

UNIVERSITY of CALIFORNIA
SANTA CRUZ

**RECONSTRUCTION OF NON-PROMPT THREE HIT TRACKS
WITH THE PROPOSED SID AT THE INTERNATIONAL LINEAR
COLLIDER**

A thesis submitted in partial satisfaction of the
requirements for the degree of

BACHELOR OF SCIENCE

in

PHYSICS

by

Christopher J. Meyer

13 June 2008

The thesis of Christopher J. Meyer is approved by:

Professor Bruce A. Schumm
Technical Advisor

David P. Belanger
Thesis Advisor

Professor David P. Belanger
Chair, Department of Physics

Copyright © by
Christopher J. Meyer
2008

Abstract

Reconstruction of Non-Prompt Three Hit Tracks With the Proposed SiD at the
International Linear Collider

by

Christopher J. Meyer

An algorithm for the reconstruction of non-prompt tracks in the outer three layers of the tracking barrel chamber is presented. It is shown to improve the efficiency by a factor of 1.5 and increase purity by a factor of 100 for 3-hit track finding. The next step is to optimize the code to search for specific signatures. The channel $slepton \rightarrow lepton + gravitino$ is presented as an example along with current work being undertaken to find this specific signature.

Contents

List of Figures	vi
List of Tables	vii
Dedication	viii
Acknowledgements	ix
1 Introduction	1
2 Theory	3
2.1 The Standard Model	3
2.1.1 The Four Forces	3
2.1.2 The Fundamental Particles	5
2.2 The Higgs Doublet	8
2.3 Supersymmetry	9
3 International Linear Collider	12
3.1 The Detector	13
3.1.1 Coordinate System	13
3.1.2 Tracking Chamber	14
3.1.3 Electromagnetic Calorimeter	16
3.1.4 Hadron Calorimeter	16
3.1.5 Muon Chamber	17
4 Terminology	18
5 Software	24
5.1 org.lcsim Class Library	24
5.2 NetBeans IDE 6.0	25
5.3 Java Analysis Studio 3	25
5.4 Monte Carlo Simulated Data Set	25

6	The Code	27
6.1	Previous Work	27
6.1.1	VXDCh eater	28
6.1.2	AxialBarrelTrackFinderZ	28
6.1.3	GarfieldTrackFinder	29
6.2	New Work	29
6.2.1	MultiDriver	30
6.2.2	The Problem	31
6.2.3	The Solution	31
7	Results	36
8	Conclusion	38
	Bibliography	41

List of Figures

2.1	The three generations of quarks with their associated leptons and neutrinos.	7
2.2	Interaction Strength of Electromagnetism, Weak, and Strong Force.	10
3.1	SSD Detector Geometry for proposed ILC	14
3.2	Cylindrical Coordinates	15
6.1	SeedExtend Matching Distributions	34
8.1	SUSY Decay	39

List of Tables

6.1	Track Finding Results for MultiDriver	30
7.1	Track Finding Results for SeedExtend	36

Dedicated to my family, whose help and motivation have gotten me where I am
today.

Acknowledgements

First and foremost I would like to thank Professor Bruce Schumm, without whose guidance and mentorship I would not have had this opportunity. Great thanks also goes to my colleague Tyler Rice who was co-author of the SeedExtend code. Credit is also due to the ILC Collaboration, in particular Dr. Norman Graf, Dr. Tim Nelson, and colleague Lori Stevens. The base your code provided was indispensable to my efforts. Thank you all for this great adventure.

1 Introduction

This thesis is on the development of pattern recognition algorithms for the International Linear Collider (ILC). Although the ILC is still only proposed there is a great deal of work being done on the design of the facility itself, the technology that will accelerate the electron/positron beams it will be colliding, and especially on the geometry of the detectors.

Work on development of pattern recognition algorithms before the design phase of the detector is complete is a crucial effort for benchmarking detector geometry. With an idea of how precise and what level of information for collisions is expected from the detector, running on simulations will provide valuable information on whether or not the detector is on track to meet experimental requirements.

This is accomplished by the Monte Carlo simulation of collisions. By generating data with known (or theorized) results, the output from different detector geometries can be created. For example, the data accessible from placing five layers of detecting devices or six layers of detecting devices can be run over with pattern recognition algorithms. If it is found that no more information is gathered with the addition of the sixth layer the final design will only include five layers. However, if there is a large difference in the quality of the results, a sixth layer can be added to improve track finding performance.

A brief introduction to the theory behind the physics of the collision products will be given, followed by an overview of the ILC and the currently proposed detector geometry. This will be followed by a description of the software and code used for benchmarking, and conclude with a presentation of our work on finding particles that only leave data on 3 layers of the central silicon tracker.

2 Theory

2.1 The Standard Model

The Standard Model is a description of what we currently think to be the most elementary group of particles from which all other matter is formed, and the forces that govern their interactions. The majority of the material presented in this section comes from *Deep Down Things* by Bruce Schumm[1].

2.1.1 The Four Forces

The four forces that mediate interaction between all matter are (in decreasing order of strength) the Strong Nuclear Force, the Electromagnetic Force, the Weak Nuclear Force, and Gravity.

Strong Nuclear Force

The strong nuclear force, or strong force, is responsible for holding protons and neutrons together in the nucleus. It is also the force by which quarks (described in following sections) are held together to form protons and neutrons. Unlike the general inverse square rule of force which asserts that the magnitude decreases as distance increases, the strong force increases its attraction

as distance increases. For a particle to interact via the strong force it must contain some amount of color charge, which comes in three types: red, green, and blue.

Electromagnetic Force

The electromagnetic force is the union of the electric and magnetic fields which are commonly observed. It is responsible for electricity and current flow and why magnets are attracted and repulsed by one other. The charge of electromagnetism is the every day charge we consider when quantifying current flow in a wire. It comes in discrete amounts, of which the electron possesses a single unit (often called e).

Weak Nuclear Force

The weak nuclear force, or weak force, is the least accessible to every day experience, although perhaps the most interesting. The weak force is responsible for beta decay of atoms, a process where a neutron turns into a proton with the emission of several other decay products. For this reason it is often referred to as the alchemist of the forces. It also has the strange property of being a non-parity conserving force; configurations that are the same up to mirror-reflection are distinct where the weak force is concerned.

Gravity

Gravity is the most commonly observed force, and often mistaken for the strongest. It is the attraction between two massive bodies such as a bowling ball and the earth. While it seems like a strong force, it is necessary to keep in mind the universe is essentially neutral and so there are little long range effects from electromagnetism to compete with gravity. By holding one small magnet above another you are able to lift the lower magnet off the ground, completely overcoming

the force exerted on it by the sizeably larger earth. Gravity is different from the other forces because as if yet it has only been observed to be an attractive force. It is also the only force which does not have a quantum theory associated with it.

Unification of Electromagnetism and the Weak Force

In the past decade the electromagnetic and weak forces have been combined under a theory to form the electroweak force, a triumph of the Standard Model. There is also strong evidence that at high enough energies the strong force would unite as well to form one single set of rules governing all interactions (see figure 2.2 and section 2.3). In any case, it is these forces that determine the behavior of the fundamental particles of nature.

2.1.2 The Fundamental Particles

The Standard Model includes 2 different types of particles, bosons and fermions. Fermions consist of half integer spin particles, while bosons have integer spin. Included in the fermion family are quarks and leptons. Bosons consist of the force carrying particles and composite hadrons. Included in this list are also all of the anti-matter partners for the above mentioned particles. An anti-matter particle has exactly the same mass as its matter partner, with all of its quantum numbers multiplied by negative one. It also has the characteristic that when a particle and its anti-matter partner are brought into contact they annihilate each other, transforming all their mass into energy via Einstein's famous relation $E = mc^2$.

While anti-matter partners are usually denoted by a bar (i.e. \bar{u} means anti- u) or flipping the charge sign (i.e. μ^+ means anti- μ^-), an anti-matter electron is commonly referred to as a positron (due to the positive charge it carries). This nomenclature is continued throughout this

thesis.

Force Carrying Particles

Force carrying particles mediate the interactions between all matter. The exchange of force can be thought of as two people playing catch. Two people (particles) are standing on an icy lake. One person (particle) throws a ball (force carrier) at the other. The person (particle) who catches the ball (force carrier) slides backwards, while the person who threw the ball (particle that ejected the force carrier) recoils in the opposite direction. Force as we know it is a subatomic game of catch.

For each force there is a different particles which carries its interaction. The strong force is mediated by gluons (named because they act to glue quarks together into a nucleon). The weak nuclear force has the W^+ , W^- , and Z^0 bosons. Electromagnetism is carried by the familiar photon (a single particle of light). Finally, gravity is theorized to be carried by the graviton, although there is no experimental confirmation for this conjecture.

Quarks

Quarks make up most of our every day matter including protons, neutrons, and many other more exotic particles. They come in six flavors: up, down, strange, charmed, top, and bottom. It is a combination of three quarks the make up the proton and neutron (up up down and down down up respectively) that are held together by the strong nuclear force.

It is because they are subject to the strong nuclear force that they are so hard to observe. Recall that the strong force increases with distance. This has the effect that any effort to tear a proton or neutron apart to get at the quarks inside will get increasingly more difficult as the quarks get farther apart. As the force increases the energy to keep them separate also increases until there

Quarks	u up	c charm	t top
	d down	s strange	b bottom
Leptons	ν_e e - Neutrino	ν_μ μ - Neutrino	ν_τ τ - Neutrino
	e electron	μ muon	τ tau
	I	II	III

The Generations of Matter

Figure 2.1: As the generation increases (I, II, III) the mass of the associated particles also increase. Things generally prefer to be in a lower energy state so the higher generations decay very quickly into their lower generation relatives. It is for this reason we must build accelerators to reach energies high enough to create these particles, then observe them before they decay to lower energy daughter particles.

is enough energy to create a quark anti-quark pair which separate and combine with the two halves of the proton or neutron to make two separate particles with no unaccompanied quarks.

Quarks are generally divided up into three generations of increasing mass-energy. Figure 2.1 shows the hierarchy, as well as the charged leptons and neutrinos (whose introduction is to come) which are associated with each group.

Baryons and Mesons

As mentioned above we do not see quarks unaccompanied due to the nature of the strong nuclear force. Instead what we see are combinations of quarks called hadrons. It is hadrons that make up the particles that we are able to observe directly on a daily basis.

There are two types of hadrons, the baryon and the meson. A baryon is a hadron made of

three quarks or three anti-quarks, while a meson consists of two quarks, a quark and an anti-quark. Since protons are made of two up quarks and a down quark they are classified as a baryon. An example of a meson is the π^+ which is composed of an up quark and an anti-down quark.

Leptons

Leptons are defined as elementary particles that do not partake in the strong nuclear force. There are two kinds of leptons, those carrying electric charge and those that are electrically neutral. Those carrying the electric charge include the electron (e^-), the muon (μ^-), and the tauon (τ^-).

Neutrinos are the uncharged partners of the above leptons. They are very light (they interact very weakly with the gravitational force) with a yet undetermined mass and come in three varieties: the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ).

While all neutrinos interact via the weak nuclear force they have the peculiar rule that they will only associate with their charged lepton partner. That is to say an electron will only interact with an electron neutrino, never with a muon or tau neutrino.

2.2 The Higgs Doublet

The Higgs doublet is an addition to the Standard Model that came when the W^+ , W^- , and Z^0 were found to have masses. In the mathematics that describes interactions between particles via the four forces the force carrying particles (the field quanta) are assumed to have zero mass. If a mass for the field quanta is simply introduced into the equations they no longer function. The predictions they make turn into nonsensical, unobserved phenomenon. Another method must be taken into account in order to incorporate the massive field quanta into the Standard Model. This method is the Higgs Field. The Higgs field is an all pervasive presence in the universe that interacts

with any particle which partakes in the weak force. Along with this field also comes a field quanta, the Higgs Boson.

One way of envisioning how this field gives particles mass is to consider the concept of a cosmic drag. For any particle that interact with the Higgs field it seems as though it is traveling through tapioca pudding. In any whichever direction it attempts to travel its movement is slowed by the viscous desert.

The amount which this tapioca pudding slows down a particle traveling through it depends on how much the particle couples with the pudding. Electrons have a low coupling, and as such are not effected very much. A top quark has a high coupling, and so is dramatically slowed. It is this coupling strength that manifests itself as mass.

Then an electron has a small mass because it has a low coupling with the Higgs field. The same holds for the heavy top quark, except it has a high coupling with the field. Introducing this doublet massive force particles are now able to exist within the Standard Model.

2.3 Supersymmetry

It is a prediction of the Standard Model that the strength of the electromagnetic force will increase as the energy of the interacting system increases. Similarly it is predicted that the weak and strong nuclear force will both decrease at higher energies. This has been experimentally shown to be true over a limited energy range, and when plotted (see figure 2.2) has the interesting property that these three forces don't quite converge at high energy (they almost do at 10^{16} GeV).

For a unified theory, if these three forces were to converge to the same point exactly it would have reached it goal. At a certain energy all of the forces would coalesce into a single force creating a grand unified force controlling everything. In the Standard Model as it stands, the forces

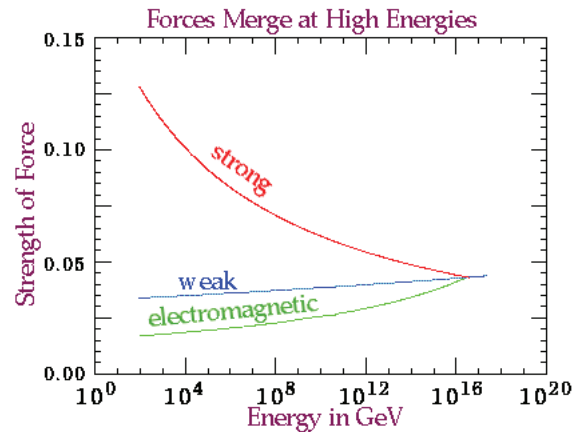


Figure 2.2: As higher energies are reached the three forces seem to coincide to a common point. Although they do not meet exactly, they are close enough to hint that some underlying cause might be behind their closeness.[3]

come close and then barely miss intersecting at a single point.

A theory that would fix this up is Supersymmetry (SUSY). Supersymmetry hypothesizes the existence of an additional set of particles (and anti-particles) to the currently accepted set described in section 2.1.2. For each of the fundamental particles observed there would be an as yet unobserved partner (a supersymmetric partner) which would have exactly the same properties except for two characteristics: it would have a different, most likely higher mass, and if the original particle were a boson, the partner would be a fermion (and vice versa).

If these particles are taken into account in extrapolations of the strength of forces in figure 2.2 it is possible the discrepancy at higher energies would disappear. The three forces of electromagnetism, weak, and strong would become the same at high enough energies. Although this energy is greatly beyond energies we can currently have access to it is not above the energy available in the universe closely following the big bang.

In order for this to work the masses would have to be close to the W , Z , and H (which has yet to be observed at the writing of this thesis). Once the energies necessary to create these

particles have been achieved, measuring the properties of these states will reflect directly on physics at the unification scale.

3 International Linear Collider

The International Linear Collider (ILC) is a proposed linear detector that will collide electron/positron beams together at up to 500 GeV in the center-of-mass frame. It is currently planned to be 35 kilometers in length with linear accelerators extending a little over 12 kilometers in opposite directions away from a central detector. While the Large Hadron Collider (LHC), a contemporary circular accelerator, will be operating at higher energies, the total energy of the collision will be divided up among the constituent quarks, anti-quarks, and gluons that comprise the protons and anti-protons the LHC beams are made up of. The ILC will step in to perform precision measurements using a much simpler electron positron collision, obtaining much higher accuracy than results from the LHC.

In league with the LHC, the ILC will potentially answer questions concerning the Higgs Boson, Supersymmetry, Extra Dimensions, Dark Matter, and the Unification of the forces. While none of these are guaranteed a definitive answer, limitations of the Standard Model point to exciting potentialities for this energy range.

3.1 The Detector

Proposed Detector plans for the ILC are composed of five main sections: the Vertex Detector, the Central Tracker Barrel, the Electromagnetic Calorimeter, the Hadron Calorimeter and the Muon Detector. The different detectors have been built with the idea of particle flow in mind, tracing the path between one section into another so the behavior of a particle can be better understood and tracking efficiency can be increased. This is achieved through fine enough Calorimeter separation that energy associated with charged particles can be separated from that of the neutral particles. General details about the different sections of the detectors come from Perricone[2].

When the electron and positron are initially fed into the collision zone they collide and decay into a shower of daughter particles. Exposed to the 5 Tesla uniform magnetic field present throughout the detector the charged particles travel in helical paths from which their momentum can be determined. Neutral particles are unaffected by the magnetic field and continue in a straight line until they are stopped in the calorimeter, or if neutrinos, leave the detector entirely. See figure 3.1.

3.1.1 Coordinate System

The coordinate system used for the detector is shown in figure 3.2 where the z-axis represents the beam pipe. Since the detectors are arrayed cylindrically around the beam pipe it is

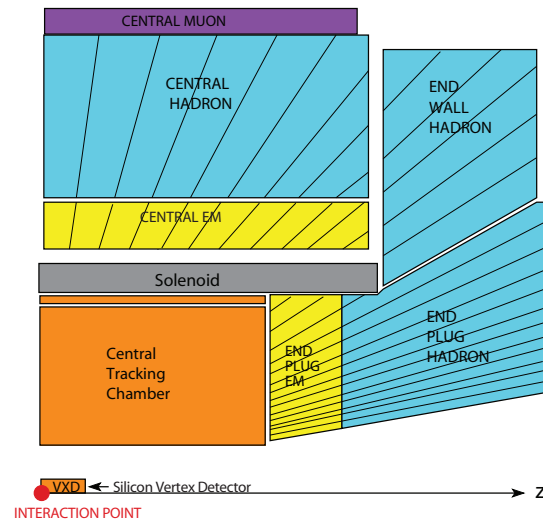


Figure 3.1: A quadrant view of the proposed detector geometry for the ILC[4]

convenient to use cylindrical coordinates that are defined by the transformations

$$r = \sqrt{x^2 + y^2}$$

$$\phi = \arctan(y/x)$$

$$z = z$$

This has the effect of transforming all x-y information onto the r - ϕ plane.

3.1.2 Tracking Chamber

There are three proposed technologies for the Tracking Chamber. One is a large gaseous chamber which is capable of tracking particles in real time. Another uses an outer chamber of gas with an inner tracker made of silicon strips. The final design would be an entirely silicon strip based chamber for both the inner and outer regions of the Tracking Chamber. The work done for this thesis uses simulated data from an all silicon tracker so we will go into more detail about this

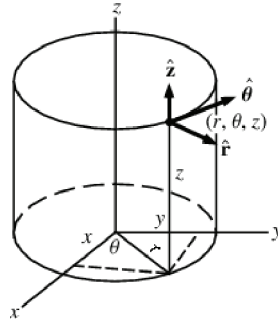


Figure 3.2: A graphical representation of cylindrical coordinates. The cylinder shown above can be thought of as one of the layers of the tracking chamber, where the z axis is the beam pipe and the origin is the interaction point.

particular device.

The all silicon Tracking Chamber is comprised of both the Vertex and Tracker Barrel Detectors. Although the Vertex has a much higher resolution the Tracker Barrel gives information over a larger region which enables accurate curvatures measurements for charged particles.

Vertex Detector

The Vertex Barrel Detector is a pixel detector comprised of a solid-state pixel technology, such as CCDs, that has yet to be chosen. It will have a resolution of better than 5 microns (5 millionths of a meter) in the r -, ϕ -, and z -directions and be comprised of five concentric barrels. The barrels will have radii of approximately 15, 23, 36, 48, and 61 millimeters from the beam pipe. With pixelization 30 times more accurate than the LHC and detectors 3-6 times closer the Vertex Detector will be better able to examine behavior immediately surrounding the electron/positron collision point.

Tracker Barrel Detector

Composed of five more concentric cylinders of silicon strip detectors the default design Tracker Barrel takes measurements at radii of approximately 221, 471, 721, 971, and 1221 millimeters. The detectors are currently proposed to be in strips approximately 10 centimeters long, giving very accurate r/ϕ information (again 5 microns) but only limited z information (about 5 centimeter precision).

3.1.3 Electromagnetic Calorimeter

Taking cue from the Stanford Linear Accelerator Center (SLAC) ILC group, a silicon-tungsten calorimeter is thought to be one of the best options for the electromagnetic (EM) calorimeter. It will begin at 1,207 millimeters from the beam pipe, about 50 millimeters away from the final layer of the tracker barrel.

The EM calorimeter will have a granularity of around 1 centimeter which will give precise enough information to separate out charged from neutral particles and match the charged particles with the path they create through the tracking chamber. It will be responsible for finding and stopping electrons, positrons, and photons, and measuring their energy with good precision.

3.1.4 Hadron Calorimeter

The hadron calorimeter will have the same design and components as the electromagnetic calorimeter but will be less finely segmented and consist of more stopping material. When a hadron passes close enough to a nucleus in this dense material it will be acted upon by the strong nuclear force which will cause an interaction and produce a chain of daughter particles. These particles will continue to interact with other nuclei creating a shower until they are low enough energy to be

absorbed by the stopping material.

3.1.5 Muon Chamber

At this point the only particles to have traversed the length of the detector and retain their momentum are muons and neutrinos. While the neutrinos will continue on, the Muon Chamber will track the muons and record their energy. The only way to view how the neutrinos behaved is to add up the energy found by the detector and notice where (in what directions) energy is missing as predicted by conservation of energy and momentum.

4 Terminology

Similar to most specialized fields high energy particle physics has a large amount of jargon specific to the experimentalists which comprise it. Most terms are self explanatory, but some are not and it is in the best interest of clarity to go through those most commonly used.

Hit

A hit is a coordinate inside the detector which represents a point where a particle was detected as it passed through a certain section of the detector. There are three types of hits used in this project, three dimensional, two dimensional, and calorimeter hits.

A three dimensional hit refers to points which have detailed information available for r-, ϕ -, and z-position. These generally come from the vertex detector since the nature of its design (pixelization) gives measurements in all three axes to about 5 micron precision.

Two dimensional hits refer to hits with information available in two dimensions, generally the tracker barrel. Although the tracker barrel gives 5 micron precision in the r and ϕ coordinates, the z direction is limited to 5 centimeter precision. This is due to the crude z segmentation of the detectors described earlier. Because of the limited nature of the measurement in the third direction these points are considered two dimensional.

A calorimeter hit is a mark caused by an energy deposition left in an electromagnetic

calorimeter segment. Each hit has three dimensions associated with it but less precision (approximately 2 millimeters in r/ϕ). Since the absorption suffers from spreading of the shower this low precision in position measurement is expected.

A hit referred to as a tracker hit is a two or three dimensional hit made in the vertex or tracker barrel detector.

Track

A track represents the path a particle is believed to have taken through the detector. It is composed of hits and generally includes the parameters used to describe the helix that best describes its path. The parameters associated with the helix are the radius of closest approach, the azimuthal angle at the point of closest approach, the curvature κ , the z position of closest approach, and the dip angle at closest approach. The tracks momentum is derived from well known relations involving curvature and the strength of the magnetic field

$$\kappa = \frac{1}{r} = \frac{300|\mathbf{B}|}{p_{\perp}c}$$

where the radius of curvature r is in meters, magnetic field \mathbf{B} is in Teslas, and $p_{\perp}c$ is in MeV, where p_{\perp} is momentum transverse to the beamline.

The radius of closest approach is the r/ϕ point on the helix which comes closest to the origin of the detector (point of closest approach or PCA) in the plane perpendicular to the beamline. The azimuthal angle is the angle in the r/ϕ plane that points towards the base of the helix. The curvature κ is a measure of how curved the helix is and is inversely related to the radius of the helix. The z distance of closest approach is the distance in z from the origin to the r/ϕ PCA on the helix. The dip angle is the angle of the track relative to the plane perpendicular to the beamline.

It should be mentioned that this parameterization of the helix is not limited to a single

rotation but continues indefinitely in the positive and negative z directions. Except for the section involving the hits associated with the track the rest of the helix is ignored (a branch cut is made).

When referred to, an N -hit track is a track which contains N tracker hits. This does not include stubs which may or may not be associated with the track.

Stub

A stub is a track made out of calorimeter hits. It has the same parameterization variables associated with it as a normal track but its curvature measurement is very imprecise. This follows from the imprecision of the curvature measurement due to the short nature of the track and low precision of the calorimeter hits.

Monte Carlo Truth Particle

The monte carlo truth particle (MC Particle) for a hit is the actual simulated particle that caused the hit. In reality this information is not available, but since the data is simulated we are able to “cheat” and get it. It is not used for pattern recognition efforts, but is only used to benchmark the success of a particular routine. The probability of finding the track associated with a given MC Particle is known as the “efficiency,” and is discussed below.

When referred to as a majority MC Particle, it is describing the MC Particle that is associate with the majority of the hits attached to a track. For example, if there are four hits from particle A, and one hit from particle B, particle A would be the majority MC particle.

An N -hit MC Particle is a particle which only traversed N layers of the detector, and thus only left behind N tracker hits.

Event

An event refers to the results from one electron/positron collision. It includes all the hit and monte carlo truth information for an electron/positron collision.

Track Hit Purity

Each track has an associated purity which is the ratio of hits associated with the majority MC particle and the total number of hits. By this definition the maximum purity for a track is 1, and for an N hit track the minimum purity is $1/N$ (corresponding to all hits coming from different MC particles).

Since a pattern recognition routine is a best guess method at finding tracks purity is a good measure of how successful the routine is on a per track basis. When running a pattern recognition algorithm a purity of 1 is strived for, but a track with a purity as low as 0.80 is considered to be found.

Fake Track

A fake track is a track that has a track hit purity less than 0.80. There are not enough hits from the same particle to give an accurate measure of the particles reconstructed momentum and path through the detector.

Fake tracks are an important consideration when evaluating the tracker performance as a way of quantitatively measuring the success of different tracking algorithms.

Findable Particle

A findable particle is a MC particle that falls under a specific set of criteria developed by Bruce Schumm, Tyler Rice, and Lori Stevens of UC Santa Cruz. These criteria determine the total

number of MC Particles we expect to find in the event. The criteria are:

- FINAL or INTERMEDIATE state particle. This means the particle is either the final particle in a decay chain, or one of the intermediates. In other words, not an unknown particle.
- Charged.
- Path Length > 50 cm
- $p_{\perp} > 0.75$ GeV
- Radius of Origin < 72 cm
- $|\cos \theta| < 0.8$

If a particle passes all of the above criteria it represents a particle we believe a pattern recognition algorithm should be able to find. If it is not found we look into the reasons why and attempt to correct them.

Acceptable Track

The criteria for acceptable tracks determine which of the tracks found by an algorithm should be considered given what data we expect to see from the event. It is a set of checks to ensure we avoid tracks that are unlikely to represent MC Particles. The criteria were also developed by Bruce Schumm, Tyler Rice, and Lori Stevens all of UC Santa Cruz. They include:

- $|\cos \theta| < 0.8$
- $p_{\perp} > 0.75$ GeV
- Distance of Closest Approach < 10 cm

- **Successfully Fit.** This means a fitting routine was able to find an equation describing the path through the collection of hits making up the track.

By imposing these restrictions on the tracks found by a pattern recognition algorithm we are able to weed out fake tracks from real ones.

Efficiency

Efficiency is a description of how many particles were successfully found out of the total number of particles in the event. In its mathematical form it is simply

$$\text{efficiency} = \frac{N_{\text{acceptable}}}{N_{\text{findable}}}$$

where $N_{\text{acceptable}}$ is the number of reconstructed tracks that pass the acceptable criteria, and N_{findable} is the number of findable MC Particles.

Reconstructible Purity

Reconstructible purity is the ratio of acceptable reconstructed tracks with a track hit purity greater or equal to 0.8 to the total number of acceptable tracks. The reconstructible purity is similar to track hit purity, but gives a measure of how successfully the algorithm performed over the entire event instead of just on a single track.

5 Software

With the large volume and complexity of the data from the ILC it is necessary to use computers to help in the analysis. The ILC Collaboration currently uses the Java coding language to perform its analysis, developing the code in the NetBeans Integrated Development Environment (IDE) and running over simulated Monte Carlo data sets in Java Analysis Studio. There is also a large frame work of code (the `org.lcsim` base libraries) that including basic classes for tracks, hits, and stubs as well as implementation code such as the driver that loads the data and provides main functionality for analysis. Together these form the tools necessary to create a track reconstruction algorithm.

5.1 `org.lcsim` Class Library

The `org.lcsim` class library is open source code available to the public online. It is where all work done by the collaboration is uploaded and is continually updated as people refine old classes and develop new algorithms for fitting, tracking, and keeping track of different information from the event files.

Included in this library is the basic driver upon which all other code is developed. Using this driver the hit information from the event is loaded into the code one event at a time. Also

included are basic classes for hits, tracks, and stubs, as well as circle and helix fitting routines, and basic swimming (extrapolation) algorithms.

5.2 NetBeans IDE 6.0

The NetBeans Integrated Development Environment (version 6.0) is used to manage the broad base of code developed by the ILC Collaboration. It provides a single suite where the author can easily view other work in the org.lcsim package and incorporate it into their own routine. It takes care of writing, compiling, and combining the compiled libraries into a jar file which can be run by Java Analysis Studio.

5.3 Java Analysis Studio 3

Java Analysis Studio 3 (JAS) provides a graphical user interface to run over event data. It supplies the events to the main driver and interprets the output into series of histograms and scatter plots easily viewed by the user.

Also included in JAS is the Event Viewer which gives a three dimensional, draggable representation of the detector geometry as well as the option to toggle on or off hits and monte carlo truth particles. This is useful for a conceptual view of how an event in the detector actually looks.

5.4 Monte Carlo Simulated Data Set

A Monte Carlo data set is created in two steps, ISAJET and GEANT. ISAJET simulates the interactions of proton/proton, proton/ant-proton, and electron/positron collisions.

After ISAJET simulates the physics the output is run through GEANT, a program which describes how the simulated particles will pass through and interact with the detector. The result is a file which contains all the hit information that would be available had the collision actually occurred with additional MC Particle information available for analyzing the efficiency and purity of the event.

6 The Code

The code used for this project falls into two categories; the old and the new. The old code was written by separate authors as stand alone packages focusing on tracker barrel reconstruction. The new code implements all these separate packages into a single package that runs through each package consecutively to form a combined tracking effort. After each package is run through the hits from the tracks found are removed from the event so that tracks are not found multiple times. A newly developed routine which focuses on calorimeter assisted pattern recognition for the last three layers of the tracker barrel (layers 3, 4, and 5) is then run, finishing off the analysis of the event. Once each event is analyzed the results are written out to a raw data file in addition to being plotted in histograms within the JAS environment.

6.1 Previous Work

The base code for the project consists of VXDCh eater, AxialBarrelTrackFinderZ, and GarfieldTrackFinder. There are also several helper routines which convert the output information of one package into a format readable by the next package. These packages are all designed to work individually, giving a best shot at finding all the tracks left in an event at the time they are run.

6.1.1 VXDCheater

VXDCheater (VerteXDetectorCheater) looks at the monte carlo truth particles for the event. If a particle leaves hits in four or more layers of the vertex detector all the hits in the event associated with this particle (but not its daughter particles) are removed from the event.

This is done because previous results have shown that vertex stub finding, track finding with four or more hits from the vertex detector, can be done at very high efficiencies and purities. Because the code for performing this pattern recognition and removal of hits from the event has not been ported to the current framework (included in the org.lcsim base libraries) the hits are removed under the assumption they will be found as a track once the code is available.

The tracks left over after this cheating process are all considered non-prompt tracks since they did not originate close to the origin and have few hits in the vertex barrel. As such all tracking algorithms following VXDCheater use only information from the tracker barrel. The original cheating code is written by Eric Wallace, a UC Santa Cruz undergraduate at the time.

6.1.2 AxialBarrelTrackFinderZ

AxialBarrelTrackFinderZ (ABTFZ) is a branch of the AxialBarrelTrackFinder code written by Tim Nelson. The original code takes all three hit combinations from the tracker barrel hit banks and makes a seed track out of every combination possible. There are several requirements they must pass before they are grouped into a seed, including a maximum angular separation in the r/ϕ plane. This acts to cut out seeds we know aren't tracks from the start and decrease run time.

Once acceptable seeds are determined, hits from the next layer are added to see if they improve or degrade the chi-square value of the circle fit. If the degradation in chi-square is not too

drastic the hit is added to the seed. The same process is then repeated for the next layer.

After being extended into all tracker barrel layers the resulting track is passed through a chi-square test, discarding the track if it is above a certain value. All tracks that are left after this cut are output to the event record.

ABTFZ is an augmentation of this code by Lori Stevens of UC Santa Cruz that checks the seeds for z-consistency after adding hits. Z-consistency assumes the motion in the r/z plane is linear, and determines if the modules that are hit are consistent with a line.

6.1.3 GarfieldTrackFinder

GarfieldTrackFinder (GTF) employs Calorimeter Assisted Tracking (CAT, like Garfield the cat). The code starts with calorimeter stubs and extends them into the tracker barrel starting at the last layer. Each hit on a subsequent layer is looked at to see if it lies within a certain distance of the predicted position. A similar chi-square test as ABTFZ is performed to see how adding each hit affects the goodness of the fit.

Written by Dmitry Onoprienko of the University of Kansas, GTF comes with an entire suite of support classes that include GarfieldHit, GarfieldTrack, GarfieldHitConverter, and more. GarfieldHit and GarfieldTrack are extensions of the basic Hit and Track classes included in the org.lcsim library, and GarfieldHitConverter takes regular hits and turns them into GarfieldHit's which are the input for GTF.

6.2 New Work

Using the previously written code, a new routine was written that combined the algorithms into one monster algorithm called MultiDriver. In addition, a specialized body of code was designed

N Hits	Find. Part.	ABTFZ Found	ABTFZ Fake	GTF Found	GTF Fake	Missed	Eff.	Rec. Purity
5	161	156	0	0	0	5	0.97	1.0
4	77	69	17	3	6	5	0.94	0.76
3	23	N/A	N/A	8	37	15	0.35	0.18

Table 6.1: Results after running MultiDriver. Results are for ABTFZ finding 4-hit tracks and assumed 10 centimeter segmentation.

to find particles which only left hits in the final three layers of the tracker barrel. Essentially weeding out all the bad three hit seeds used in ABTFZ, it was dubbed SeedExtend.

6.2.1 MultiDriver

To improve on pattern recognition the above three routines were combined to form a package called MultiDriver. MultiDriver first runs VXDCheater, then ABTFZ, then GTF. After each step, hits on found tracks are removed before the next step begins so the same track is not found twice. The results are read out to histograms in JAS which give the efficiency and purity of the entire driver as well as analysis on each of the components separately.

Table 6.1 shows the results after running MultiDriver. For four and five hit tracks the results show over 94% of tracks being found with no fake 5-hit tracks found and an acceptable number of fake 4-hit tracks. It should be noted the purity for 4-hit tracks can be improved to above 95% if ABTFZ is only allowed to find 5-hit tracks. In this case GTF finds a large number of the 4-hit findable particles but efficiency is decreased by a noticeable amount.

Examining three hit tracks shows that not only is the efficiency very low, but the fake tracks outnumber the good tracks by a factor of 4.6:1. Since certain SUSY signatures are thought to occur between the second and third layer of the tracker barrel this needs to be improved.

6.2.2 The Problem

The problem is three hit tracks are not being found. Since any randomly chosen three hits form a circle the chi-square goodness of fit test does nothing to differentiate between good and bad tracks (both have perfect chi-square). When GTF attempts to find tracks it starts at the stubs and traces them inwards, using the stub as a fourth hit to increase the purity of three hit tracks found. It is the way this extra information is used that decreases its effect on three hit tracks.

When stubs are fit they do not use all the information available to them. An average is first taken of calorimeter hits closest to the origin to find a base point and a base direction, but no curvature (κ) information is derived. It is with these average values from the first half of the stub that a track is traced out to attach tracker hits.

This method works well for layers that are close to the stub, but as the distance increases more and more area must be swept out to account for ignoring κ in the stub calculations. The other restriction with this method is the lack of precision data from the calorimeter. With 2 millimeters error in the position of calorimeter hits the curvature measurement is very inaccurate, giving little information about κ .

While GTF is able to sort out some bad three hit tracks from all the candidates, there is not enough information or precision available with this method to successfully cut out all of the bad tracks.

6.2.3 The Solution

The solution is to take the redeeming qualities of both the previous pattern recognition routines (ABTFZ and GTF) and combine them. To find three hit tracks it is necessary to find a way to weed out bad tracks. All of the tracks that don't arise from a particle traveling through the

detector need to be thrown out so that only the pure tracks remain.

ABTFZ starts by creating three hit seeds from tracker barrel hits. The circles fit through these hits are very accurate since each hit has x-y information precise to five microns. GTF takes the additional information offered by the electromagnetic calorimeter and uses it as extra information. Combining these two methods yields the routine SeedExtend.

SeedExtend

SeedExtend starts with a set of three hit seeds that include every possible combination of hits found on the last three layers of the tracker barrel (layers 3, 4, and 5). The list of seeds is reduced with an initial round of cuts that focus on the behavior a plausible track would show. The hits must be within 0.3 radians of each other, the momentum derived from the curvature of the fitted circle must be less than 0.5 GeV (billion electron volts), and the distance of closest approach in the r/ϕ plane must be less than 200 millimeters.

This initial set of cuts reduces the number seeds by a factor of 20. Next the list of stubs is taken from the event file. The cluster of hits associated with each stub is taken and re-fit, with a good stub typically containing between 20 and 60 calorimeter hits. This is done so that each hit is taken into account when the helix for the stub is found. In order to get the best helical fit result the three points closest to the origin (the base), farthest from the origin (the end), and in the middle of the stub (half way in between base and end), are weighted by a factor of 10 in the chi-square fit. This process takes longer than the averaging method that is used by GTF but gives us valuable curvature information.

Once the event has a list of likely seeds and stubs it tries to match them up. Instead of starting with the inaccurate helix fit from the stub, the seeds are taken as the starting point of the

track. Each seed is then extrapolated to the position it enters the EM calorimeter, after which each stub track cycles through all of these entry positions looking for a match.

To determine which stubs result from potential seed candidates there are four comparisons; base difference, ϕ difference, z difference, and kappa ratio. The base difference is the difference in the r/ϕ plane from the point where the seed enters the EM calorimeter to the base of the stub. The ϕ difference is the difference in angle between the helices tangent line projected into the r/ϕ plane. Z difference is the spatial separation in the z-direction. Kappa ratio is the difference in seed and stubs curvature all divided by the curvature of the seed, $|\kappa_{seed} - \kappa_{stub}|/\kappa_{seed}$.

The values used to decide whether seeds and stubs matched were found using single muon events. These are Monte Carlo simulated data sets which place muons at the center of the detector with momenta of different magnitude and direction. The result is a single, clean path through the detector with minimal noise. The experimental results of this process are shown in figure 6.1. The results were the base difference was never more than 2 millimeters, the phi difference was never more than 300 milliradians, the z difference was always below 200 millimeters and the kappa ratio cut off was around 10.

For the first three the placement of the cut was clean and easily visible from the data. The kappa difference was found to be less helpful since the radius of curvature from the stubs had a large error. This is again due to the limited precision of the calorimeter hits as well as the minimal length the curvature measurement is taken over.

After passing a stub through this comparison with all of the seeds in the event there is generally no match, but in the case where the seed represents an actual particle one or more matches can occur. In the case that there is only one match the track is added to the found track list. When multiple seeds are found to match up with a single stub the seed which has the least difference in

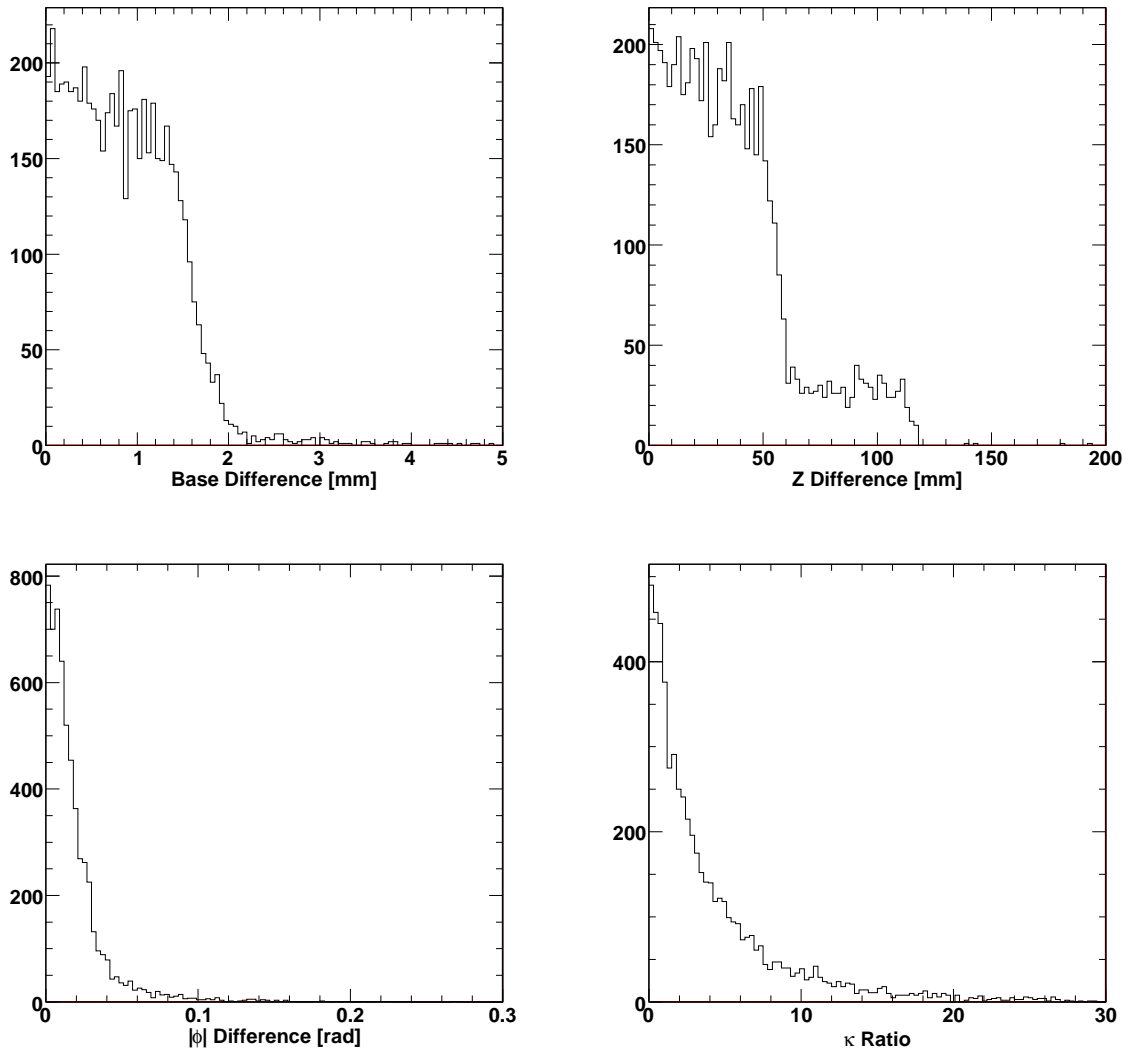


Figure 6.1: The distributions used to determine cuts for matching seed tracks to stub tracks.

ϕ is chosen.

Once a list of all successfully matched tracks is made event clean up is performed. This involves taking all the hits for each track and removing them from the event hit banks. This is

done in case another routine is developed further down the road which will try to find tracks after SeedExtend runs. The tracks are then added to the event track lists so that analysis routines will have easy access to them.

The code will be available to the public online in the org.lcsim CVS repository, *:pserver:anonymous@cvs.freehep.org:/cvs/lcd*, located in the *src/org/lcsim/contrib/scipp* directory. It is not currently uploaded at the writing of this thesis but is planned to be released within the next two months.

7 Results

Table 7.1 shows the results from running SeedExtend after ABTFZ has had a chance to weed out 4- and 5-hit tracks. This is done to reduce the number of possible three hit seeds SeedExtend must run through. Since the calorimeter is being used to assist in tracking SeedExtend is only expected to find tracks which actually enter the electromagnetic calorimeter. For this reason only tracks which have hits on the fifth tracking layer are considered for efficiency measurements. This extends the findable MC truth particles to include those originating before 721 millimeters. In accordance, the number of findable 3-hit particles is decreased from 23 to 19 since we don't expect to find any 3-hit tracks that did not enter the EM calorimeter.

Out of the 19 particles that were considered findable 3-hit tracks 12 of them were found by SeedExtend giving an efficiency of 92%. At the same time there was only 1 fake track for 12 good tracks, giving a dramatically increased purity ratio.

There were two other tracks found by SeedExtend that did not correspond with 3-hit

N Hits	Acc. Part.	SE Found	SE Fake	Eff.	Rec. Pur.
3	19	12	1	0.92	0.92
4	8	2	0	0.25	1.0

Table 7.1: Results after running SeedExtend following MultiDriver. Acceptable 4-hit particles are the missed 4-hit tracks from running ABTFZ. 10 centimeter segmentation is assumed.

particles, but instead with 4-hit particles left behind by ABTFZ. Since it is unknown which 4 layers the remaining findable 4-hit tracks traversed it is impossible to compare the efficiency measurement with previous results. This is a good result none the less since we did nothing to actively find 4-hit tracks.

8 Conclusion

SeedExtend greatly improved on the efficiency of previous attempts to find 3 hit particles by over a factor of 1.5. What is more impressive is the increase in purity from 0.12 good tracks for every fake track to 12 good tracks for every fake track (an increase by a factor of 100). However, it should be noted that these results come from comparison to packages that were not designed for the sole purpose of finding 3-hit particles. It is also still true that the efficiency of the algorithm remains on the low end at 55% for all findable three hit particles.

However, in comparison with the total number of tracks in an event, three hit tracks do not make up a majority or even a large minority. The efficiency of finding three hit tracks in standard ILC collisions is not that important for event reconstruction since they comprise such a small amount out of the much larger number of total tracks. Instead it is necessary to find them to focus on the interesting physics for which 3-hit tracks could play a significant role. This is the purpose for developing this more general algorithm for 3-hit track finding.

Now that such a framework is in place it can be fine tuned to search for certain channels which are thought to occur in the outer layers of the tracking chamber. One such example is the decay

$$slepton \rightarrow lepton + gravitino$$

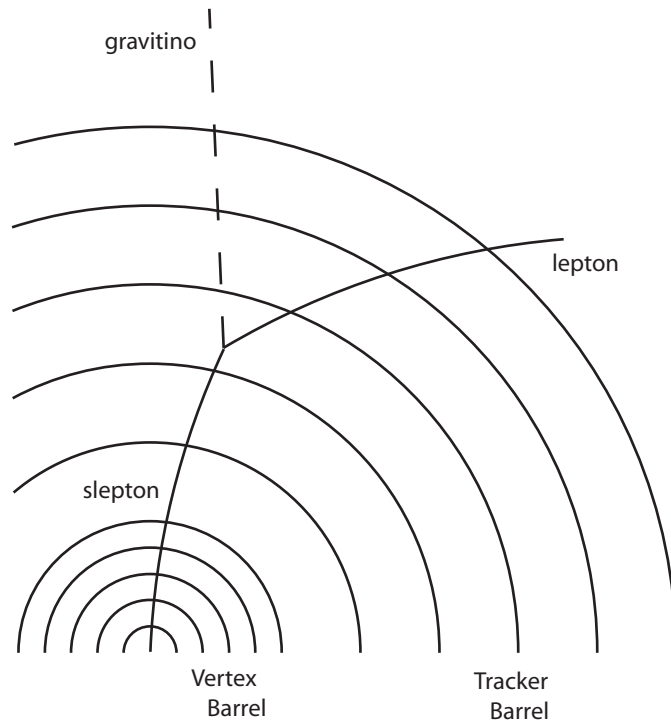


Figure 8.1: The decay channel of $slepton \rightarrow lepton + gravitino$. Since the gravitino doesn't leave a mark in the detector it looks as if one particle traversed the tracking chamber making a discontinuous change in direction at some point in the tracker barrel.

which is a tell tale sign of supersymmetry. In this decay all that is visible to the detectors are the slepton and the lepton, creating a kink at the point where the slepton decays (see figure 8.1). In this case the path of the lepton will appear as a track found by SeedExtend, while the slepton will be found by an inside out vertex stub tracking algorithm. Once found these two tracks can be traced to their intersection and the kink can be reconstructed indicating an slepton decay. The information this give us will shed direct light on the physics of grand unified theory scales. As of the writing of this thesis this work is currently being undertaken by a group led by Bruce Schumm with results to come within the next year.

At this point the concept of 3 hit track finding has been proven; it now remains to be

improved upon. As special processes which require three hit tracks are thought of the code need only be fine tuned to reveal the new physics waiting just beyond our current energy limitations.

Bibliography

- [1] Schumm, Bruce. *Deep Down Things*. Maryland: The John Hopkins University Press 2004

- [2] Perricone, Mike. “Super-Fast Super-Sensitive Detectors”. *Symmetry Magazine*, Volume 02, September 2005.

- [3] Chin, Lawrence C.. *A Thermodynamic Interpretation of History*.
<<http://www.geocities.com/therapeuter2002/layereduniverse.html>> (15 May 2007)

- [4] Fermi National Accelerator Laboratory. *The Collider Detector at Fermilab*.
<http://www-cdf.fnal.gov/experiment/drawings/detector_drawings.html> (15 May 2007)