The Other Side of Particle Physics: Computation and Analysis

Uttam Paudel
Physics 129
University of California, Berkeley

November 23rd, 2010
In the beginning...

- **Discovery of Atomic Structure:**
  - "It was almost as if you fired a 15 inch shell into a piece of tissue paper and it came back and hit you.”__E. Rutherford

### Table: LXXIX. The Scattering of α and β Particles by Matter and the Structure of the Atom. By Professor E. Rutherford, F.R.S., University of Manchester.

<table>
<thead>
<tr>
<th>Metal</th>
<th>Atomic weight</th>
<th>$\alpha$</th>
<th>$\alpha / A^2 Z^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>207</td>
<td>62</td>
<td>508</td>
</tr>
<tr>
<td>Gold</td>
<td>197</td>
<td>67</td>
<td>242</td>
</tr>
<tr>
<td>Platinum</td>
<td>195</td>
<td>63</td>
<td>232</td>
</tr>
<tr>
<td>Tin</td>
<td>119</td>
<td>34</td>
<td>296</td>
</tr>
<tr>
<td>Silver</td>
<td>108</td>
<td>27</td>
<td>241</td>
</tr>
<tr>
<td>Copper</td>
<td>64</td>
<td>14.5</td>
<td>225</td>
</tr>
<tr>
<td>Iron</td>
<td>56</td>
<td>19.2</td>
<td>250</td>
</tr>
<tr>
<td>Aluminium</td>
<td>27</td>
<td>3.4</td>
<td>243</td>
</tr>
</tbody>
</table>

Average: 233

\[
N(\theta) = \frac{N_i nLZ^2k^2e^4}{4r^2KE^2\sin^4(\theta/2)}
\]

- $N_i =$ number of incident alpha particles
- $n =$ atoms per unit volume in target
- $L =$ thickness of target
- $Z =$ atomic number of target
- $e =$ electron charge
- $k =$ Coulomb's constant
- $r =$ target-to-detector distance
- $KE =$ kinetic energy of alpha
- $\theta =$ scattering angle
...the good old days

MARCH 15, 1933

PHYSICAL REVIEW

The Positive Electron

CARL D. ANDERSON, California Institute of Technology, Pasadena, California
(Received February 28, 1933)

Out of a group of 1300 photographs of cosmic-ray tracks in a vertical Wilson chamber 15 tracks were of positive particles which could not have a mass as great as that of the proton. From an examination of the energy-loss and ionization produced it is concluded that the charge is less than twice, and is probably exactly equal to, that of the proton. If these particles carry unit positive charge the curvatures and ionizations produced require the mass to be less than twenty times the electron mass. These particles will be called positrons. Because they occur in groups associated with other tracks it is concluded that they must be secondary particles ejected from atomic nuclei.

Editor

The time-consuming operation of searching exposed photographic plates for interesting “events” was conducted by a large group of young girls, officially known as “scanners” but unofficially called “Cecil’s Beauty Chorus”, without whom discoveries would certainly not have been made.

Fig. 1. A 68 million volt positron (fo = 1.1 × 10^9 gauss cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron (fo = 7.5 × 10^9 gauss cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.
"The finder of a new elementary particle used to be rewarded by a Nobel Prize, but such a discovery now ought to be punished by a $10,000 fine" — Willis Lamb
...and the new era begun

**Difficulties:**
- Particles with higher availability had already been discovered before 50’s
- Remaining were either rare or at high energy range or unstable, thus studying HEP using photographic plates became virtually impossible.

**Good News:**
- Large scale experiments
- Computerized data collections (i.e. more data)
  - BABaR has accumulated ~1 petabyte ($10^6$ GB) of data
  - Fermilab: 10TB/day ~ data stored in the entire library of congress with 530 miles of bookshelves
...but with more data comes more responsibilities

- Extremely difficult to analyze and handle large data sets
- Signals are usually buried by unwanted processes
- Typical discovery involves finding 2-100 signature events (with $5\sigma$ C.L ~ 99.9999%)
Ways to extract information

- Using kinematics
  - Conservation of Energy, Momentum, Charge, and other Quantum Numbers
  - But.. it is still not good enough to extract 1 event out of 10,000,000 events
  - High chance of getting biased and being fooled by fake signals
    - Ex: particle “Oops-Leon” initially though to be a new particle

- Blind Analysis
  - Performed without looking at the answer
  - Reduces or eliminates experimenter’s bias

*Classic Example: “Pepsi Challenge”*
Blind Analysis

- One Scenario
  - Choice of experimental selection or cut
  - Cut value could be chosen arbitrarily,
  - Results depend upon the choice of Sensitivity value

- Another Scenario
  - Measurement of rare signals
  - Experimentalist could be tempted (with a good intention) to remove a few extra “background-like” events
  - Could potentially overestimate the branching fraction
Methods

- **Hidden Signal Box:**
  - Signal region is hidden during the analysis (if known)
  - Analysis is conducted on the sidebands or on Monte Carlo data
  - Once the analysis method, selection cuts, and background estimates are fixed, the box is open

- **Fractional Signal:**
  - When an experimenter is expecting to see certain % or # of signals
  - An unknown fraction \( F \) of the total signal \( N \) is given for analyses
    \[
    N' = N \times F
    \]
  - \( F \) is revealed after the completion of the analysis with the fractional data, no bias on the number \( N \)
    - Ex: Sudbury Neutrino Observatory, where the experimenters wish to report the total number \( N \) of certain neutrinos
Example

Search for a Light Scalar Higgs Boson in the Radiative Transition of $\Upsilon(1S) \to \gamma A^0 \to \tau \tau$ in the BABaR dataset!

A schematic diagram of Higgs decayed from Upsilon(1S).
Decay is possible, only if $2m_\tau < m_{A^0} < m_{\Upsilon(1S)}$

- NMSSM prediction: A light scalar CP-odd Higgs boson (~a few GeV) from the decay of $\Upsilon(nS)$ resonances.
- We analyzed the decay of $\Upsilon(1S)$. Branching Fraction Prediction $\sim 10^{-6} - 10^{-4}$
- Data taken at B-Factory Babar
Data Samples

- **Signal:**
  - Monte Carlo (MC) simulation of Higgs in the broad mass range 4-9.45 GeV.

- **Backgrounds:**
  - Ups3S-OnPeak – 28.5 fb⁻¹.
  - GenericUpsilon2S – MC broad spectrum of \( \Upsilon(2S) \)
  - Largely dominated by unwanted processes such as:
    - \( e e \rightarrow e e \mu \mu \gamma \); \( e e \rightarrow e e \pi \pi \gamma \); \( e e \rightarrow \gamma \tau \tau \).
    - \( \Upsilon(1S) \rightarrow \gamma \tau \tau \), \( \Upsilon(1S) \rightarrow \gamma \mu \mu \), \( \Upsilon(1S) \rightarrow \gamma e e \).

- Collected during March and April of 2008.
Overview

- **Split variables into 3 groups – Pions, Photon, Higgs**
  - **Pion variables**: $\Upsilon(2S) \rightarrow \pi^+\pi^- \Upsilon(1S)$, identified by searching for two low-momentum pions ($\pi$);
  - **Photon variables**: $\Upsilon(1S) \rightarrow \gamma A^\circ$, identified by detecting a high energy photon.
  - **Higgs variables**: $A^\circ \rightarrow \tau^+\tau^-$ decays identified by 2 charged particles and significant amount of missing energy.

- **Pre-selection requirements**: # Charged tracks: 4, # of EMC crystal hits by photon, $\text{gammaCrys}$, Number of $\pi^0 = 0$
  - Standard selection criteria based on the detector
Variables

Example of variables used in the analysis:
1) Angle between 2 leptons in the CM frame
2) Energy of the lepton system
5) Distance between lepton vertex and ee beam
6) UPolarCosTh: Cosine angle between the lepton plane and the beam
9) UChi2Prob: Probability of $\tau^+ \tau^-$ coming from the same vertex.
10) $H_{\text{mass}}^2 = (P_i - P_f)^2$: where $P_f$ sums over the two charged pions and the photon, and $P_i$ is the $e^+e^-$ system

Figure: Vertex helps to distinguish signal from the background processes such as $Y(1S) \rightarrow \gamma \tau \tau$. 
Figure 4.1: Schematic view of a reaction of the type $e^+e^- \rightarrow \tau(\rightarrow \ell \bar{\nu}_\ell \nu_\tau)\tau(\rightarrow \pi/K\pi^0\pi^0\nu_\tau)$ with $\ell = e, \mu$ in the center-of-mass system. Only the final state particles indicated in black are reconstructed either directly (charged particles, solid lines) or using their decay products ($\pi^0$ mesons, dashed lines). The event is divided into two hemispheres by a plane perpendicular to the thrust axis (Equation 4.2). Due to the boost of the $\tau$ leptons, the decay products of each $\tau$ are constrained to its respective hemisphere. Signal events (Equation 4.1) hence contain exactly one track in each hemisphere. According to the charged particle which the hemispheres contain, they are called signal hemisphere and tag hemisphere.
Training Tools

- Utilize Machine Learning Techniques
  - Pattern Reorganization
  - Merge several variables and output a single one
    - Ex: Random Forest, Neural Network
  - Similar to how our brain works

- Toolkit for Multivariate Data Analysis (TMVA)
Figure: Signal vs. Background
Optimization

- Goal: Reasonably good signal efficiency with highest possible background rejection
- Optimize FOM for the variables using
  \[ \text{FOM} = \frac{S}{1.5 + \sqrt{B}}. \]
  \( S \) = signal, \( B \) = background
- Repeat the process until the cut value converges
Final Case Study

- Apply the optimized selection cuts and calculate signal efficiency,
- Understand PDF for background and signal,
- Study systematic errors,
- Unblind a fraction of the real data

\[ f(x; \alpha, n, \tilde{x}, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right), & \text{for } \frac{x-x_0}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-x_0}{\sigma})^{-n}, & \text{for } \frac{x-x_0}{\sigma} \leq -\alpha \end{cases} \]

where

\[ A = \left(\frac{n}{\alpha}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \]

\[ B = \frac{n}{|\alpha| - |\alpha|} \]

**Fig:** Probability distribution function for the MC and background

- Resonance, i.e., particle created from electron-positron collision

**HM^2** Generic Generic Gen Pdf

\[ \chi^2 / \text{ndf} = 0.439 \]

\[ \alpha = 0.0064 \pm 0.0049 \]

\[ a = 0.014 \pm 0.017 \]
Unblinding

- In BABaR
  - Working Group (AWG) reviews the analysis before unblinding,
  - a description of the cut optimization and background characterization procedures;
  - the expected number of background events in the signal box,
  - the signal efficiency from Monte Carlo or control samples;
  - the expected statistical sensitivity;
  - an estimation of systematic errors.

Final Procedure

- Fit the real data with the PDF
- Search for a peak/resonance
- Either make a new discovery or put constrains on the existing theories
- Publication
References

• Rutherford's Original paper:
  http://www.math.ubc.ca/~cass/rutherford/rutherford681.html

• The Positive Electron: C. Anderson, Caltech, Physical review, vol.43, 1933
  http://prola.aps.org/abstract/PR/v43/i6/p491_1

• Historical Timeline:
  http://particleadventure.org/other/history/smt.html

  http://prl.aps.org/abstract/PRL/v74/i14/p2632_1


• Blind Analysis in Particle Physics: A. Roodman, SLAC, PHYSTAT2003, 2003

• ROOT: http://root.cern.ch/drupal/

• TMVA: http://tmva.sourceforge.net/

• BABaR Collaboration, SLAC: http://www.slac.stanford.edu/BF/
Questions, Comments, Suggestions?