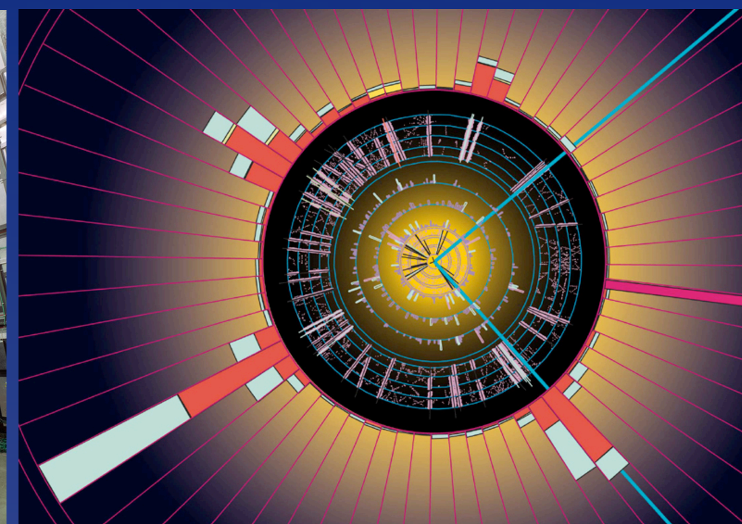
Experimental Particle Physics Analysis with *R*





Adam L. Lyon

Fermi National Accelerator Laboratory

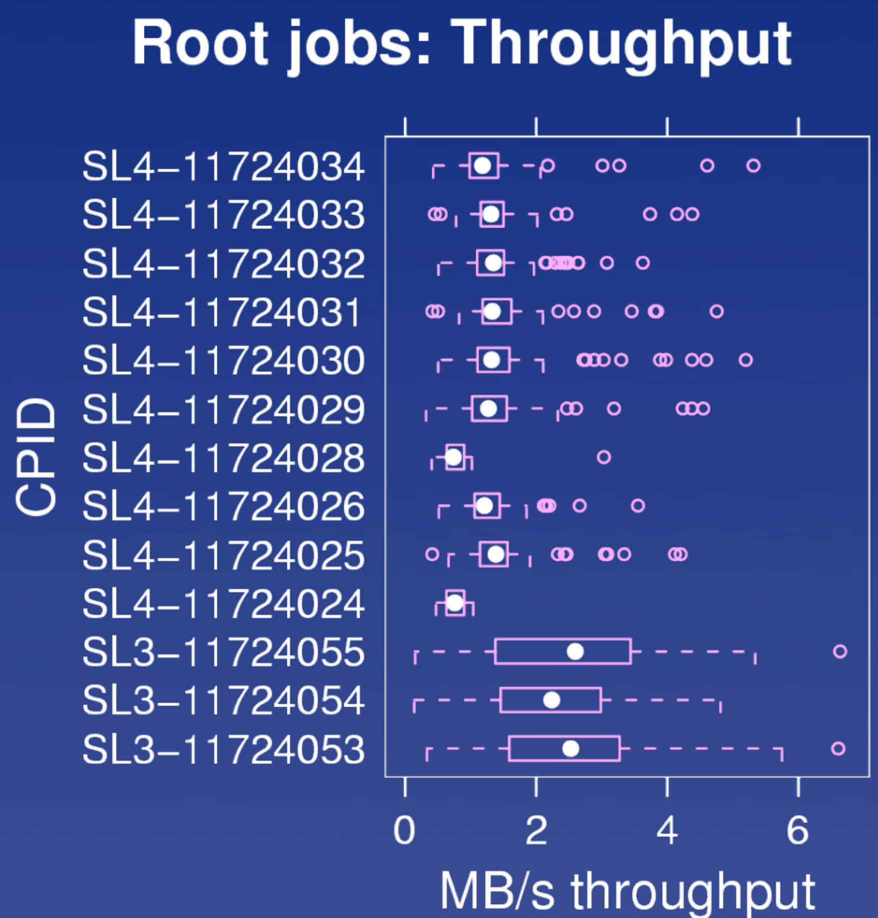
useR! 2007, Iowa State University

My story

A Crash Course in Particle Physics

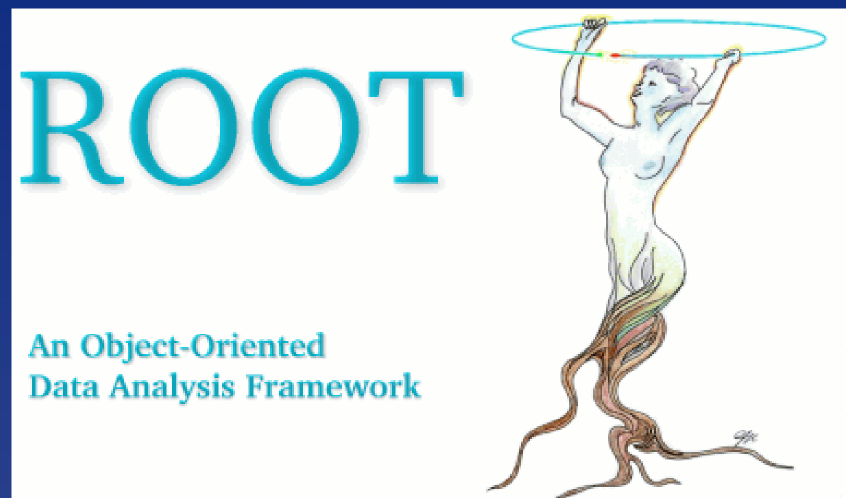
RootTreeToR

- I use *R* for analysis of a data handling system



- Easy to import data
  - Text files
  - Oracle DB
- Easy to manipulate data
- Easy interactive exploration
  - *S* Language
- Easy to make complicated plots
  - *Lattice*
  - *ggplot*

- For Experimental Particle Physics, *Root* is the ubiquitous data analysis tool (root.cern.ch)



- Command language is CINT (like interpreted C++)
- **Data format optimized for large data sets**
- Plotting library not as advanced as *R*
- Sophisticated data analysis not quite interactive (Complicated analysis requires compile/run step)

Want to use *R* to analysis EPP data in *Root* format...

RootTreeToR

A Crash Course in Particle Physics

(fasten your seat belts)

- What are the fundamental pieces of matter?
- How do they work?

- What are the fundamental pieces of matter?
- How do they work?

$p \quad n \qquad e$

- What are the fundamental pieces of matter?
- How do they work?

$$u \;\; d \qquad e$$

- What are the fundamental pieces of matter?
- How do they work?

$u \quad d$ $\qquad e$

Quarks $\qquad$ Leptons

- What are the fundamental pieces of matter?
- How do they work?

$u \quad d$ $e \quad \nu_e$

Quarks   Leptons

- What are the fundamental pieces of matter?
- How do they work?

t	b	τ	ν_τ
c	s	μ	ν_μ
u	d	e	ν_e
Quarks		Leptons	

- What are the fundamental pieces of matter?
- How do they work?

t	b	τ	ν_τ	Weak	$W\ Z$
c	s	μ	ν_μ	Strong	g
u	d	e	ν_e	Electromagnetic	γ
Quarks		Leptons		Force carriers	

- What are the fundamental pieces of matter?
- How do they work?

?? Higgs H

t	b	τ	ν_τ	*Weak*	$W\ Z$
c	s	μ	ν_μ	*Strong*	g
u	d	e	ν_e	*Electromagnetic*	γ
Quarks		Leptons		Force carriers	

- What are the fundamental pieces of matter?
- How do they work?

The Standard Model

?? Higgs H

t	b	τ	ν_τ	*Weak*	$W\ Z$
c	s	μ	ν_μ	*Strong*	g
u	d	e	ν_e	*Electromagnetic*	γ
Quarks		Leptons		Force carriers	

- Is there more? We think so!

Supersymmetry? Technicolor?
Extra dimensions?
Anomalous couplings?

?? Higgs H

Gravity?

t b	τ ν_τ		*Weak*	W Z
c s	μ ν_μ		*Strong*	g
u d	e ν_e		*Electromagnetic*	γ
Quarks	Leptons		Force carriers	

- Goals:
  - Determine the fundamental particles
  - Understand their properties & interactions
  - Does the Standard Model describe observations?
  - Is the Standard Model part of a more encompassing theory?


Fermilab Tevatron Collider

$$E = mc^2$$

$$E = mc^2$$

$$E = mc^2$$

$$E = mc^2$$

Quantum Mechanics
PROBABILITIES

Quantum Mechanics

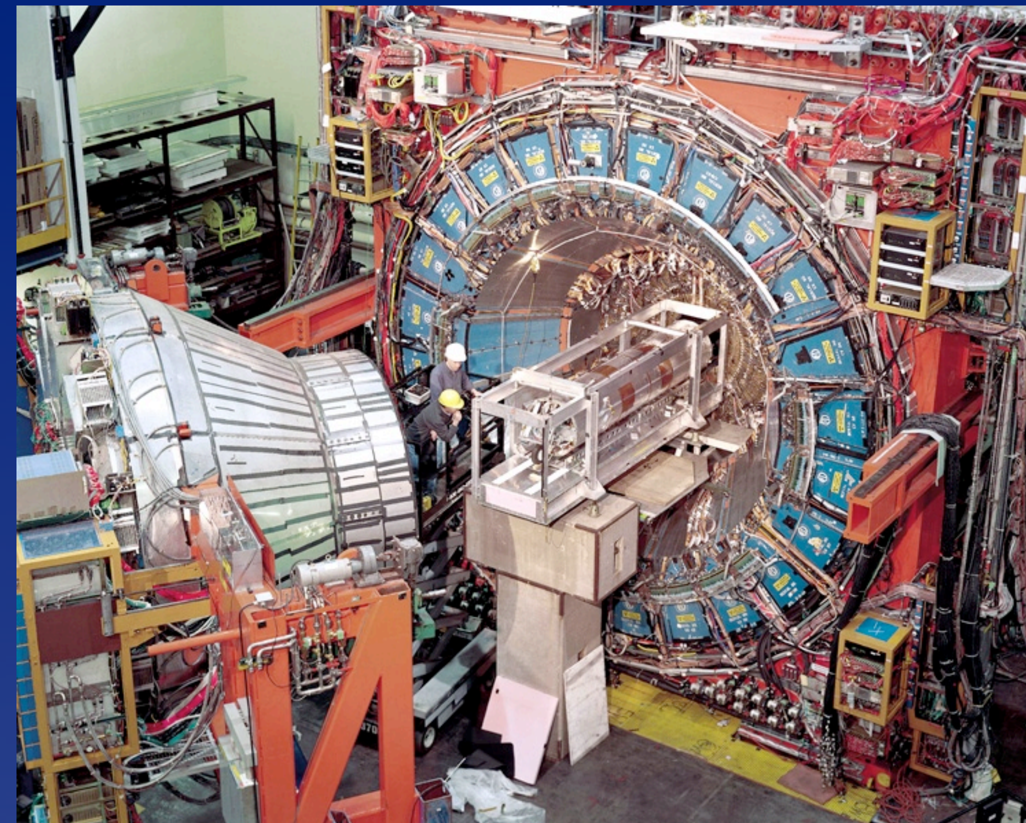
PROBABILITIES

Quantum M[illegible]chanics
PRO[illegible]BILITIES

Quantum Mechanics
PROBABILITIES

- The higher energy the collisions
  - The more massive the particles that can be produced

- The higher the collision rate (and the longer we run)
  - The rarer the interactions that can be observed



<u>Detectors</u>

We use the very big to learn about the very small!

Experimental Discrimination between Charge $2e/3$ Top Quark and Charge $4e/3$ Exotic Quark Production Scenarios

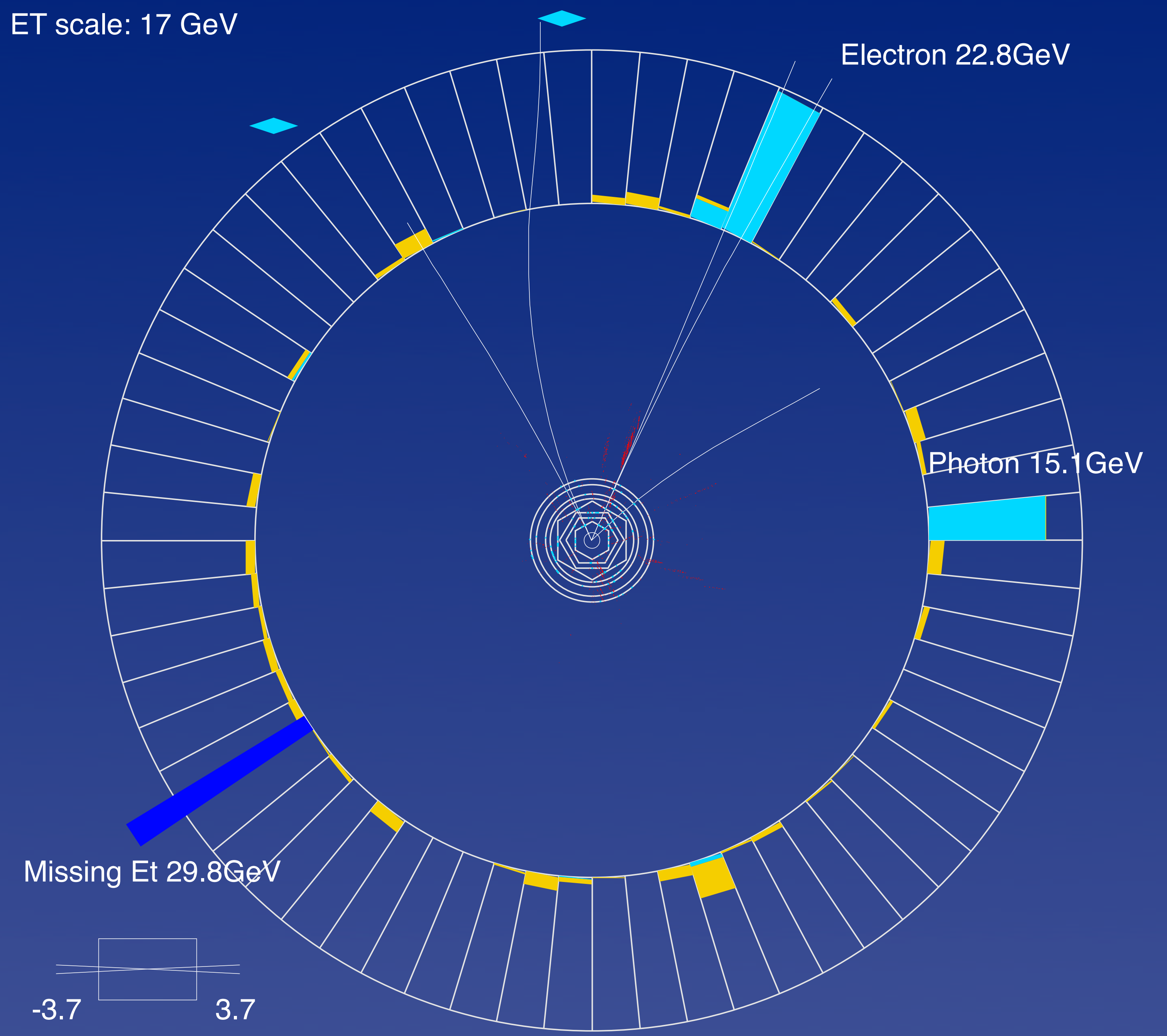
V. M. Abazov,[36] B. Abbott,[76] M. Abolins,[66] B. S. Acharya,[29] M. Adams,[52] T. Adams,[50] M. Agelou,[18] S. H. Ahn,[31] M. Ahsan,[60] G. D. Alexeev,[36] G. Alkhazov,[40] A. Alton,[65] G. Alverson,[64] G. A. Alves,[2] M. Anastasoaie,[35] T. Andeen,[54] S. Anderson,[46] B. Andrieu,[17] M. S. Anzelc,[54] Y. Arnoud,[14] M. Arov,[53] A. Askew,[50] B. Åsman,[41] A. C. S. Assis Jesus,[3] O. Atramentov,[58] C. Autermann,[21] C. Avila,[8] C. Ay,[24] F. Badaud,[13] A. Baden,[62] L. Bagby,[53] B. Baldin,[51] D. V. Bandurin,[60] P. Banerjee,[29] S. Banerjee,[29] E. Barberis,[64] P. Bargassa,[81] P. Baringer,[59] C. Barnes,[44] J. Barreto,[2] J. F. Bartlett,[51] U. Bassler,[17] D. Bauer,[44] A. Bean,[59] M. Begalli,[3] M. Begel,[72] C. Belanger-Champagne,[5] L. Bellantoni,[51] A. Bellavance,[68] J. A. Benitez,[66] S. B. Beri,[27] G. Bernardi,[17] R. Bernhard,[42] L. Berntzon,[15] I. Bertram,[43] M. Besançon,[18] R. Beuselinck,[44] V. A. Bezzubov,[39] P. C. Bhat,[51] V. Bhatnagar,[27] M. Binder,[25] C. Biscarat,[43] K. M. Black,[63] I. Blackler,[44] G. Blazey,[53] F. Blekman,[44] S. Blessing,[50] D. Bloch,[19] K. Bloom,[68] U. Blumenschein,[23] A. Boehnlein,[51] O. Boeriu,[56] T. A. Bolton,[60] G. Borissov,[43] K. Bos,[34] T. Bose,[78] A. Brandt,[79] R. Brock,[66] G. Brooijmans,[71] A. Bross,[51] D. Brown,[79] N. J. Buchanan,[50] D. Buchholz,[54] M. Buehler,[82] V. Buescher,[23] S. Burdin,[51] S. Burke,[46] T. H. Burnett,[83] E. Busato,[17] C. P. Buszello,[44] J. M. Butler,[63] P. Calfayan,[25] S. Calvet,[15] J. Cammin,[72] S. Caron,[34] W. Carvalho,[3] B. C. K. Casey,[78] N. M. Cason,[56] H. Castilla-Valdez,[33] D. Chakraborty,[53] K. M. Chan,[72] A. Chandra,[49] F. Charles,[19] E. Cheu,[46] F. Chevallier,[14] D. K. Cho,[63] S. Choi,[32] B. Choudhary,[28] L. Christofek,[59] D. Claes,[68] B. Clément,[19] C. Clément,[41] Y. Coadou,[5] M. Cooke,[81] W. E. Cooper,[51] D. Coppage,[59] M. Corcoran,[81] M.-C. Cousinou,[15] B. Cox,[45] S. Crépé-Renaudin,[14] D. Cutts,[78] M. Ćwiok,[30] H. da Motta,[2] A. Das,[63] M. Das,[61] B. Davies,[43] G. Davies,[44] G. A. Davis,[54] K. De,[79] P. de Jong,[34] S. J. de Jong,[35] E. De La Cruz-Burelo,[65] C. De Oliveira Martins,[3] J. D. Degenhardt,[65] F. Déliot,[18] M. Demarteau,[51] R. Demina,[72] P. Demine,[18] D. Denisov,[51] S. P. Denisov,[39] S. Desai,[73] H. T. Diehl,[51] M. Diesburg,[51] M. Doidge,[43] A. Dominguez,[68] H. Dong,[73] L. V. Dudko,[38] L. Duflot,[16] S. R. Dugad,[29] D. Duggan,[50] A. Duperrin,[15] J. Dyer,[66] A. Dyshkant,[53] M. Eads,[68] D. Edmunds,[66] T. Edwards,[45] J. Ellison,[49] J. Elmsheuser,[25] V. D. Elvira,[51] S. Eno,[62] P. Ermolov,[38] H. Evans,[55] A. Evdokimov,[37] V. N. Evdokimov,[39] S. N. Fatakia,[63] L. Feligioni,[63] A. V. Ferapontov,[60] T. Ferbel,[72] F. Fiedler,[25] F. Filthaut,[35] W. Fisher,[51] H. E. Fisk,[51] I. Fleck,[23] M. Ford,[45] M. Fortner,[53] H. Fox,[23] S. Fu,[51] S. Fuess,[51] T. Gadfort,[83] C. F. Galea,[35] E. Gallas,[51] E. Galyaev,[56] C. Garcia,[72] A. Garcia-Bellido,[83] J. Gardner,[59] V. Gavrilov,[37] A. Gay,[19] P. Gay,[13] D. Gelé,[19] R. Gelhaus,[49] C. E. Gerber,[52] Y. Gershtein,[50] D. Gillberg,[5] G. Ginther,[72] N. Gollub,[41] B. Gómez,[8] A. Goussiou,[56] P. D. Grannis,[73] H. Greenlee,[51] Z. D. Greenwood,[61] E. M. Gregores,[4] G. Grenier,[20] Ph. Gris,[13] J.-F. Grivaz,[16] S. Grünendahl,[51] M. W. Grünewald,[30] F. Guo,[73] J. Guo,[73] G. Gutierrez,[51] P. Gutierrez,[76] A. Haas,[71] N. J. Hadley,[62] P. Haefner,[25] S. Hagopian,[50] J. Haley,[69] I. Hall,[76] R. E. Hall,[48] L. Han,[7] K. Hanagaki,[51] P. Hansson,[41] K. Harder,[60] A. Harel,[72] R. Harrington,[64] J. M. Hauptman,[58] R. Hauser,[66] J. Hays,[54] T. Hebbeker,[21] D. Hedin,[53] J. G. Hegeman,[34] J. M. Heinmiller,[52] A. P. Heinson,[49] U. Heintz,[63] C. Hensel,[59] K. Herner,[73] G. Hesketh,[64] M. D. Hildreth,[56] R. Hirosky,[82] J. D. Hobbs,[73] B. Hoeneisen,[12] H. Hoeth,[26] M. Hohlfeld,[16] S. J. Hong,[31] R. Hooper,[78] P. Houben,[34] Y. Hu,[73] Z. Hubacek,[10] V. Hynek,[9] I. Iashvili,[70] R. Illingworth,[51] A. S. Ito,[51] S. Jabeen,[63] M. Jaffré,[16] S. Jain,[76] K. Jakobs,[23] C. Jarvis,[62] A. Jenkins,[44] R. Jesik,[44] K. Johns,[46] C. Johnson,[71] M. Johnson,[51] A. Jonckheere,[51] P. Jonsson,[44] A. Juste,[51] D. Käfer,[21] S. Kahn,[74] E. Kajfasz,[15] A. M. Kalinin,[36] J. M. Kalk,[61] J. R. Kalk,[66] S. Kappler,[21] D. Karmanov,[38] J. Kasper,[63] P. Kasper,[51] I. Katsanos,[71] D. Kau,[50] R. Kaur,[27] R. Kehoe,[80] S. Kermiche,[15] N. Khalatyan,[63] A. Khanov,[77] A. Kharchilava,[70] Y. M. Kharzheev,[36] D. Khatidze,[71] H. Kim,[79] T. J. Kim,[31] M. H. Kirby,[35] B. Klima,[51] J. M. Kohli,[27] J.-P. Konrath,[23] M. Kopal,[76] V. M. Korablev,[39] J. Kotcher,[74] B. Kothari,[71] A. Koubarovsky,[38] A. V. Kozelov,[39] J. Kozminski,[66] D. Krop,[55] A. Kryemadhi,[82] T. Kuhl,[24] A. Kumar,[70] S. Kunori,[62] A. Kupco,[11] T. Kurča,[20,*] J. Kvita,[9] S. Lammers,[71] G. Landsberg,[78] J. Lazoflores,[50] A.-C. Le Bihan,[19] P. Lebrun,[20] W. M. Lee,[53] A. Leflat,[38] F. Lehner,[42] V. Lesne,[13] J. Leveque,[46] P. Lewis,[44] J. Li,[79] Q. Z. Li,[51] J. G. R. Lima,[53] D. Lincoln,[51] J. Linnemann,[66] V. V. Lipaev,[39] R. Lipton,[51] Z. Liu,[5] L. Lobo,[44] A. Lobodenko,[40] M. Lokajicek,[11] A. Lounis,[19] P. Love,[43] H. J. Lubatti,[83] M. Lynker,[56] A. L. Lyon,[51] A. K. A. Maciel,[2] R. J. Madaras,[47] P. Mättig,[26] C. Magass,[21] A. Magerkurth,[65] A.-M. Magnan,[14] N. Makovec,[16] P. K. Mal,[56] H. B. Malbouisson,[3] S. Malik,[68] V. L. Malyshev,[36] H. S. Mao,[6] Y. Maravin,[60] M. Martens,[51] R. McCarthy,[73] D. Meder,[24] A. Melnitchouk,[67] A. Mendes,[15] L. Mendoza,[8] M. Merkin,[38] K. W. Merritt,[51] A. Meyer,[21] J. Meyer,[22] M. Michaut,[18] H. Miettinen,[81] T. Millet,[20] J. Mitrevski,[71] J. Molina,[3] N. K. Mondal,[29] J. Monk,[45] R. W. Moore,[5] T. Moulik,[59] G. S. Muanza,[16] M. Mulders,[51] M. Mulhearn,[71] L. Mundim,[3] Y. D. Mutaf,[73] E. Nagy,[15] M. Naimuddin,[28] M. Narain,[63] N. A. Naumann,[35] H. A. Neal,[65] J. P. Negret,[8] P. Neustroev,[40] C. Noeding,[23] A. Nomerotski,[51] S. F. Novaes,[4] T. Nunnemann,[25] V. O'Dell,[51] D. C. O'Neil,[5] G. Obrant,[40] V. Oguri,[3] N. Oliveira,[3] N. Oshima,[51] R. Otec,[10]

G. J. Otero y Garzón,[52] M. Owen,[45] P. Padley,[81] N. Parashar,[57] S.-J. Park,[72] S. K. Park,[31] J. Parsons,[71] R. Partridge,[78] N. Parua,[73] A. Patwa,[74] G. Pawloski,[81] P. M. Perea,[49] E. Perez,[18] K. Peters,[45] P. Pétroff,[16] M. Petteni,[44] R. Piegaia,[1] J. Piper,[66] M.-A. Pleier,[22] P. L. M. Podesta-Lerma,[33] V. M. Podstavkov,[51] Y. Pogorelov,[56] M.-E. Pol,[2] A. Pompoš,[76] B. G. Pope,[66] A. V. Popov,[39] C. Potter,[5] W. L. Prado da Silva,[3] H. B. Prosper,[50] S. Protopopescu,[74] J. Qian,[65] A. Quadt,[22] B. Quinn,[67] M. S. Rangel,[2] K. J. Rani,[29] K. Ranjan,[28] P. N. Ratoff,[43] P. Renkel,[80] S. Reucroft,[64] M. Rijssenbeek,[73] I. Ripp-Baudot,[19] F. Rizatdinova,[77] S. Robinson,[44] R. F. Rodrigues,[3] C. Royon,[18] P. Rubinov,[51] R. Ruchti,[56] V. I. Rud,[38] G. Sajot,[14] A. Sánchez-Hernández,[33] M. P. Sanders,[62] A. Santoro,[3] G. Savage,[51] L. Sawyer,[61] T. Scanlon,[44] D. Schaile,[25] R. D. Schamberger,[73] Y. Scheglov,[40] H. Schellman,[54] P. Schieferdecker,[25] C. Schmitt,[26] C. Schwanenberger,[45] A. Schwartzman,[69] R. Schwienhorst,[66] J. Sekaric,[50] S. Sengupta,[50] H. Severini,[76] E. Shabalina,[52] M. Shamim,[60] V. Shary,[18] A. A. Shchukin,[39] W. D. Shephard,[56] R. K. Shivpuri,[28] D. Shpakov,[51] V. Siccardi,[19] R. A. Sidwell,[60] V. Simak,[10] V. Sirotenko,[51] P. Skubic,[76] P. Slattery,[72] R. P. Smith,[51] G. R. Snow,[68] J. Snow,[75] S. Snyder,[74] S. Söldner-Rembold,[45] X. Song,[53] L. Sonnenschein,[17] A. Sopczak,[43] M. Sosebee,[79] K. Soustruznik,[9] M. Souza,[2] B. Spurlock,[79] J. Stark,[14] J. Steele,[61] V. Stolin,[37] A. Stone,[52] D. A. Stoyanova,[39] J. Strandberg,[41] S. Strandberg,[41] M. A. Strang,[70] M. Strauss,[76] R. Ströhmer,[25] D. Strom,[54] M. Strovink,[47] L. Stutte,[51] S. Sumowidagdo,[50] A. Sznajder,[3] M. Talby,[15] P. Tamburello,[46] W. Taylor,[5] P. Telford,[45] J. Temple,[46] B. Tiller,[25] M. Titov,[23] V. V. Tokmenin,[36] M. Tomoto,[51] T. Toole,[62] I. Torchiani,[23] S. Towers,[43] T. Trefzger,[24] S. Trincaz-Duvoid,[17] D. Tsybychev,[73] B. Tuchming,[18] C. Tully,[69] A. S. Turcot,[45] P. M. Tuts,[71] R. Unalan,[66] L. Uvarov,[40] S. Uvarov,[40] S. Uzunyan,[53] B. Vachon,[5] P. J. van den Berg,[34] R. Van Kooten,[55] W. M. van Leeuwen,[34] N. Varelas,[52] E. W. Varnes,[46] A. Vartapetian,[79] I. A. Vasilyev,[39] M. Vaupel,[26] P. Verdier,[20] L. S. Vertogradov,[36] M. Verzocchi,[51] F. Villeneuve-Seguier,[44] P. Vint,[44] J.-R. Vlimant,[17] E. Von Toerne,[60] M. Voutilainen,[68,†] M. Vreeswijk,[34] H. D. Wahl,[50] L. Wang,[62] M. H. L. S Wang,[51] J. Warchol,[56] G. Watts,[83] M. Wayne,[56] M. Weber,[51] H. Weerts,[66] N. Wermes,[22] M. Wetstein,[62] A. White,[79] D. Wicke,[26] G. W. Wilson,[59] S. J. Wimpenny,[49] M. Wobisch,[51] J. Womersley,[51] D. R. Wood,[64] T. R. Wyatt,[45] Y. Xie,[78] N. Xuan,[56] S. Yacoob,[54] R. Yamada,[51] M. Yan,[62] T. Yasuda,[51] Y. A. Yatsunenko,[36] K. Yip,[74] H. D. Yoo,[78] S. W. Youn,[54] C. Yu,[14] J. Yu,[79] A. Yurkewicz,[73] A. Zatserklyaniy,[53] C. Zeitnitz,[26] D. Zhang,[51] T. Zhao,[83] B. Zhou,[65] J. Zhu,[73] M. Zielinski,[72] D. Zieminska,[55] A. Zieminski,[55] V. Zutshi,[53] and E. G. Zverev[38]

(D0 Collaboration)

[1]*Universidad de Buenos Aires, Buenos Aires, Argentina*
[2]*LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil*
[3]*Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[4]*Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
[5]*University of Alberta, Edmonton, Alberta, Canada, Simon Fraser University, Burnaby, British Columbia, Canada, York University, Toronto, Ontario, Canada, and McGill University, Montreal, Quebec, Canada*
[6]*Institute of High Energy Physics, Beijing, People's Republic of China*
[7]*University of Science and Technology of China, Hefei, People's Republic of China*
[8]*Universidad de los Andes, Bogotá, Colombia*
[9]*Center for Particle Physics, Charles University, Prague, Czech Republic*
[10]*Czech Technical University, Prague, Czech Republic*
[11]*Center for Particle Physics, Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[12]*Universidad San Francisco de Quito, Quito, Ecuador*
[13]*Laboratoire de Physique Corpusculaire, IN2P3-CNRS, Université Blaise Pascal, Clermont-Ferrand, France*
[14]*Laboratoire de Physique Subatomique et de Cosmologie, IN2P3-CNRS, Universite de Grenoble 1, Grenoble, France*
[15]*CPPM, IN2P3-CNRS, Université de la Méditerranée, Marseille, France*
[16]*IN2P3-CNRS, Laboratoire de l'Accélérateur Linéaire, Orsay, France*
[17]*LPNHE, IN2P3-CNRS, Universités Paris VI and VII, Paris, France*
[18]*DAPNIA/Service de Physique des Particules, CEA, Saclay, France*
[19]*IPHC, IN2P3-CNRS, Université Louis Pasteur, Strasbourg, France and Université de Haute Alsace, Mulhouse, France*
[20]*Institut de Physique Nucléaire de Lyon, IN2P3-CNRS, Université Claude Bernard, Villeurbanne, France*
[21]*III. Physikalisches Institut A, RWTH Aachen, Aachen, Germany*
[22]*Physikalisches Institut, Universität Bonn, Bonn, Germany*
[23]*Physikalisches Institut, Universität Freiburg, Freiburg, Germany*

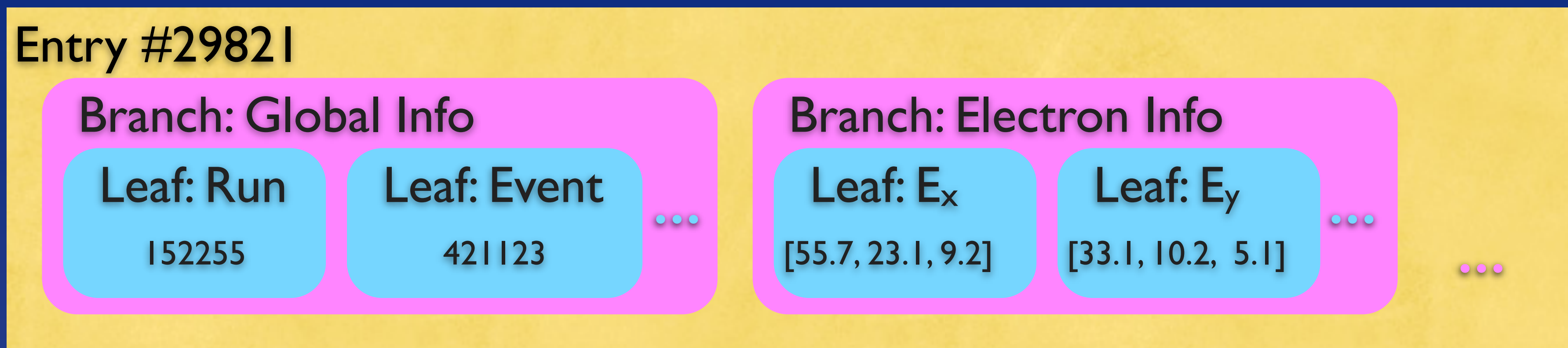
Run 169224 Event 663831



RootTreeToR

- Remember: each event is independent from all others
- We keep ~400 columns of data about each event
- We have over 2 billion events (“rows”)
- A typical data sample (400M events) is 16 TB!
- Typically skim down to 10,000s of events for intense analysis
- *Root* is ubiquitous in the field
    - i/o optimized for large data sets
    - C++ OO (compiled and interpreted [sort of])

- Data in Root Tree (like a hierarchical database)
  - An entry represents an event (=collision)



- Data accessed directly or via wrapper classes imported into Root

- Interactive: `myTree.Draw("electron.px()")`
  - Complicated algorithms are difficult to do interactively (using / programming boundary)
  - Trees are static

- End up writing big C++ programs
    - Run set of functions per event
    - Accumulate histograms
    - Perhaps write out smaller specialized Root Trees
- Not interactive!
- Difficult to explore the data
- Lots of tweaking and re-running

- Strategies for importing data into *R*

- Translate Root Tree to into something *R* can directly read (text file, DB). But too slow and painful.

- *REALIZATION*: Don't need all of the data all of the time!
  - Is why Root's optimization works well

- Try an "apply" function (e.g. *rootApply*)
  - An *R* function is called for each tree entry
  - Root data for entry is passed in as data frame
  - Returns list of results
  - Return lists are aggregated into a data frame
  - But too slow; Large overhead from *R*
  - Could make apply function C, but that's not *R*

- <u>Use Root for what Root is good at: I/O</u>
  - e.g. the Draw command is very fast (only reading one or two leaves)

- Use Root functions (TTreePlayer) to create an *R* data frame with data from specified leaves
  - Loop from within Root instead of *R*
  - Fill data frame from C++ code
  - Any data Root Tree::Draw can read, this can read
  - No overhead from *R* *(!!)*
  - Full power of *R* can analyze the data frame
  - Once task is complete, repeat with other aspects of data

- *RootTreeToR Package*

- RootTreeToR
  - Find it at http://www.phystat.org
  - Requires Root to already be installed
  - To get started do `?RootTreeToR`

  - Use the `toR` command to bring data from Root into *R* (uses TTreePlayer)

  - **Use** `toRUser` **command to call your C++ function that fills the data frame**
    - For complicated data manipulations within Root
    - Package provides a C++ class to represent an *R* data frame (prevents mistakes)

- <u>Summary</u>

- *R has* lots to offer for analyses of EPP data
  - Advanced plotting
  - Advanced statistics
  - Advanced data manipulations with *S* and packages (e.g. interpolation of irregularly spaced data ... akima)
  - An efficient interactive data manipulation language

- Use *Root* for its i/o strength

- Realize one does not need all of the data all of the time