many-electron problems and determining accurate molecular orbitals, which are the eigenfunctions of the Schrödinger equation for a molecule, has been the main task of quantum chemists for many years. Approximate methods have been developed to solve the Schrödinger equation, since it is intractable computationally to find the exact solution for a many-electron system. One approximate method that is used frequently to solve the Schrödinger equation is based on Hartree-Fock theory. In the present work, Hartree-Fock is not the main method used, but it is crucial to introduce it to explain the methods on which this work is based.

Consider a trial function in the form of a single N-electron Slater-determinant, which obeys the Pauli exclusion principle,

$$\mid \Psi\rangle = \hat{O}\hat{A} \mid \phi_1\phi_2...\phi_\alpha...\phi_N\rangle \ . \tag{4–6}$$

Here $\hat{O}$ is the spin projector operator that ensures that the wave function remains an eigenfunction of the spin-squared operator ($\hat{S}^2$), $\hat{A}$ is the antisymmetrizer, ϕ_α is a one-electron wave function that represents the molecular orbital, and Dirac notation has been adopted.

The molecular orbitals can be expanded as a linear combination of atomic orbitals ψ_α ,

$$\phi_i = \sum_u \chi_u C_{ui} = XC_i \ , \tag{4–7}$$

which constitute the basis set for the calculation.

Ψ is varied with respect to C following the variational principle to minimize the expectation value of the electronic Hamiltonian, $\hat{H}$ (where the electronic subscript has been dropped for simplicity) to give the following expression for the effective one-particle Fock operator, f,

$$f \mid \phi_i\rangle = [h + \sum_{j=1}^{N} J_j - K_j] \mid \phi_i\rangle = \sum_{j=1}^{N} \epsilon_{ji} \mid \phi_j\rangle \ , \tag{4–8}$$

where h represents the first two terms of equation (4–4) and J and K are the coulomb and exchange operators respectively. Using a unitary basis to diagonalize the Hermitian matrix, ϵ, with matrix elements ϵ_{ji} yields the canonical Hartree-Fock equation,

$$f\phi_i = \epsilon_i \phi_i \ . \tag{4–9}$$

From this equation the following generalized-eigenvalue expression can be obtained,

$$FC = SCE \ , \tag{4–10}$$

where F is the Fock matrix, C is a square matrix containing the molecular orbital coefficients, S is the overlap matrix and E is the energy matrix containing the orbital energies ϵ_i.

The Fock matrix elements are,

$$\begin{aligned} F_{uv} &= \langle \chi_u \mid f \mid \chi_v \rangle = \langle u \mid f \mid v \rangle \\ &= H_{uv} + \sum_{s,t} P_{st} [\langle us \mid vt \rangle - \frac{\langle us \mid tv \rangle}{2}] \ , \end{aligned} \tag{4–11}$$

P is the density matrix,

$$P_{uv} = \sum_a C_{ua} C_{va} n_a \ , \tag{4–12}$$

where n_a is the occupation number.

The overlap matrix elements are,

$$S_{uv} = \langle \chi_u \mid \chi_v \rangle = \langle u \mid v \rangle \ . \tag{4–13}$$

A self-consistent solution of the previous equations is known as the Hartree-Fock, Self-Consistent Field (SCF) methodology.

These *ab initio* expressions for the Fock matrix will be used below to explain additional details of the methods used in this research.

CHAPTER 5
MATH, FIGURES, AND TABLES

5.1 Text Flow: Problems and Solutions

Let's face it. LaTeX is used because you can type very complicated formulas and equations without ever touching a mouse! And in most situations, where the spacing requirements are not as demanding as the UF Theses and Dissertation requirements, LaTeX's love of white space would not pose a problem.

However, you ARE producing this document for the UF Graduate School and their spacing requirements ARE quite demanding. Pages, other than the last page of a chapter, are supposed to be full. LaTeX has a tendency to break a page early rather than split a display element (equation array, align, theorem, table, figure, etc.). To help in this matter there are three lines in the preamble of the ufsampleETD.tex file. These are:

```
\renewcommand{\topfraction}{0.85}
\renewcommand{\textfraction}{0.1}
\renewcommand{\floatpagefraction}{0.75}
```

However, even with these commands LaTeX likes to break pages rather than equation arrays, theorems, postulates, proofs, etc. If you find that these elements are causing pages to break badly you can un-comment the command `\allowdisplaybreaks` and with any luck, that will cure the problem. If not, it may be necessary to break the displays manually which is always the last resort and only done just before final submission.

5.2 Equation Notes

Again, one of the major reasons for using LaTeX in the first place is how it handles the mathematics displays. However, there are a few details you need to be aware of:

- If you place an extra carriage return before or after an equation LaTeX will place some extra "paragraph" spacing around the equation. Try to avoid this.
- If you want to explain your notation in the equation with a list of statements such as; "where $Z = \gamma$," do so in paragraph form rather than a vertical list.

- While every equation does not have to be labeled, determining which items should and should not be labeled can get tricky. Since LaTeX generates the labels automatically it's usually best to let it label as it sees fit. In equation arrays that leads to a final result it's certainly acceptable to suppress the labels on preliminary steps by using the `\nonumber` command as long as the last line of the procedure has a label.

5.3 Tables, Figures, and Subfigures

5.3.1 Tables

Typically, the standard LaTeX table environment is used. Table captions need to be above the table, and typically there should not be any vertical lines in the table. Formatting tables in a landscape page is explained in section 5.4.

Table 5-1. A sample Table

First	Second	Third
12	45	26
17	32	93
text	51	can be there too.
	28	Figures too - a cat.
	000	and a mouse!

5.3.2 Figures

You can either use `\includegraphics` or \epsfig commands to include your figures. The appropriate packages are included in the packages.tex file. For additional precaution, a copy of `rotating.sty` is included in the template. Please do not delete this file. Note that the caption for the figures is below the figure. The figures in the table above was inserted with \epsfig. Below is a sample file with `\includegraphics`:

5.3.3 Subfigures

In addition to the standard LaTeXoptions for scaling and rotation, the `rotating` package has additional options for turning and rotating both text and figures/tables. Please look at the documentation of this package for further details.

Figure 5-1. LaTeX2ϵ logo, resized for no reason. This caption is being extended in order to test that it has the correct indentation.

For subfigures, please use `subfigure` command (see chapter4.tex for code). We have made some slight modifications to the subfigure.sty file to match the Editorial Office specification so make sure it is in the same folder as the ufsampleETD.tex file when you compile your document.

A

B

D

C

Figure 5-2. Tom and Jerries? A) Mouse 1 B) mouse 2 C) Hungry Cat D) mouse 3

There is some fancy formatting possible with `subfigure`. For instance, it is possible (but not suggested) to list the captions of each subfigure in the List of Figures in the table of contents. Please look at the documentation of `subfigure` package for details.

5.4 Formatting in Landscape Mode

There are many ways to format figures and tables in a landscape. Depending on what you want to use, you can use one of the following enviroments:

- The `landscape` environment
- The `sidewaysfigure` environment
- The `sidewaystable` environment

5.4.1 The `landscape` environment

The `landscape` environment starts by default a new page, because it changes the two lengths \paperwidth and \paperheight. Typically you will use the `landscape` if you want to have an **entire** section or a subsection in landscape mode (i.e. both text and figures/tables). This should not be used for a just single figure or a table (Use `sidewaysfigure` and `sidewaystable` described in sub-section 5.4.2 instead). In the chapter4.tex file we demonstrate how to force a page break in the Table of Contents. Something that often needs to be done but is not documented in most LaTeX tutorials.

The following sub-subsection is included in an environment like:

```
\begin{landscape}
\subsubsection{Sample Landscape Page}\label{ps}
Note that even though we are in landscape mode,
only the text part is in landscape.
The header and footer (for example the page number) are still
in portrait mode. This $\backslash$begin$\{$landscape$\}$
environment is part of the lscape package. Look at the documentation
of this package for further details and options.
\end{landscape}
```

The \begin{landscape} immediately starts the new page, a lot of vertical whitespace, like the one on this page, maybe possible.

The subsubsection command also represents a common error in dissertation/thesis construction. There is only one heading at that level. Whenever a section is divided, it must be divided into two segments - otherwise there is no reason to introduce another category.

5.4.1.1 Sample Landscape Page

It's Important to remember that no headings or paragraph text should actually be on a landscaped page. Only the Figure or Table and their captions. In addition, the landscaped page should be rotated in the pdf so the text on the landscaped page is readable without turning your head sideways. Note that even though we are in landscape mode, only the text part is in landscape, the header and footer (that implement, for example the page number) are still in portrait mode. This `landscape` environment is part of the `lscape` package. Look at the documentation of this package for more details and options.

5.4.2 `sidewaysfigure` and `sidewaystable` Environments

With large figures and tables with lots of columns, it is sometimes necessary to rotate them to landscape mode. The `sidewaysfigure` and the `sidewaystable` environments from the `rotating` package can be used for this. Sample code for these two environments are give below:

A Landscape Figure:

```
\begin{sidewaysfigure}
\centerline{\epsfig{figure=images/figurename.eps, scale=0.5}}
\caption{Your Caption for the figue}
\end{sidewaysfigure}

Note: Change value of scale to change your figure size.
      1 is 100%, i.e. original size. You can go beyond 1
      if your file is in a scalable vector graphics format
      like .eps
```

A Landscape Table:

```
\begin{sidewaystable}
\centering %optional
\begin{tabular}{rl}
\end{tabular}
\caption{Your Caption for the table}
\end{sidewaystable}
```

The above code will produce a figure/table rotated by 90° in the counter-clockwise direction, and will also rotate the caption accordingly as per the Editorial Office requirements. Both these environments are "intelligent" in the sense that they will put your figure/table in a new landscape page and will **NOT** leave empty whitespace before the figure like the `landscape` environment. The following 'Lorem Ipsum' paragraphs demonstrate this.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Integer ante. Ut tincidunt ultrices turpis. Phasellus nonummy pulvinar sem. Donec sem nisl, rhoncus eu, porttitor in, blandit nec, arcu. Vestibulum tincidunt ante. Pellentesque quis massa. Proin vehicula feugiat turpis. Aenean at tellus sed justo ornare dictum. Nullam sit amet libero nec lorem sodales cursus. Donec tortor nulla, convallis in, suscipit in, posuere at, nunc.

Figure 5-3. You can turn the world on its heels.

Aliquam tortor risus, ultricies sed, eleifend in, congue quis, justo. Pellentesque egestas orci non urna. Phasellus ligula. Ut nonummy. Suspendisse potenti. Donec posuere justo quis eros. In erat. Nunc aliquam metus sed dui. Fusce justo felis, posuere a, elementum non, semper eget, mi. Morbi iaculis lorem at sem. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Phasellus velit. Maecenas libero tortor, pharetra id, dictum ac, lacinia vestibulum, urna. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In libero nunc, fringilla a, condimentum lobortis, consequat eget, quam. Phasellus eget nisi. Maecenas risus ligula, euismod a, tristique non, sagittis eu, quam. Donec metus nunc, varius ut, lacinia sit amet, pellentesque ac, mauris. Nulla mollis aliquam metus.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Integer ante. Ut tincidunt ultrices turpis. Phasellus nonummy pulvinar sem. Donec sem nisl, rhoncus eu, porttitor in, blandit nec, arcu. Vestibulum tincidunt ante. Pellentesque quis massa. Proin vehicula feugiat turpis. Aenean at tellus sed justo ornare dictum. Nullam sit amet libero nec lorem sodales cursus. Donec tortor nulla, convallis in, suscipit in, posuere at, nunc. Aliquam tortor risus, ultricies sed, eleifend in, congue quis, justo. Pellentesque egestas orci non urna. Phasellus ligula. Ut nonummy. Suspendisse potenti. Donec posuere justo quis eros. In erat. Nunc aliquam metus sed dui. Fusce justo felis, posuere a, elementum non, semper eget, mi. Morbi iaculis lorem at sem. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Phasellus velit. Maecenas libero tortor, pharetra id, dictum ac, lacinia vestibulum, urna. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In libero nunc, fringilla a, condimentum lobortis, consequat eget, quam. Phasellus eget nisi. Maecenas risus ligula, euismod a, tristique non, sagittis eu, quam. Donec metus nunc, varius ut, lacinia sit amet, pellentesque ac, mauris. Nulla mollis aliquam metus.

Table 5-2. The Same Table as 5-1, but in landscape mode

First	Second
12	26
17	93
text	can be there too.
	Figures too - a cat.
	and a mouse!

5.5 Some bugs and fixes

A quirk in the LaTeX2ϵ template is the centering of table and figure captions ... which the editorial office will not accept. This is actually only a problem for captions that are less than the width of the paper(within the margins that is). A fix has already been implemented in the template for this issue.

However, if you find your short captions are being centered in spite of the new caption package options, try using the following codes, which each differ ever so slightly depending if the caption is for a table or figure. Inserting the given code in the table or figure environments just after you declare the start of that environment for each table or \landscape environment figure that has a short caption that is being centered:

For Tables:

```
  \makeatletter
\long\def\@makecaption#1#2{%
  \vskip\abovecaptionskip
  \sbox\@tempboxa{#1: #2}%
  \ifdim \wd\@tempboxa >\hsize
    #1: #2\par
  \else
    \global \@minipagefalse
    \hb@xt@\hsize{\box\@tempboxa\hfil}%
  \fi}
\makeatother
```

For Figures:

```
\makeatletter
\long\def\@makecaption#1#2{%
  \vskip\abovecaptionskip
  \sbox\@tempboxa{#1: #2}%
  \ifdim \wd\@tempboxa >\hsize
    #1: #2\par
  \else
    \global \@minipagefalse
    \hb@xt@\hsize{\box\@tempboxa\hfil}%
  \fi
  \vskip\belowcaptionskip}
\makeatother
```

CHAPTER 6
GENERAL THESIS TIPS

- BACK UP YOUR THESIS. Often you will not realize for days or weeks that important paragraph or page is missing. Make recovery as easy as possible by keeping a dated backup of each writing session. Then copy those backups to at least two locations other than your hard drive: your home server, gmail account, thumb drive, the options are wide and numerous. There is no excuse for not backing up the most important document of your academic career.

- Start your bibliographic database the day you start reading. Keep it up to date and annotate it, so you know where it came from (library, Internet, public library, professor), whether you've read it, and where you want to cite it. This will make the writing process less frustrating and creating the bibliography seamless.

- Think of thesis formatting as a form of productive procrastination. Please don't put it off until the last week.

- BACK UP. No, seriously. It's not "if" your hard drive fails, it's "when." Not to scare you or anything, but it's a good habit, like buckling your seat belt or not leaving your laptop unattended. You really don't want to wish you had taken that small precaution.

- Keep the editable original of each graphic you want to include in your thesis in one folder. Later you may need to change a graphic quickly and having the editable original makes it easy. For graphs, keep the original Excel/JMP/Stata document, not a PDF. For photographs, keep a high resolution copy. For drawings and illustrations, keep the original document.

- Use the timesaving benefits of LaTeX from the first day. Cross references can refer to tables, graphics, and chapters so you do not have to update references as your thesis changes. Use comments to make notes about what needs to added or changed.

- Enjoy the experience! And get some sleep, food and relaxation on occasion. Thousands of people did this before you; you can do this too.

CHAPTER 7
CONCLUSION

Acquiring information for a risk averse manager is never a trivial task. One might think the imposed risk from motivating another task would naturally force the manager to collect more information. But the contract design is subtle. If the imposed risk on the manager is too small, he will not acquire information, but just invest in the project; if the imposed risk is too high, the manager will not do the work either, but just forgo the project.[1] This paper shows that auditing, when conducted properly, can help create efficient incentives for the manager to acquire information.

An audit of the manager's acquired information is beneficial because it aligns the manager's incentives in acquiring useful information and in making a proper investment decision. But if the board of directors cannot commit not to use the disclosed information to renegotiate the initial contract, an extensive audit may exacerbate the control problem. Therefore, if the audit technology is not highly effective in identifying misstatements, the auditor may only want to verify whether the manager's report is consistent with his investment decision, but allow the manager to keep the finer details private. This arrangement is beneficial because it also reins in the self-interested behavior coming from the owner's side. Together with the manager's behavior, it depicts an interesting balance in equilibrium.

Financial reports contain information that is useful for future decision-making. The consequences of past decisions are also recorded in financial reports. The auditor verifies the reported information thus serves as a monitoring device of managers' decisions. It is important for the auditor to understand the manager's dynamic decisions making. The manager's decision in financial reporting is correlated with his decision in productive

[1] Laux [2004] pointed out that motivating the manager to implement another task cannot automatically provide the manager the right incentive to collect information. It is always necessary to motivate information acquisition explicitly.

actions, which in turn affects the underlying resource allocation. Ex post uses of reported information influence the manager's ex ante incentives to acquire information and make a proper investment decision. It is also important to recognize the economic consequences of standard setting. The rules on financial reporting and auditing change the preparers' behavior. Standard-setters are not regulating nature but rational economic agents.

The results echo the line of literature that provides explanations for earnings management based on the effect of renegotiation (a violation of the Revelation Principle's assumptions).[2] My analysis focuses on how renegotiation affects the manager's investment behavior when the audited information is endogenously acquired. Audit technology determines how much information is to be disclosed. In contrast, a perfect audit in my model results in a first-best scenario and the manager's information should be disclosed. The intuition is that the manager's shirking means no useful information is produced and thus will be detected in the audit. The adverse effect of renegotiation is amplified only when the audit technology is relatively ineffective so using the auditor's report alone is not informative enough to provide incentives for the agent to work. My analysis emphasizes the subtlety of the audit function and therefore the solution is "interior"—motivating accurate financial reports may or may not be efficient. Auditors' judgments are the centerpiece.

Some argue that we impose too much responsibility on the auditor. The auditor's job is to "check whether a reported number is correct". The auditor does not ask why and how the number is generated. As the auditor assesses audit risk and materiality before performing substantive tests of transactions, however, she is concerned about management. SAB 99 advises the auditor to investigate the manager's incentives carefully as opposed to setting some mechanical materiality threshold. Accounting firms hire

[2] See Demski and Frimor [1999], Christensen, Demski and Frimor [2002], Christensen, Feltham and Sabac [2004] and Gigler and Hemmer [2004].

experts to audit R&D contracts because they have superior knowledge to evaluate the manager's performance. Auditing is not a simple task in that it requires the auditor formulate judgments. The deeper the auditor understands managerial decisions, the easier the auditor reaches a correct conclusion.

Some worry about the auditor's incentives if she is provided more discretion. That is a valid concern. There is another round of incentive problems. One problem is how the PCAOB evaluates the auditor's work when their opinion is at odds with the auditor's judgments. Auditor's exposure to legal liability forces standard setters to consider simple, rules-based standards that permits less discretion. The audit fees, market competition, etc., all influence the auditor's behavior. A model with multiple players would be more appropriate to address these questions. But our model provides a salient structure of the audit function. More importantly, we point out the gap in the understanding of financial reporting and auditing. If financial reporting is a sophisticated communication process, auditing should help to serve this goal. We first provide a benchmark where an ideal auditor should perform, then we search for feasible mechanisms to induce the auditor to perform as we hope. After all, the questions boil down to the trade off between revealing information enforced by an auditor and the resulting concerns for efficiency. The main message from my study is well reflected here: the optimal auditor's choice depends on the context that creates the incentive nexus and there is no panacea for all the reporting issues. Auditors are expected to rely on judgments to deliver high-quality work.

Besides revelation mechanisms, there are other mechanisms that may make the discovered information useful. Contract renegotiation takes into consideration newly discovered information. Thus the manager's investment incentives are better aligned with the current situation. In this way, the spillover effect between information acquisition and investment decisions can be isolated. Efficiency is strictly improved. However, the result hinges on the assumption that there is no additional cost to invest in the project. An extension would drop this assumption and examine the spillover effect between the two

tasks. We might find that interim contract renegotiation can disrupt the synergy between the two tasks.

APPENDIX: THIS IS THE FIRST APPENDIX

And the appendix text goes here. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas eget magna. Aenean et lorem. Ut dignissim neque at nisi. In hac habitasse platea dictumst. In porta ornare eros. Nunc eu ante. In non est vehicula tellus cursus suscipit. Proin sed libero. Sed risus enim, eleifend in, pellentesque ac, nonummy quis, nulla. Phasellus imperdiet libero nec massa. Ut sapien libero, adipiscing eu, volutpat porttitor, ultricies eget, nisi. Sed odio. Suspendisse potenti. Duis dolor augue, viverra id, porta in, dignissim id, nisl. Vivamus blandit cursus eros. Maecenas sit amet urna sit amet orci nonummy pharetra.

Praesent cursus nibh et mauris. In aliquam felis sit amet ligula. Nulla faucibus nisl eget nisl. Aliquam tincidunt. Mauris eget elit sed massa luctus posuere. Pellentesque suscipit. In odio urna, semper ut, convallis ut, porta et, nibh. Nulla sodales metus nec velit posuere gravida. Cras tristique. Etiam urna risus, accumsan ut, placerat sed, iaculis id, est.

Nullam mi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Duis vitae metus in massa hendrerit rhoncus. Fusce tortor justo, laoreet eu, facilisis at, gravida et, felis. Donec imperdiet mollis erat. Integer tempus nulla ac lorem. Fusce porttitor. Aenean quis arcu. Morbi consectetuer, leo eu mollis elementum, urna massa malesuada risus, euismod tempor lorem elit ut mauris. Cras elit orci, facilisis ac, mattis iaculis, cursus ac, augue. Donec eget nisl. Pellentesque fermentum sodales nibh. Vivamus non risus. Donec est libero, tincidunt sit amet, pretium vitae, blandit sed, tellus. Nunc diam risus, interdum sed, laoreet quis, varius ac, turpis. In et purus eget nibh vehicula rhoncus. Aenean et neque. Praesent nisl nisi, tempus quis, nonummy ac, auctor a, neque. Suspendisse et metus. Suspendisse non metus eu mauris auctor sagittis.

REFERENCES

[1] Adams, W. W., Eby, R. K., and McLemore, D. E. *The Materials Science and Engineering of Rigid-Rod Polymers*. Pittsburgh, PA: Materials Research Society: Pittsburgh, 1989.

[2] Agarwal, S., Johns, L. E., and Narayanan, R. "Growth curves in precipitation and solidification.", 2006. Journal of Crystal Growth (submitted).

[3] Andrews, Donald W.K. "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space." *Econometrica* 68 (2000): 399–405.

[4] Antonio, D. and Chaouche, M. "Sedimentation of a Sphere in a Suspension of Neutrally Buoyant Fibers." *J. Rheol.* 46 (2002): 749–759.

[5] Bailey, D. H. and Swarztrauber, P. N. "The fractional Fourier transform and applications." *SIAM Rev.* 33 (1991).3: 389–404.

[6] Boek, E. S., Coveney, P. V., Lekkerkerker, H. N. W., and Van Der Schoot, P. "Simulating the Rheology of Dense Colloidal Suspensions Using Dissipative Particle Dynamics." *Phys. Rev. E* 55 (1997): 3124–3133.

[7] Brady, J. F. "Computer Simulation of Viscous Suspensions." *Chem. Eng. Sci.* 56 (2001): 2921–2926.

[8] Chwang, A. T. and Wu, T. Y. "Hydromechanics of Low-Reynolds-Number Flow .1. Rotation of Axisymmetric Prolate Bodies." *J. Fluid Mech.* 63 (1974): 607–622.

[9] Datta, Somnath. "On a modified bootstrap for certain asymptotically nonnormal statistics." *Statistics & Probability Letters* 24 (1995): 91–98.

[10] DiCiccio, Thomas J. and Romano, Joseph P. "On adjustments based on the signed root of the empirical likelihood ratio statistic." *Biometrika* 76 (1989).3: 447–456.

[11] Diller, Antoni. *LaTeX Line by Line*. New York, New York: John Wiley & Sons, 1993.

[12] Hall, Alastair R. *Generalized Method of Moments*. New York: Oxford University Press, 2005.

[13] Sato, T. and Teramoto, A. "Dynamics of Stiff-Chain Polymers in Isotropic Solution: Zero-Shear Viscosity of Rodlike Polymers." *Macromolecules* 24 (1991): 193–196.

[14] van Bruggen, M. P. B., Lekkerkerker, H. N. W., Maret, G., and Dhont, J. K. G. "Long-Time Translational Self-Diffusion in Isotropic and Nematic Dispersions of Colloidal Rods." *Phys. Rev. E* 58 (1998): 7668–7677.

BIOGRAPHICAL SKETCH

This section is where your biographical sketch is typed in the bio.tex file. It should be in third person, past tense. Do not put personal details such as your birthday in the file. Again, to make a full paragraph you must write at least three sentences.