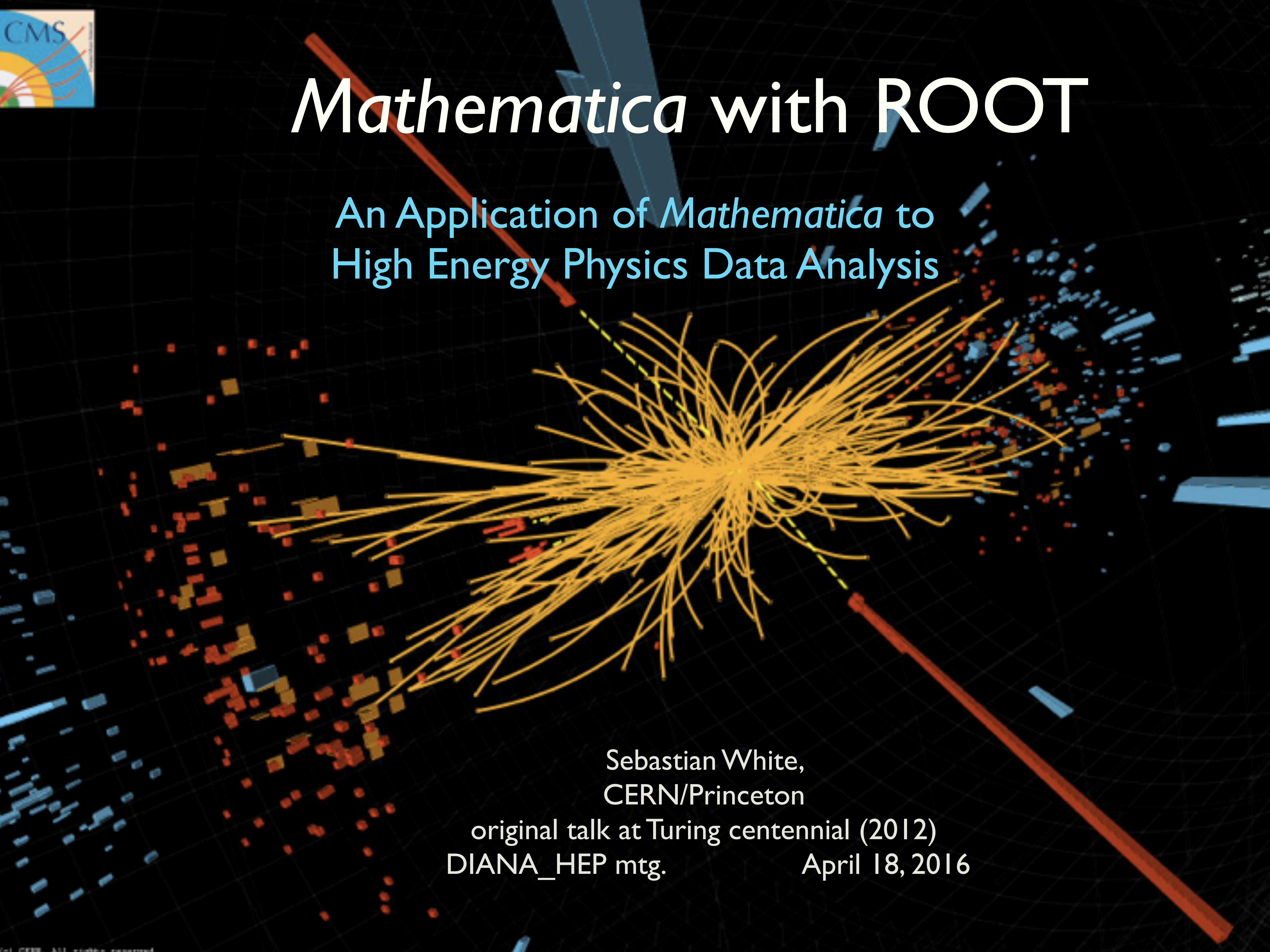


Mathematica with ROOT

An Application of *Mathematica* to High Energy Physics Data Analysis

Sebastian White,
CERN/Princeton

original talk at Turing centennial (2012)

DIANA_HEP mtg. April 18, 2016

“Mathematica with ROOT”

in collaboration with Ken Hsieh, WR

released Feb. 2011 on following sites:

- CERN ROOT page: http://root.cern.ch/drupal/content/mathematica-importer-root-data-files
- Wolfram.com/infocenter: http://library.wolfram.com/infocenter/Articles/7716/
- Feynman Computing Center: http://cd-docdb.fnal.gov/cgi-bin/DisplayMeeting?conferenceid=522
- “arxiv” preprint server (with Tom Throwe, BNL): http://arxiv.org/abs/1102.5068

significant help from the ROOT Developer Team: Fons Radermakers (CERN), Philippe Canal(Fermilab) and others (Fine, Brun, Nevski, etc.)

and from WR: Peter Overman, James Mulnix and initial encouragement from Stephen Wolfram.

suggested by Peter Overman that we follow this with a blog about the HEP context. This talk is that blog.

Disclaimer: I am a “detector guy” -not a “computing guy”. (I led the construction of one of the ATLAS detectors-ZDC)

“Mathematica with ROOT” imports data into Mathematica by linking to the CERN libraries with MATHLINK.

It is capable of dealing with all of the commonly used data generated by the large Physics and Astronomy community that has adopted the CERN technologies.

This project coincided with release 8 of Mathematica so we benefitted from the customizable import/export features new in 8.

Timeline:

- August 2010: Initial contact with Stephen Wolfram, who suggested Peter Overmann as point of contact. Initial meetings with Peter and Ken Hsieh (both of whom have a background in theoretical physics).
- Sept. 2010: Interaction between ROOT team and WR to understand potential legal issues with LGPL(GNU Lesser General Public License).

•Oct.-Dec.2010: Following release of Mathematica 8, work begins in earnest
•Jan. 2011: significant help in stress testing the importer from Tom Throwe(BNL). Further testing on ATLAS production data (D3PD), typically using ~5 GB data file. Completion of document.
•Feb. 24, 2011: Visit to Urbana campus, release on multiple venues, presentation at Feynman Computing Center, Fermilab.

additional comments:

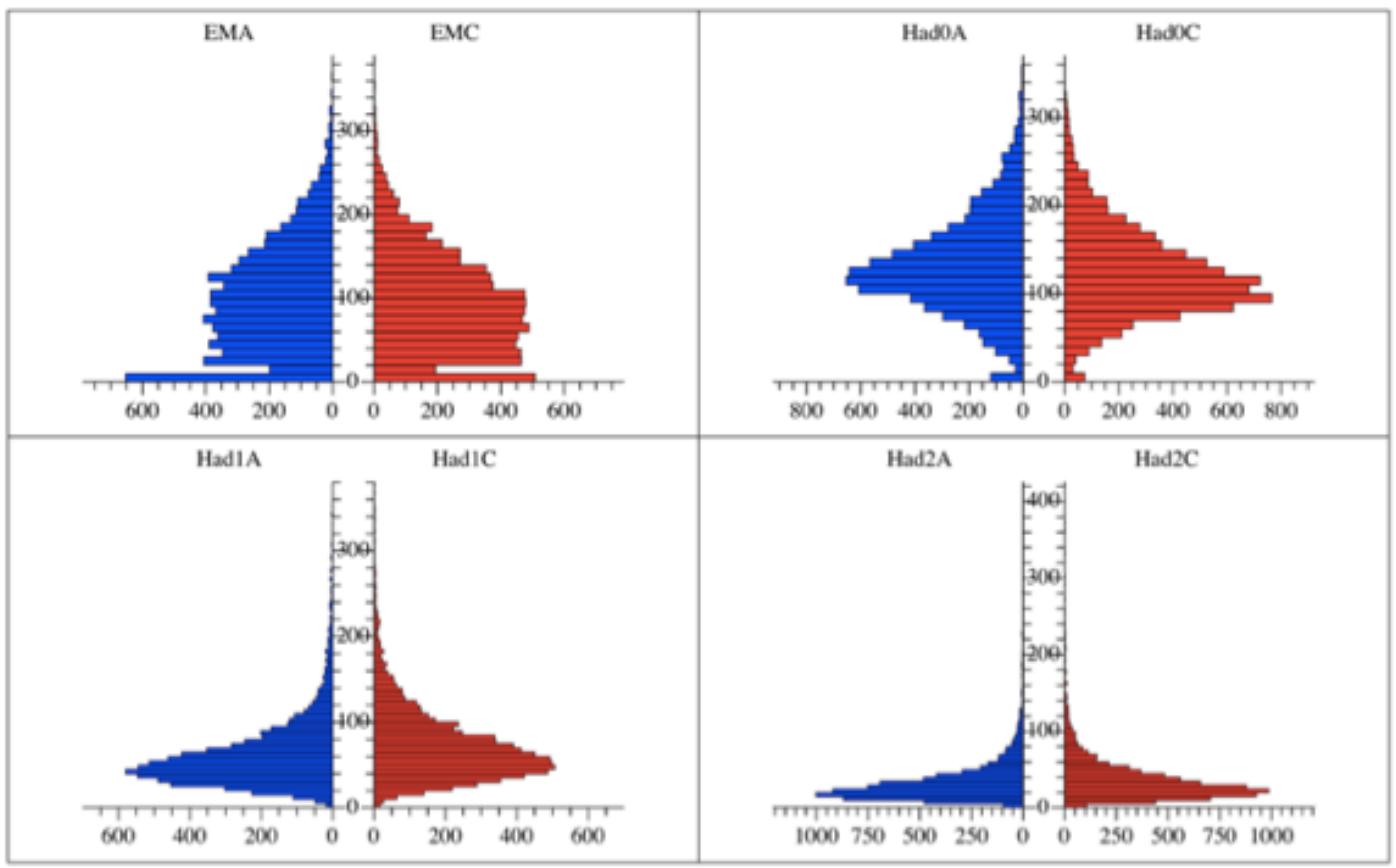
•package tested on most commonly used platforms: Windows (32 and 64 bit), Mac OS (32 and 64 bit), LINUX
•released with “idiot’s guide” for Mac Users
•most people able to get up and running on examples from CERN ROOT pages in ~2 hrs.
•interesting to see what people from ROOT side found neat about being able to do things in Mathematica

Fons chose the following example, from a 5 GB ATLAS D3PD I was analyzing. Not sure why. For me, I like working with Mathematica because things like "PairedHistogram" turn up when you need them.

In plots below I am checking gain balancing in layers of 2 calorimeters on either side of the ATLAS collision point, 280 m apart.

ZDC Analysis

```
GraphicsGrid[{{PairedHistogram[EMASignalHG, Style[EMCSignalHG, Red],
     ChartBaseStyle → Blue, ChartLabels → {"EMA", "EMC"}],
   PairedHistogram[HD0ASignalHG, Style[HD0CSignalHG, Red],
     ChartBaseStyle → Blue, ChartLabels → {"Had0A", "Had0C"}]},
  {PairedHistogram[HD1ASignalHG, Style[HD1CSignalHG, Red],
     ChartBaseStyle → Blue, ChartLabels → {"Had1A", "Had1C"}],
   PairedHistogram[HD2ASignalHG, Style[HD2CSignalHG, Red], ChartBaseStyle → Blue,
     ChartLabels → {"Had2A", "Had2C"}]}}, ImageSize → 560, Frame → All]
```




```
ZDCAHG = ZDCAmodHG.ZDCARelGain; ZDCCHG = ZDCCmodHG.ZDCCRelGain;
```

matrix and vector operations often useful in data reconstruction. Makes for readable code(last line).

Test Drive using examples from CERN ROOT Demonstrations (from our paper)

In[8]:=
```
(* This imports the metadata of a given tree *)
Import["cernstaff.root", {"ROOT", "TTreeMetadata", "T"}]
```

Out[8]=
```
{{Category, Category, Int_t, 3354}, {Flag, Flag, UInt_t, 3354},
 {Age, Age, Int_t, 3354}, {Service, Service, Int_t, 3354},
 {Children, Children, Int_t, 3354}, {Grade, Grade, Int_t, 3354},
 {Step, Step, Int_t, 3354}, {Hrweek, Hrweek, Int_t, 3354}, {Cost, Cost, Int_t, 3354},
 {Division, Division, Char_t, 3354}, {Nation, Nation, Char_t, 3354}}
```

In[9]:=
```
(* import the branch information of a TTree *)
header = {"Name", "Title", "Data Type", "Entries"};
branchinfo = Import["cernstaff.root", {"ROOT", "TTreeMetadata", "T"}];
Grid[Join[ {header}, branchinfo ], Frame → All]
```

Out[11]=

Name	Title	Data Type	Entries
Category	Category	Int_t	3354
Flag	Flag	UInt_t	3354
Age	Age	Int_t	3354
Service	Service	Int_t	3354
Children	Children	Int_t	3354
Grade	Grade	Int_t	3354
Step	Step	Int_t	3354
Hrweek	Hrweek	Int_t	3354
Cost	Cost	Int_t	3354
Division	Division	Char_t	3354
Nation	Nation	Char_t	3354

- **Importing only parts of one particular branch.**

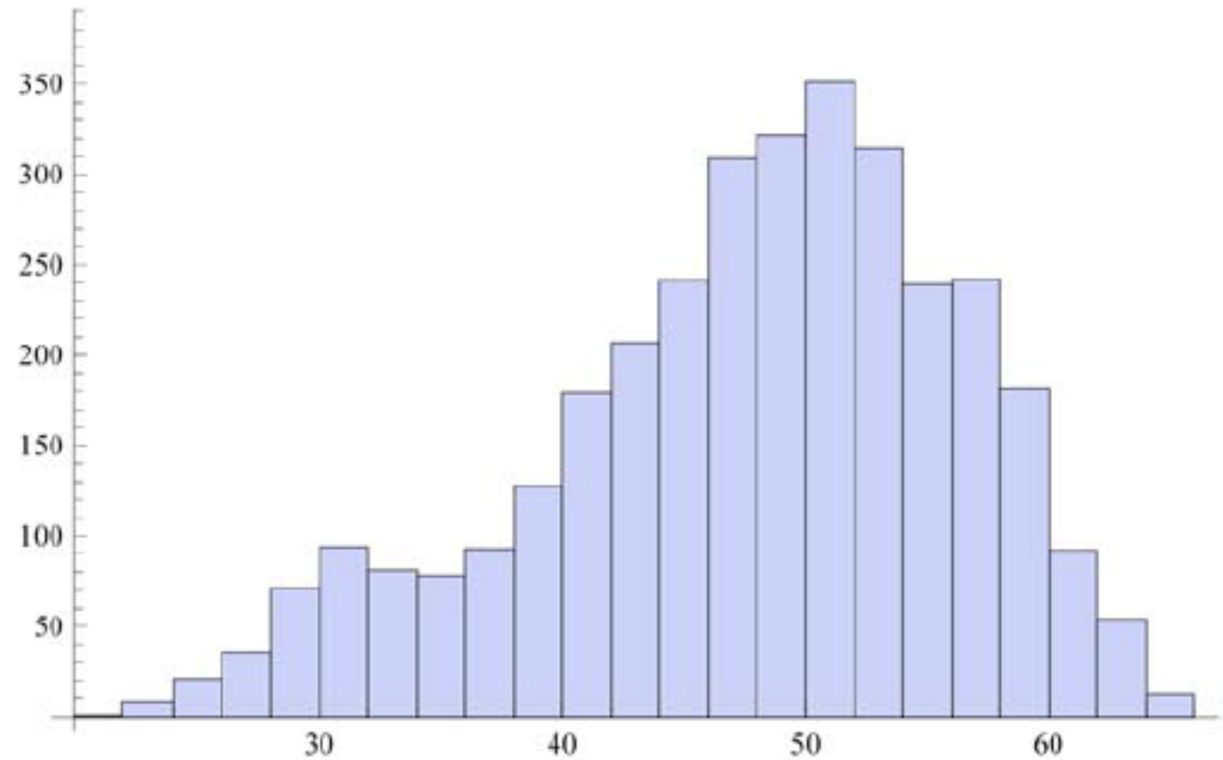
In[14]:=
```
dat = Import["cernstaff.root", {"ROOT", "TTreeData", "T", "Age"}, "Range" → {11, 20}]
```

```
Partial branch import with bounds: {11, 20}
```

Out[14]=
```
{51, 54, 54, 46, 54, 57, 55, 55, 57, 51}
```

In[12]:=
```
(* This imports the Age branch and plots it as with Histogram[] *)
dat = Import["cernstaff.root", {"ROOT", "TTreeData", "T", "Age"}];
Histogram[dat]
```

Out[13]=



In[2]:=
```
file = "cernstaff.root"
```

Out[2]=
```
cernstaff.root
```

In[3]:=
```
data = Import[file, {"ROOT", "TTreeData", "T"}, "Range" → {1, 2}]
```

Out[3]=
```
{{202, 530}, {15, 15}, {58, 63}, {28, 33}, {0, 0}, {10, 9},
 {13, 13}, {40, 40}, {11975, 10228}, {PS, EP}, {DE, CH}}
```

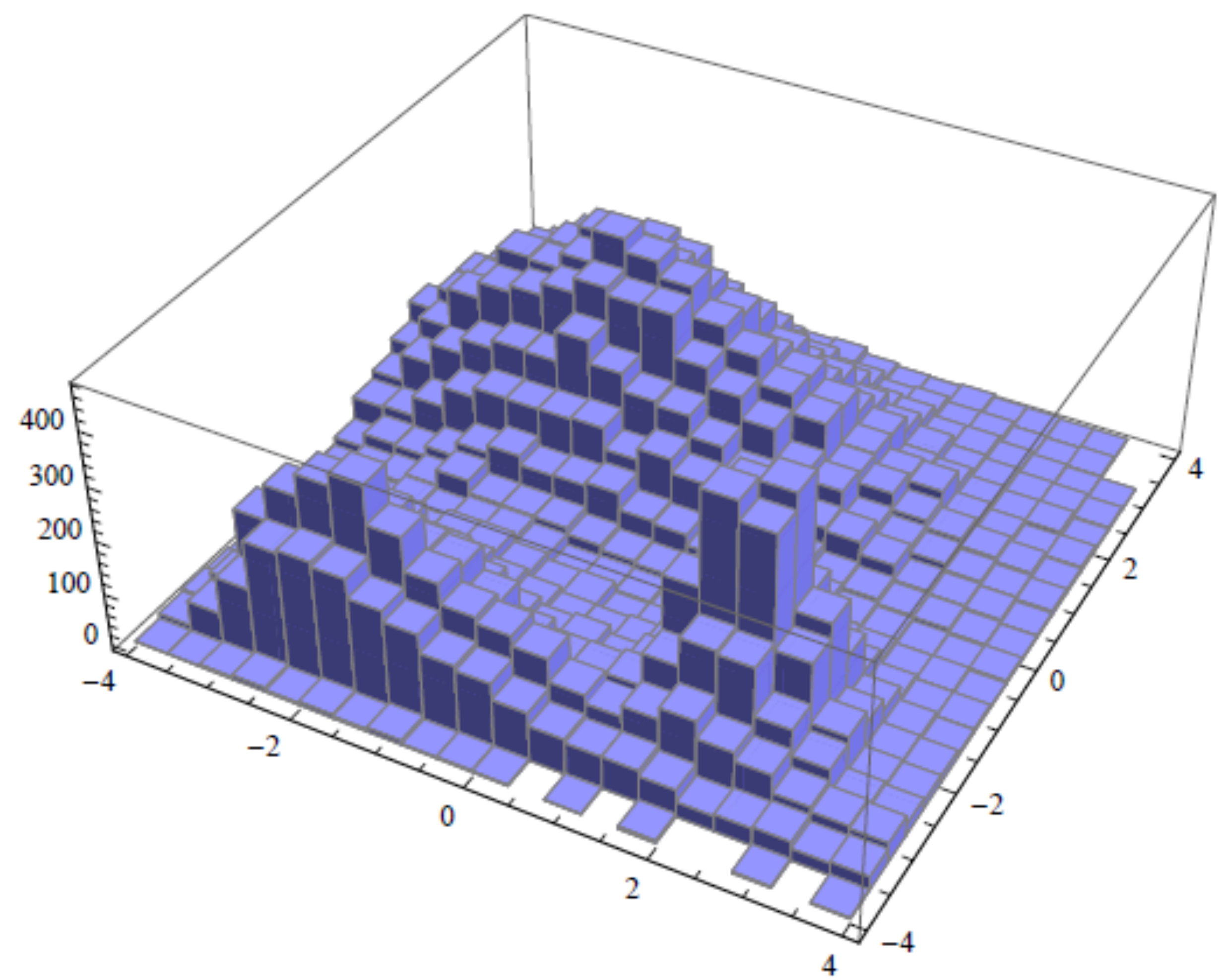
In[4]:=
```
Transpose[data]
```

Out[4]=
```
{{202, 15, 58, 28, 0, 10, 13, 40, 11975, PS, DE},
 {530, 15, 63, 33, 0, 9, 13, 40, 10228, EP, CH}}
```

In[5]:=
```
Import[file, {"ROOT", "TTreeData", "T", {"Service", "Division", "Children"}},
 "Range" → {11, 15}]
```

Out[5]=
```
{{29, 31, 29, 25, 26}, {PS, PS, PS, PS, PS}, {0, 2, 0, 0, 1}}
```

```
In[5]:= Import["cernstaff.root", {"ROOT", "Keys"}]

Out[5]= {{T, CERN 1988 staff data, TTree}}

In[6]:= Import["demo.root", "ROOT"]

Out[6]= {{h0, histo nr:0, TH1F}, {h1, histo nr:1, TH1F}, {h2, histo nr:2, TH1F},
         {h3, histo nr:3, TH1F}, {h4, histo nr:4, TH1F}, {h5, histo nr:5, TH1F},
         {h6, histo nr:6, TH1F}, {h7, histo nr:7, TH1F}, {h8, histo nr:8, TH1F},
         {h9, histo nr:9, TH1F}, {h10, histo nr:10, TH1F}, {h11, histo nr:11, TH1F},
         {h12, histo nr:12, TH1F}, {h13, histo nr:13, TH1F}, {h14, histo nr:14, TH1F}}

In[7]:= Import["th2f.root", "ROOT"]

Out[7]= {{h2, xygaus + xygaus(5) + xylandau(10), TH2F}}

In[35]:=
(* Import the histogram directly as a Graphics *)
graphics3D = Import["th2f.root", {"ROOT", "TH2FGraphics", "h2"}]

Out[35]=
```



In[32]:=

```
(* This imports the histogram data of a given TH2F object. *)
data = Import["th2f.root", {"ROOT", "TH2FData", "h2"}];
```

The data is of the form:
{ {x1, Δx1, count1, error1}, {x2, Δx2, count2, error2}, ... }

In[33]:=

```
(* show first 5 entries in a grid *)
head = {"x", "Δx", "y", "Δy", "count", "Δcount"};
Grid[Join[{head}, Take[data, 5]], Frame → All]
```

Out[34]=

x	Δx	y	Δy	count	Δcount
-4.	0.4	-4.	0.4	1.	1.
-4.	0.4	-3.6	0.4	11.	3.31662
-4.	0.4	-3.2	0.4	16.	4.
-4.	0.4	-2.8	0.4	12.	3.4641
-4.	0.4	-2.4	0.4	9.	3.

a nice synopsis by Ken(from paper outline):

Strength of ROOT
- 20+ years of tools specialized for high-energy physics*
- C++ interpreter offers interactivity
- efficient-handling of large datasets, out-of-core file i/o

Strength of Mathematica
- symbolic manipulation
- interactivity and ease of use
  -- new CDF technology allows one to share dynamic/interactive results!
- generalized data structure allows flexible, powerful manipulation of data
- everything-in-one-package: unified framework for all fields of studies

What's in the paper/package
- uses Mathematica modular Import/Export framework (New-In-8!): clean, easy-to-use interface
- open-source, easily extensible
- MathLink-based converter allows one to interface ROOT libraries with Mathematica

The specifics of the package
- retrieves ROOT files tree and map structures
- out-of-core import of n-tuple branches and leaves
- automatic handling of data types: simply evaluate
  Import[file, {"ROOT", tree, branch}]
  without specifying data type!
- import of histogram data to take advantage Mathematica's visualization tools

**many colleagues from my generation consider the "MINUIT" fitting package key to their attachment to ROOT*

Ken's synopsis (cont.)

Working with large (10s of GBs) datasets
- one can emulate a typical ROOT work flow
 -- parse tree/map structure of data to find relevant dataset
 -- import only the relevant datasets for analysis
 -- interactivity of Mathematica vs C++ interpreter
- but much easier in Mathematica: no need to set up tree, data structure, free resource
 -- we do that for you under the hood!
- out-of-core features tested against a 5GB ATLAS file

Large datasets are the direction in the future
- data is everywhere, not just high energy physics, and the amount of data is exploding!
- Wolfram is a leader in innovation of how one thinks about data*

**Perhaps a good opportunity for interaction with the HEP community at: http://www.wolframdatasummit.org/2012/)*

Comments about HEP context

- a good update is http://www.chep2012.org/ held last month in NYC (see ie Joe Incandela's talk)

- A significant player in HBOOK->PAW->ROOT evolution was the Physics Analysis Software Support and Maintenance Requirements Group (http://www.fnal.gov/projects/runii/passuma/reqs-only.htm) which was considering the options, including commercial ones, for supporting Run II at Fermilab in the 1990's. Commercial packages didn't seem ready, at the time, to handle complex HEP data structures.

- A very large fraction of the LHC experiment effort goes into software development. ATLAS offline framework is the result of 10 years of work by ~600 physicists.

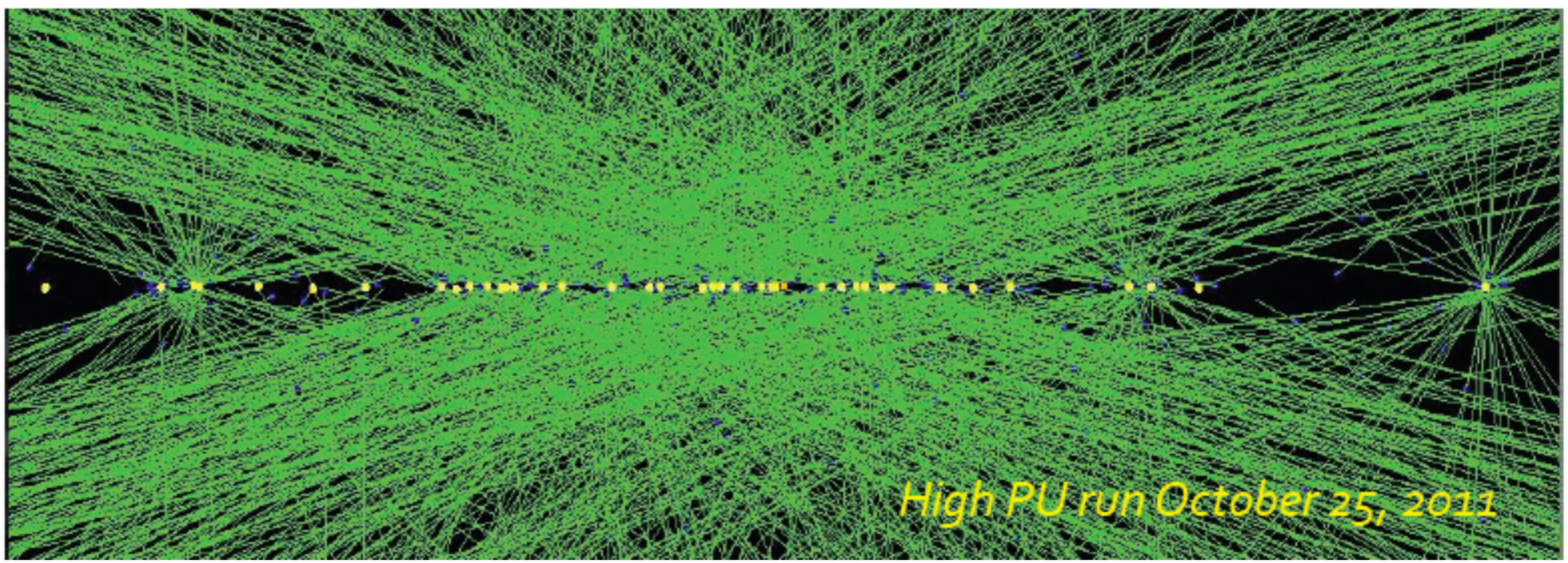
- Challenges in developing professional quality code in such a distributed organization are common to other fields of scientific research also. See:

"Computational science: ...Error…why scientific programming does not compute".Nature 467, 775-777 (2010)Zeeya Merali - particularly comments by David Rousseau, software coordinator for ATLAS.

my personal take on *Mathematica*

I first considered adopting it for computing in 2006 when my co-PI on an LDRD for long baseline neutrino experiment design, Milind Diwan, was doing all simulations in Mathematica.

Following year I started work on tools for dealing with pileup at full LHC intensity:



Above is a reconstruction of tracks in CMS from 1 LHC bunch crossing. Everything happens in <0.5 nanoseconds! Typically there are ~30 collision events in a frame this year!

in :"On the Correlation of Subevents in the ATLAS and CMS/Totem Experiments",S. White- http://adsabs.harvard.edu/abs/2007arXiv0707.1500W
All modeling (starting from the LHC design book) was done in Mathematica

in the past 5 years I've written a series of papers on physics modeling for LHC, RHIC and a future electron-Nucleus Collider, often with Mark Strikman, a high energy theorist, -all using Mathematica
-I've had mixed reaction to including the Mathematica code in these papers.
-also limited success in starting these papers from Latex exports from Mathematica

in 2010 Vitaly Yakimenko and I proposed a new secondary electron beam design based on Rutherford/Hofstadter scattering:
"Energy Calibration of Underground Neutrino Detectors using a 100 MeV electron accelerator"http://adsabs.harvard.edu/abs/2010arXiv1004.3068W
This originated from recreational activity in connection with outreach for the Rutherford Centennial. Really enabled by the ease of doing things in Mathematica. It has turned out to be very useful for developing fast timing detectors. The "single electron project".

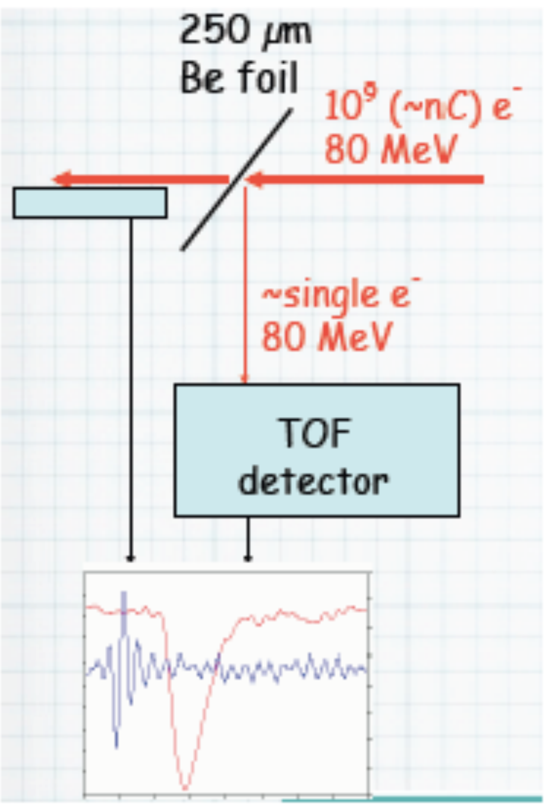
Question:How many electrons are scattered into a 1 cm^2 detector 30 cm from a 0.5mm Be target at 90° from a 10^9 e^- primary beam?
Answer: 1 !!

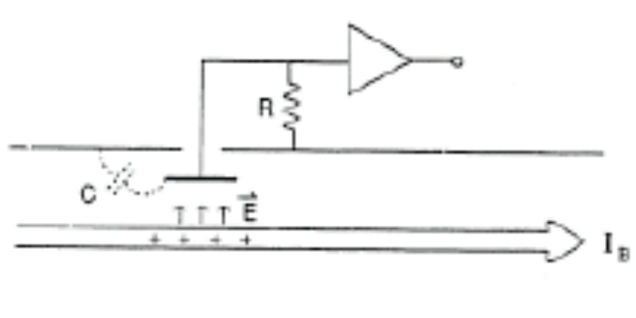
Testbeams used to characterize APD based timing detector

1. Single electron project at ATF
2. PSI (~200 MeV muons and electrons)
3. Frascati BTF <500 MeV electrons, tertiary beam from DAFNE Linac

5. **Energy Calibration of Underground Neutrino Detectors using a 100 MeV electron accelerator** / White, Sebastian ; Yakimenko, Vitaly
An electron accelerator in the 100 MeV range, similar to the one used at BNL's Accelerator test Facility, for example, would have some advantages as a calibration tool for Argon neutrino detectors. [...]
arXiv:1004.3068. - 2010.

rates calculated based on Hofstadter's data



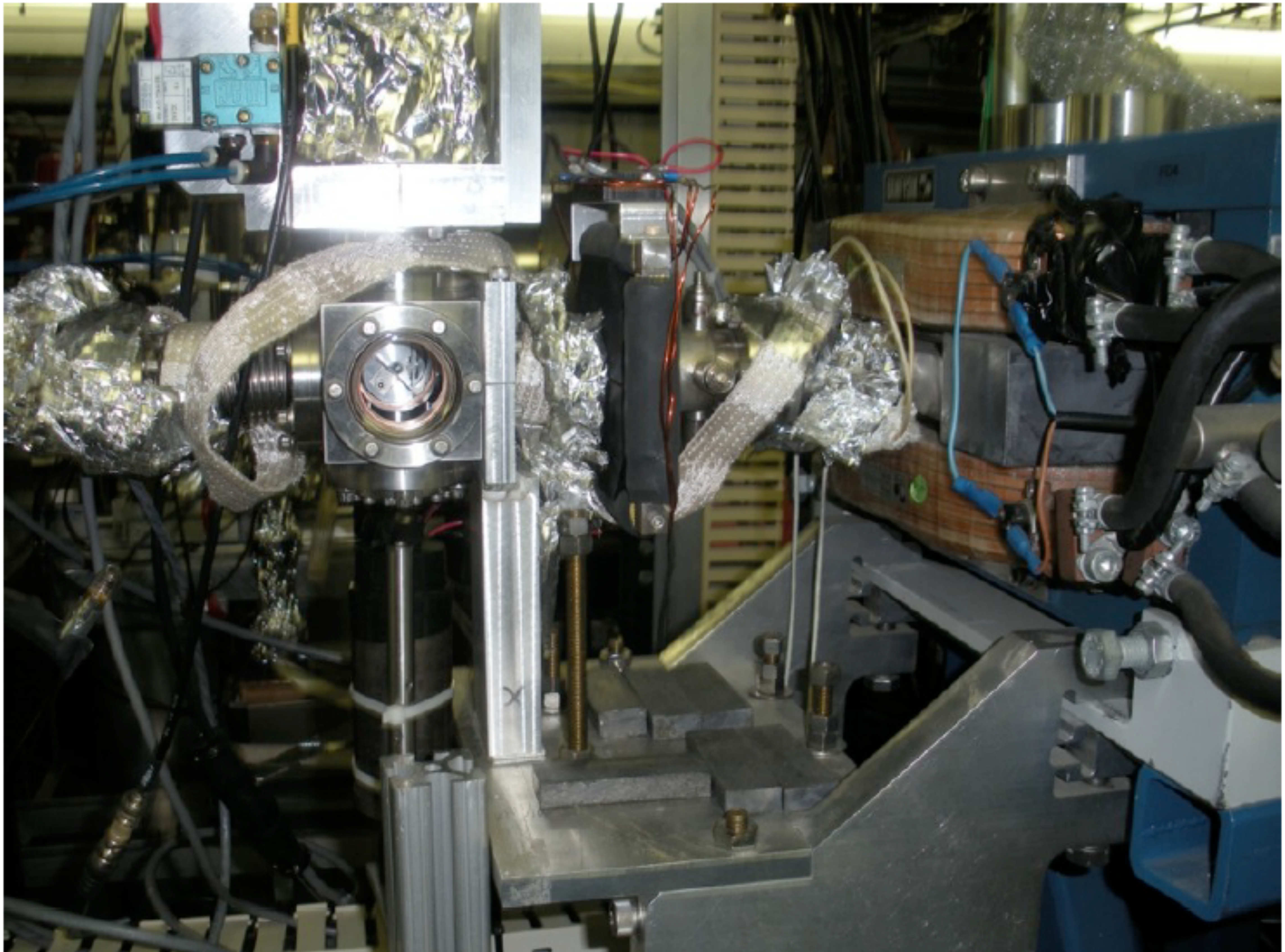


- a unique feature of ATF beam is 3 picosec bunch length(streak camera)
- could this be exploited to evaluate fast timing detectors?
- common technique for secondary beam design is successive dispersion and collimation
- this requires real estate

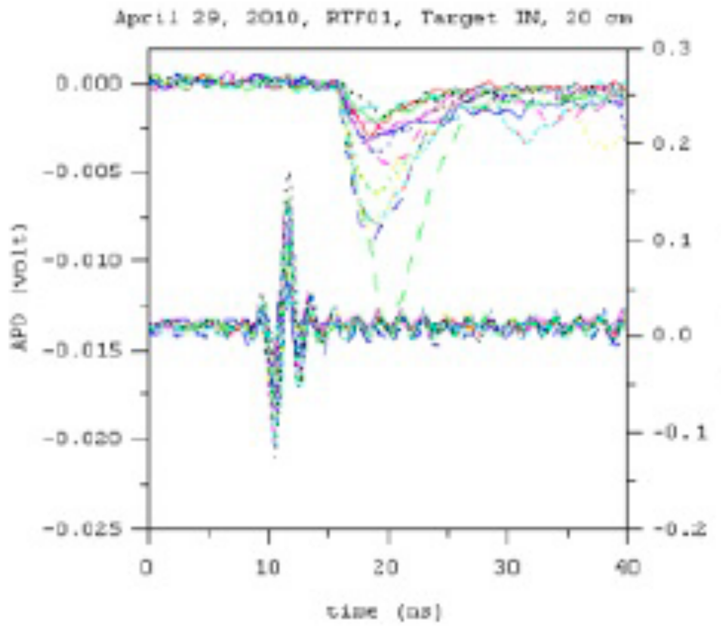
Vitaly

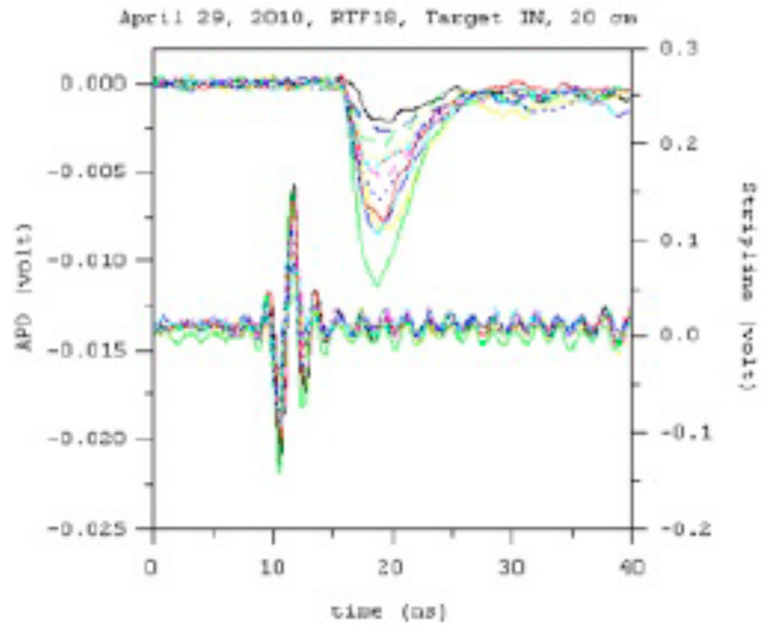
Kirk, Thomas, Misha

the beamline

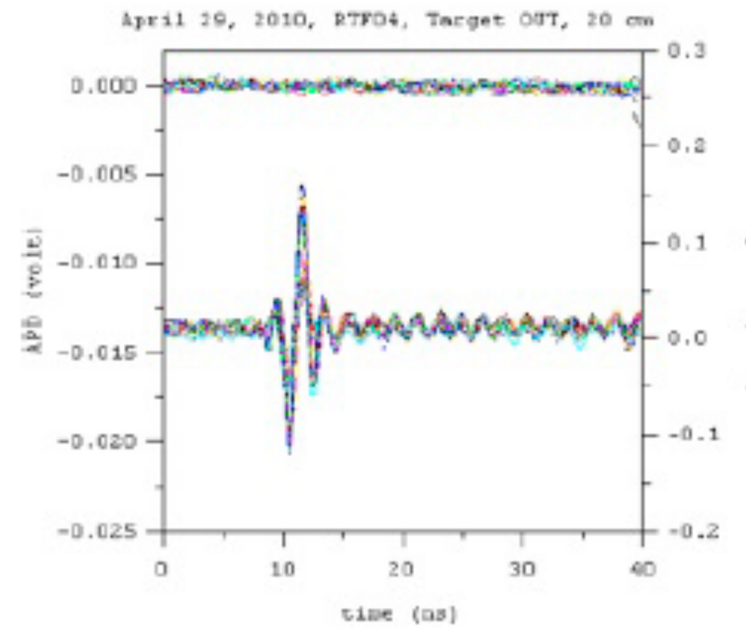


Target IN 20 cm

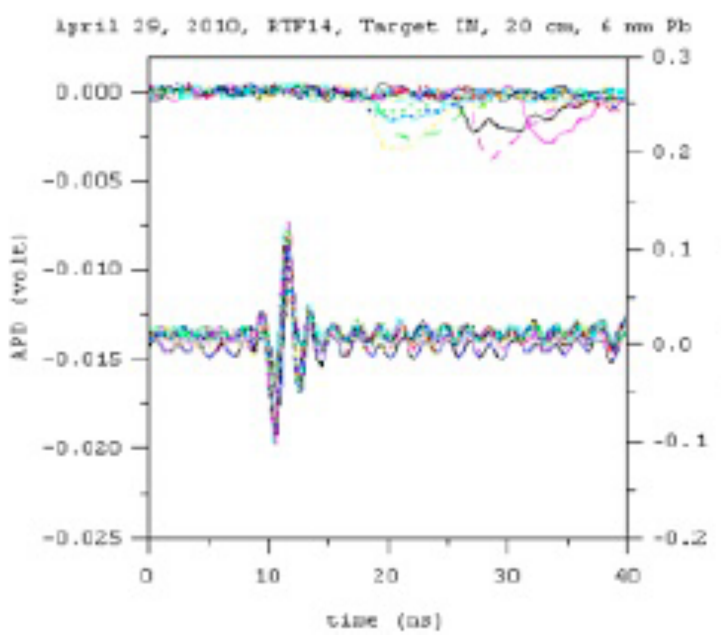


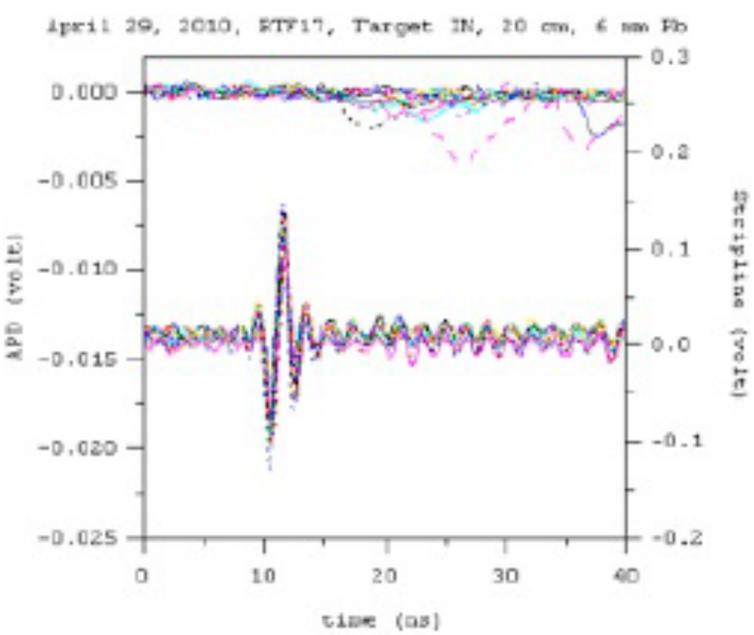


Target OUT 20 cm

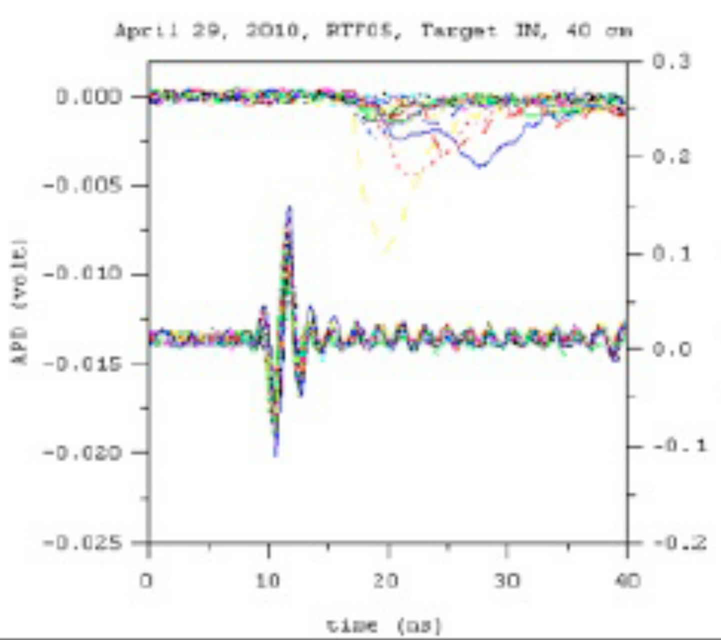


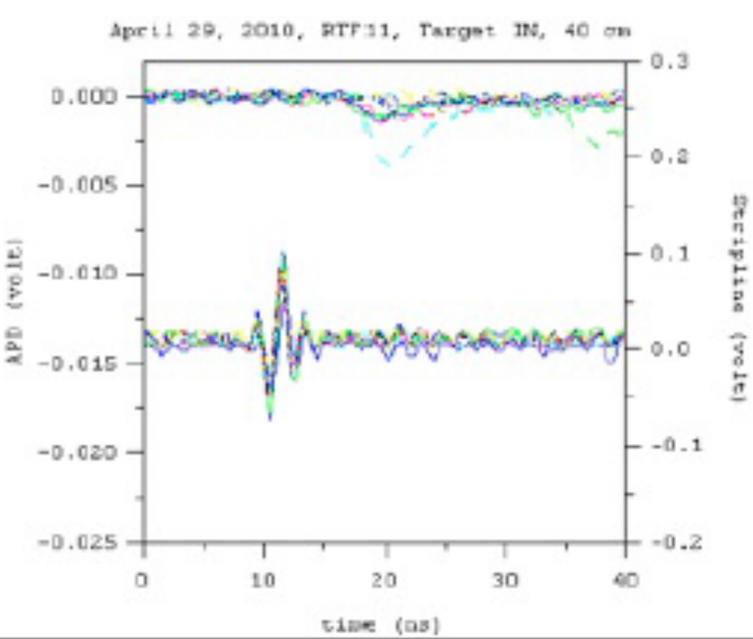
Target IN 20 cm, 6 mm Pb





Target IN 40 cm

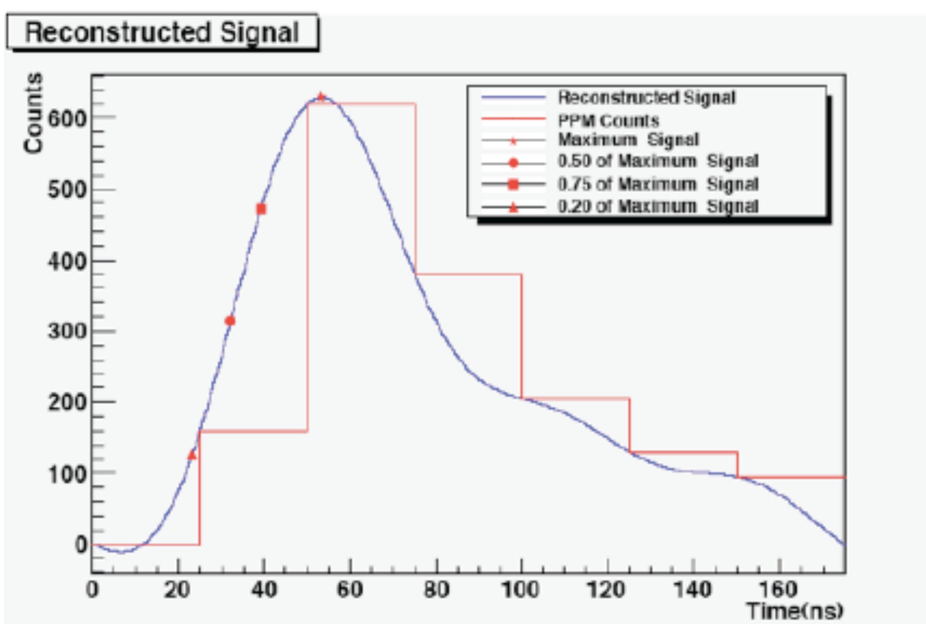




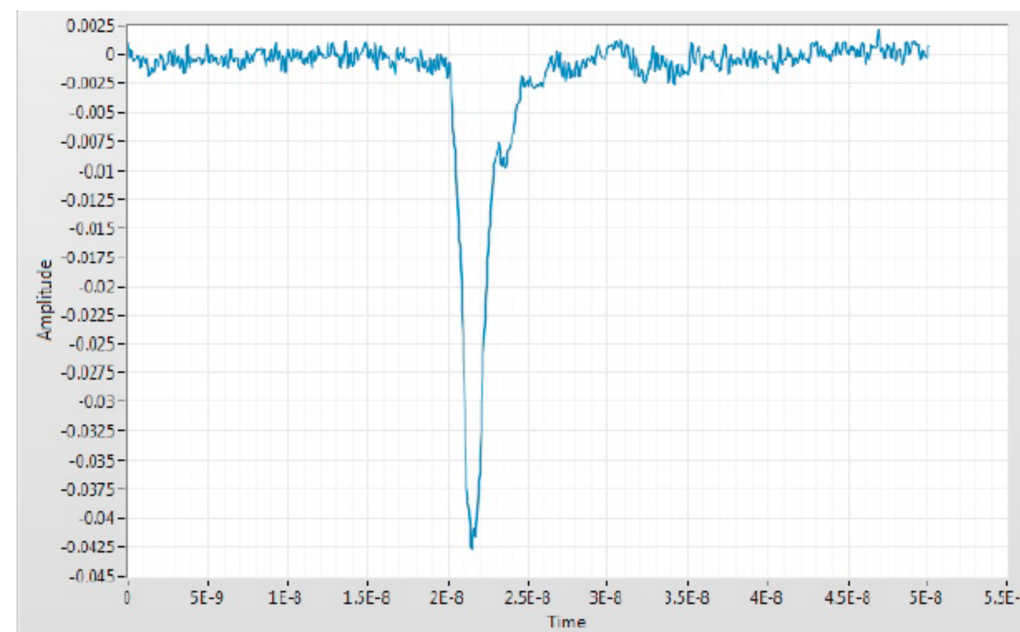
In summer 2010 the ATLAS Zero Degree Calorimeter(ZDC) came on line with 7 TeV proton-proton collisions. I developed ZDC reconstruction (from sparsely sampled waveforms) using Mathematica in collaboration with Soumya Mohapatra and Andrei Poblaguev. This was enabled by Mathematica's interactive tools and availability of ie, the "sinc" function. see:"Very Forward Calorimetry at the LHC - Recent results from ATLAS", S. White http://library.wolfram.com/infocenter/Articles/7716/

the above web-site also contains notebooks on LHC modeling and signal reconstruction.

With ZDC construction completed we have returned to developing technologies to deal with pileup at LHC.



->



ZDC:~0.1 nsec time resolution

Deep depleted APD used as charged particle detector:
<10 picosecond time resolution

(analysis of 40 GSa/s scope data)

Outlook

•“Mathematica with ROOT” was released at a very busy time for LHC experiments.
•Many hits on the web site, many people know about it (particular interest among engineers involved in the LHC experiments).
•Younger physicists are deep in Higgs search and mostly committed to mainstream tools.
•Senior physicists, who often don’t have time to commit to learning full-blown tools adopted by experiments, would appreciate simpler framework to work with data. This would enable better interaction with students and postdocs.
•Many physicists spend most of their time working with simple n-tuples in a simpler framework- often with tools that are still around from when they were grad students. Mathematica easily replaces these.
•“Mathematica with ROOT” should allow the same ease of use with the full data structure.
•With the upcoming Long Shutdown of the LHC (Spring 2013) there is an opportunity to re-visit dealing with inefficiencies in the current approach to supporting computing at the LHC. There is an initiative to explore a common framework for analysis work-flow and this might provide also incentive for collaborating with developers of professional scientific software:

“ laboratories, and country contributions. Constrained budgets and completing external projects are reducing the effort available for development and sustainable operations. CMS and ATLAS need to evaluate ways to reduce the needs while maintaining a strong program. A common infrastructure to support analysis is a step in the direction of reducing development and maintenance effort and thereby improving the overall sustainability of the systems.

perhaps this crisis is an opportunity