# Gentoo Prefix for KamLAND

June 18, 2014

# 1 Abstract

In this KamLAND technical note, we are going to discuss about data analysis tools newly available on Gentoo Prefix at abby/count computing cluster in RCNS.

The tools are Python centric instead of C/C++.

# 2 Gentoo Prefix

Gentoo Prefix [1] is variant of Gentoo meta-distribution which installs under normal user privilege a full Gentoo in a directory offset of a Unix-like OS, such as HPUX, AIX, GNU/Linux, Interix, Mac OS X, Solaris and Cygwin.

Gentoo is a source Linux distribution [2]. "source" in the sense that usually you compile everything out of source code. Gentoo has a package manager called *portage*, written in Python together with bash, containing instructions to build the packages. For source code is naturally available to user, it is more straightforward to tune software, as often needed in scientific research.

## 2.1 Try it out on abby

abby is a set of frontend servers for computational cluster in RCNS. If you already have an account on abby, you can try Gentoo Prefix out at once. In this note, we are going to use *offline* account for demonstration. Be sure to substitute with your own account.

On abby:

---

[1]http://prefix.gentoo.org
[2]http://www.gentoo.org

```
$ ~benda/gento/startprefix
Entering Gentoo Prefix /gpfshome/home/benda/gnto
[...]
*
* Welcome to Abby01
*
```

As you can see, Gentoo Prefix is installed by user "benda", as normal user
without root privilege. It is shared among abby/count by the IBM GPFS
cluster file system in RCNS, so that analysis tools developed in Gentoo
Prefix can run on LSF distributed computation platform.

On abby, the administrators have added many dirty tricks with LD_LIBRARY_PATH,
which is considered harmful [3]. It interferes with libraries in Gentoo Prefix,
so we unset it (e.g. on bash):

```
$ unset LD_LIBRARY_PATH
```

Now you are open to the world of Gentoo and newest software on this
planet.

# 3   A survey of bleeding edge tools

## 3.1   gcc-4.7

At the time of writing (Apr 2013), we provide gcc-4.7 with Gentoo Prefix.
The version of default gcc, however, is 4.1.2, released in 2007 [4].

New versions of gcc has optimization for newer CPUs. The Xeon X5650
on abby has Intel Corei7 features, most significantly sse4_1 and sse4_2
instruction sets which could accelerate vector manipulation. gcc-4.7 could
make use of these new features.

Software in Gentoo Prefix has been compiled with C(XX)FLAGS "-O2
-march=corei7".

In Gentoo Prefix, we have gcc-4.7 ready to use:

```
$ g++ --version
g++ (Gentoo 4.7.2 p1.3, pie-0.5.5) 4.7.2
Copyright (C) 2012 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

---

[3]When should I set LD$_{LIBRARYPATH}$?    http://linuxmafia.com/faq/Admin/ld-lib-
path.html

[4]GCC 4.1 Release Series http://gcc.gnu.org/gcc-4.1/

## 3.2 emacs-24

Emacs is very popular in RCNS. The version of emacs on abby, however, is 21.4, released in 2005 [5]. There are many new useful features in the past 8 years of emacs, most notably mature CJK (Chinese Japanese Korean) support. Emacs24 is available in Gentoo Prefix (1).



Welcome to GNU Emacs, one component of the GNU/Linux operating system.

Emacs Tutorial          Learn basic keystroke commands
Emacs Guided Tour       Overview of Emacs features at gnu.org
View Emacs Manual       View the Emacs manual using Info
Absence of Warranty     GNU Emacs comes with *ABSOLUTELY NO WARRANTY*
Copying Conditions      Conditions for redistributing and changing Emacs
Ordering Manuals        Purchasing printed copies of manuals

To start...    Open a File    Open Home Directory    Customize Startup
To quit a partially entered command, type Control-g.

This is GNU Emacs 24.3.1 (x86_64-pc-linux-gnu)
 of 2013-04-02 on abby01
Copyright (C) 2013 Free Software Foundation, Inc.

U:%%-  *GNU Emacs*      All L17     (Fundamental)
Shell command: import emacs24.png

Figure 1: emacs24 welcome screen

## 3.3 perl-5.16

Perl is a scripting computer language that is famous for one liners, and often quoted as Swiss army knife for *ix.

perl-5.8.8 on abby is released in 2006 [6]. When you want to write sophisticated file handling tools in KamLAND, make sure to check out perl-5.16 for newest perl features for convenience.

---

[5]Emacs 21.4 released http://lists.gnu.org/archive/html/info-gnu-emacs/2005-02/msg00000.html

[6]Perl 5.8.8 Release Announcement http://dev.perl.org/perl5/news/2006/perl-5.8.8.html

## 3.4 root-5.34

The newest version of CERN ROOT is available:

```
$ root

  *********************************************
  *                                           *
  *        W E L C O M E  to  R O O T         *
  *                                           *
  *   Version   5.34/05  14 February 2013     *
  *                                           *
  *  You are welcome to visit our Web site    *
  *          http://root.cern.ch              *
  *                                           *
  *********************************************

ROOT 5.34/05 (tags/v5-34-05@48582, Feb 14 2013, 23:27:57 on linuxx8664gcc)

CINT/ROOT C/C++ Interpreter version 5.18.00, July 2, 2010
Type ? for help. Commands must be C++ statements.
Enclose multiple statements between { }.
root [0]
```

It is compiled with Python binding:

```
$ python -c "from ROOT import TFile; print type(TFile)"
<class 'ROOT.TFile_meta'>
```

## 4  Python

Python is a script computer language that is easy for prototyping and fast development. It is also famous for its cooperativeness with other languages. C++ classes and functions are easily exposed to and called by Python. As a consequence, it becomes popular in data analysis as a "glue" language to integrate exotic formats of data and different kinds of programs.

Traditionally data analysis tools in KamLAND are written in C/C++ and CERN ROOT. Python has nice interfaces to reuse those tools and libraries.

Python in Gentoo Prefix is 2.7.3, while the default version is 2.4.3.

4

Python wrappers of KamLAND tools are generated with Py++, which in turn uses boost-python. We can call present tools from Python without compilation. This is very useful for exploring and peeking around data. The following examples will give you an intuitive idea.

## 4.1 SF

SF stand for serial file, "serial" in that I has an interface to read a chain of files. It is a format used to save waveform and TQ data.

For example, we want to get NSumMax of event with timestamp 9006539845 from run 10000, which is a (surprise!) pile-up. This can be done interactively.

```
$ ipython
[...]

In [1]: import sfc

In [2]: rtq = sfc.rtq(10000)

In [3]: tq = rtq.searchTs(9006539845)

In [4]: print tq.hd.nsMax
72
```

## 4.2 KVF

KVF stands for "KamLAND Vector File" used in RCNS, which records parameters such as vertex and energy of every event. It is the format used after vertex and energy fitters.

Lets get the A2 energy of the previous event (10000:9006539845).

```
In [5]: import vf

In [6]: ci = vf.kvr(10000)

In [7]: print ci.searchTs(9006539845).EnergyA2
0.266774564981
```

## 4.3 KAT

Kat stands for "KamLAND Analysis Tools" used in RCNS, which hold all the main algorithms in KamLAND data analysis. It is an encyclopedia of KamLAND tricks (therefore hard to understand).

Lets see if we could reproduce the result obtained from KVF.

```
In [9]: import kt

In [10]: kt.prepare(10000)

In [11]: tqo = kt.tqh()

In [12]: tqo.fill(tq)

In [13]: rst = kt.av2(tqo)

In [14]: print rst[0].engA2
0.211471825838
```

Oops! We cannot reproduce it! Relax, it is common in KamLAND, we often cannot reproduce our previous result because someone else may have updated something without realizing it will affect a unexpected library. If the difference counts, trace the origin. If not, forget about it.

To conclude, let me show a snip of code making fanny plots with matplotlib. Note that details are omitted, so this is not trivially reproducible. (2)

```
plt.hist(spp, bins=np.arange(-600, 100, 10), alpha=0.5, label="cluster %d: %.2fMeV" % (c

dist = plt.hist(tq.tme, bins=np.arange(-600, 100, 10), histtype='step', label="raw")
plt.vlines(splt, 0, max(dist[0]))
plt.legend(loc='upper left')
plt.title('twin event fitting %d:%d' % (rN, ev.hd.ts))
plt.xlabel('time(ns)')
plt.ylabel('count')
pp.savefig()
```

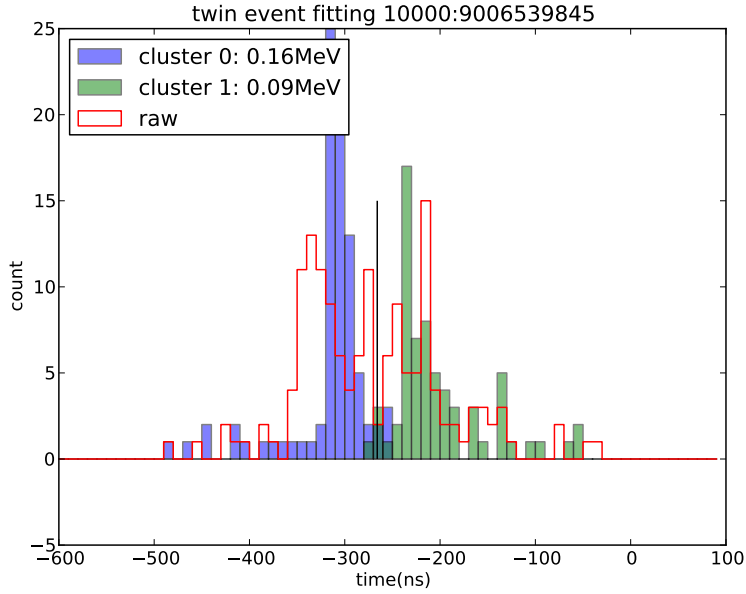# 5 Future Plan

Geant4 9.6 with KLG4sim and Python binding.

Figure 2: fitting result of pile-up event timestamp 9006539845 of run 10000.

# 6    Security Concerns

When you use Gentoo Prefix, keep in mind that the tools are compiled by the user 'benda' on abby. Be sure not to pass sensitive data or private information to it.

# 7    Concluding Remarks

This technical note gives an alternative way to process KamLAND data without touching KamLAND C/C++ libraries directly. This will save time for physicists to concentrate more on exotic and crazy toys and games.