

MCNEG 2004 at NPL

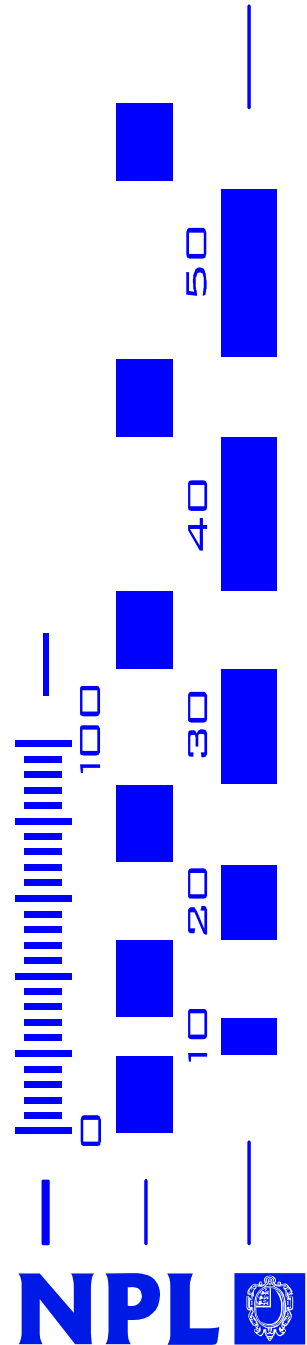
Distributing EGS on the NPL United Devices Grid

Simon Duane

National Physical Laboratory, UK

simon.duane @ npl.co.uk

16th March 2004



What's a grid?

- Like electrical power distribution
 - Producers and consumers linked in a transparent way
 - Don't need to know who burnt the oil (or where)
- Computing grid
 - Producers have cpu / disk / memory
 - Consumers have computational tasks
- Producers = donors / desktop pc owners ... the victims
- Consumers = grid users ... me, Hugo, et al.

Let's get together...

Some pre-history: hardware for MC at NPL

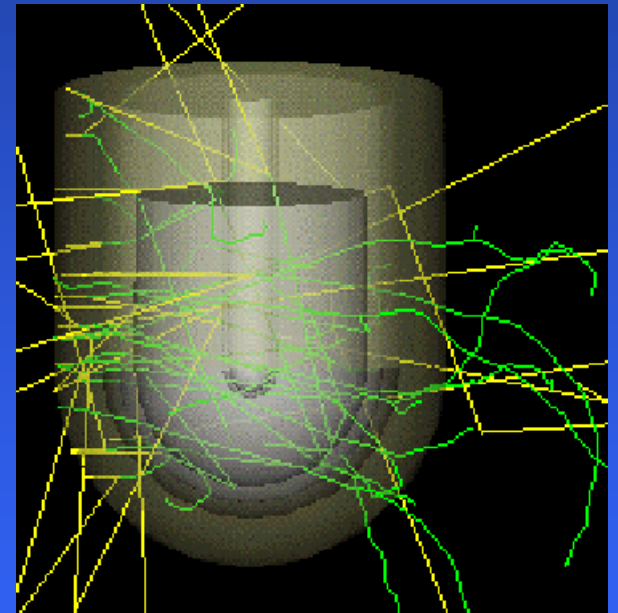
- 1987 – VAX 11/785 (finance)
 - 1988 – microVAX II (dosimetry)
 - 1989 – Meiko Computing Surface (dosimetry)
 - Initially 4 processors, then 28 processors
 - 1990 onwards – PCs (dosimetry)
 - DOS + Pharlap 386 + Lahey F77 (not networked)
 - Win3.1, Win3.11, Win95 (networked)
 - Linux (dual boot desktops)
 - Linux (dedicated)
 - 2002 up to 9 Linux boxes dedicated to MC
 - But not a proper cluster...
- => We know about parallel execution

Parallelizing Monte Carlo simulations

- In principle, no problem:
 - we need billions of histories anyway
- In practice, need to
 - split task at start
 - Independent random sequences – we use Marsaglia-Zaman (easy to generate and label 10^8 sequences)
 - Merge results at end
 - Dose calculation - combine and improve statistics
 - Phase-space generation – concatenate files
- Multiple instances of the executing program are independent (they don't need to talk to one another)

output data: fixed SIZE problems (e.g. dose)

- Can always be made to work:
 - Volume of data io is fixed
 - Increase job duration to make comms/comp large enough



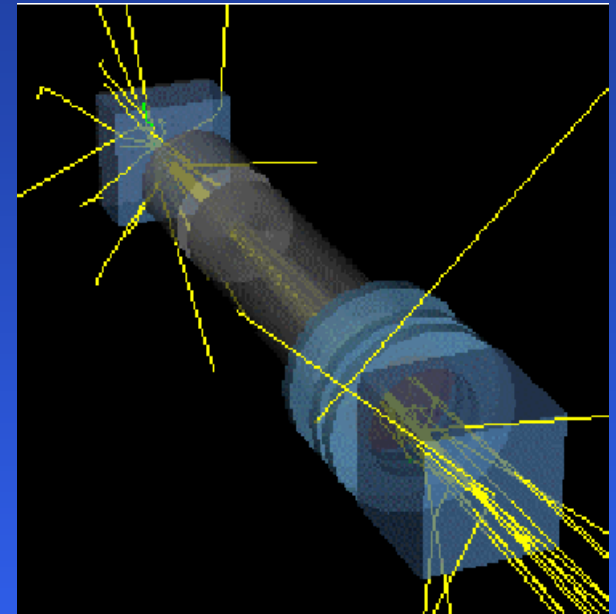
DOSCHAM

output data: fixed RATE problems (e.g. phase space)

- May or may not be worth it:
 - Electron beams – no
 - Photon beams – maybe

(for our linac simulations, anyway)

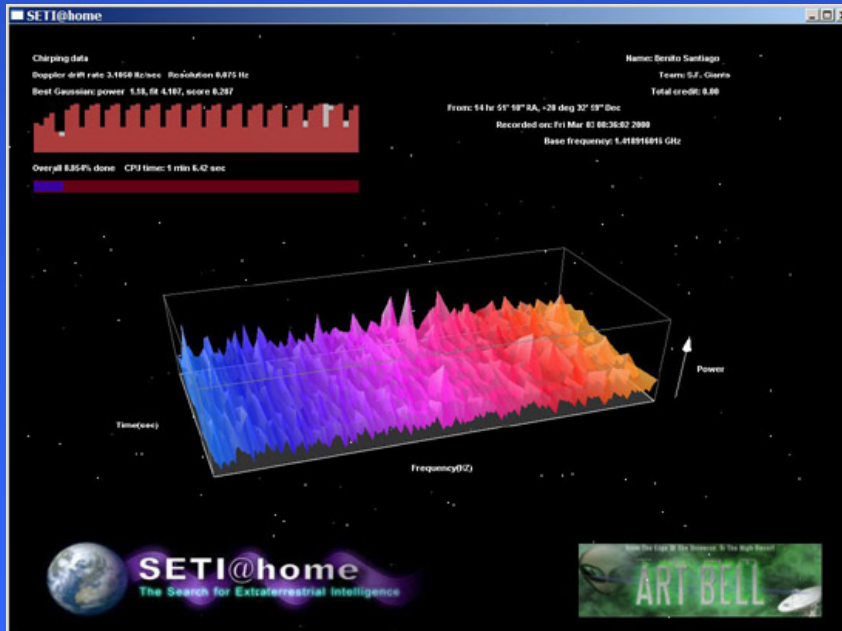
NPLLINAC and DOSCHAM:
EGS4/PRESTA usercodes written by
David Shipley



NPLLINAC

Grid examples

- GIMPS
 - Great Internet Mersenne Prime Search:
- Seti@home
 - You've seen the screen saver:



The screenshot shows a Microsoft Internet Explorer browser window. The title bar reads 'New Scientist - Microsoft Internet Explorer provid...'. The address bar shows 'http://www.newscientist.com/news/news'. The main content area displays the following text:

The World's No.1 Science & Technology News Service

Largest prime number ever is found
15:11 02 December 03
NewScientist.com news service

A 26-year-old graduate student in the US has made mathematical history by discovering the largest known prime number.

The new number is 6,320,430 digits long. It took just over two years to find using a distributed network of more than 200,000 computers.

Michael Shafer a chemical engineering student at Michigan State University used his office computer to contribute spare processing power to the Great Internet Mersenne Prime Search (GIMPS). The project has more than 60,000 volunteers from all over the world taking part.

"I had just finished a meeting with my advisor when I saw the computer had found the new prime," Shafer says. "After a short victory dance, I called up my wife and friends involved with GIMPS to share the great news."

Prime numbers are positive integers that can only be divided by themselves and one. Mersenne primes are an especially rare type of prime that take the form $2^p - 1$, where p is also a prime number. The new number can be represented as $2^{20,996,011} - 1$. It is only the 40th Mersenne prime to have ever been found.

The NPL UD grid – (i)

- The desktop machines run MS Windows NT / 2000 / XP
 - UD also allow linux, or Sun, or AIX, or any mixture...
- Linux server(s)
 - We have two, located centrally.
- Secure
 - Optional encryption – we don't need it.
- Unobtrusive
 - Users unaware their pc being used by someone else

The NPL UD grid – (ii)

- 650 staff, approx 600 desktop PCs
 - Mostly 3 year life, so reasonably current models
 - Mostly idle, most of the time
- All networked, using managed switches
 - 1Gbs (backbone – not sure exactly where)
 - 100 Mbs (new building)
 - 10Mbs (old buildings)
- PCs at NPL are configured with a standard disk image
 - Includes UD Grid agent software (since late 2003)
 - >200 PCs have agent installed

The NPL UD grid – (iii)

- 2 linux servers (could have been one)
 - Filestore (DB2 database)
 - Grid management services (5)
 - Poll server
 - Dispatch server
 - Realm server
 - RPC server
 - File server
- 100 agent licenses (with option for more...)
 - Dispatch server limits number of devices that can be actively running jobs.

(show slide from UD training)

What happens?

- Each device runs agent as service
 - i.e. on boot, independent of user login to Win2k
- Every 2 mins:
 - Agent:
“hello – I’m waiting for something to do”
 - Poll Server:
“thanks – do this” [sends program and data files]
 - or
 - Agent:
“hello – I’m waiting for something to do”
 - Poll Server:
“thanks – there’s nothing on at the moment – come back a bit later”

Setting up and running code on the grid

- Start with fortran source in CygWin (copy from linux?)
- Make all io to current directory (if not already)
- Implement split and merge
- Compile (e.g. g77 in CygWin)
- Copy to another directory
 - The executable, cygwin1.dll, any input data files, the UD loader.exe
- Build a UD program module and data packages
 - persistent and workunit datas
- Run in the testagent (on local PC)
- Upload to UD server and launch job(s)

Grid user interface - options

- Web-based
 - Interactive
 - User-friendly
 - Good for getting to know what's there
- Use XML-RPC or SOAP
 - Programmable – C++, Java, etc
 - Scriptable – perl, python, etc

Web based – login

Management Console: Login - Microsoft Internet Explorer provided by NPL v1.5

File Edit View Favorites Tools Help

Address <https://udserver.npl.co.uk/login.ud> Go Links

GRID MP
PLATFORM

UNITED DEVICES

Management Console

User Authentication

User Name	<input type="text" value="sd1"/>
Password	<input type="password" value="*****"/>
Mode	Simple Interface ?

© 2000-2004 United Devices, Inc. All Rights Reserved. Mon, 15 Mar 2004 17:55:42 UTC

Local intranet

Web based – console

Grid MP platform: Home - Microsoft Internet Explorer provided by NPL v1.5

File Edit View Favorites Tools Help

Address <https://udserver.npl.co.uk/>

GRID MP
PLATFORM

UNITED DEVICES

Start Home User: sd1 | Help | Logout

MANAGEMENT CONSOLE

Welcome Simon Duane, to udserver.npl.co.uk running Grid MP platform v4.0.3106

PLATFORM MANAGEMENT

- Services
 - ▶ Service Dashboard
 - ▶ Manage Services
- Agents
 - ▶ Manage Agent Versions
- Devices
 - ▶ [Manage Device Groups](#)
 - ▶ [Manage All Devices](#)
- Users
 - ▶ Manage User Groups
 - ▶ Manage Users

WORKLOAD MANAGEMENT

- Applications
 - ▶ [Manage Applications](#)
 - Application Creation Wizard
- Programs
 - ▶ [Manage Programs](#)
 - Program Creation Wizard
- Jobs
 - ▶ Go to Job #
 - ▶ [Manage Jobs](#)
 - ▶ [Update Jobs](#)
 - [Job Creation Wizard](#)
- Errors
 - ▶ [List Errors](#)

DATA MANAGEMENT

- Data Sets
 - ▶ [Manage Data Sets](#)
 - [Data Set Creation Wizard](#)
- Reports
 - ▶ View Reports

© 2000-2004 United Devices, Inc. All Rights Reserved. Mon, 15 Mar 2004 17:52:04 UTC

Local intranet

Web based – manage jobs

Jobs - Microsoft Internet Explorer provided by NPL v1.5

File Edit View Favorites Tools Help

Address <https://udserver.npl.co.uk/workload/jobs/jobs.ud>

GRID MP
PLATFORM

UNITED DEVICES

Start [Home](#) > [Workload](#) > [Jobs](#) User: [sd1](#) | [Help](#) | [Logout](#)

Jobs

List of all Jobs

Actions

- [Create Job](#)
- [Job Creation Wizard](#)
- [View Job Distribution by State](#)
- [Update Jobs](#)
- [Delete Jobs](#)

Filter/Search Options

Job ID	Description	Application	Creator	State	Runnable	Workunits (done/total)	Results (succ/unsucc/Errors)	Priority
168	npllinac job	NPLLINAC	sd1	Completed	No	1/1 (100.00%)	1/0/3	10
167	npllinac job	NPLLINAC	sd1	Completed	No	1/1 (100.00%)	1/0/0	10
166	npllinac job	NPLLINAC	sd1	Completed	No	1/1 (100.00%)	1/0/0	10
165	npllinac job	NPLLINAC	sd1	Completed	No	100/100 (100.00%)	100/0/162	10
164	npllinac job	NPLLINAC	sd1	Completed	No	100/100 (100.00%)	100/0/130	10
163	npllinac job	NPLLINAC	sd1	Completed	No	100/100 (100.00%)	100/0/71	10
162	npllinac job	NPLLINAC	sd1	Enabled	Yes	0/0 (0.00%)	0/0/0	10
161	npllinac job	NPLLINAC	sd1	Completed	No	10/10 (100.00%)	10/0/0	10
160	npllinac job	NPLLINAC	sd1	Completed	No	1/1 (100.00%)	1/0/0	10
159	npllinac job	NPLLINAC	sd1	Completed	No	10/10 (100.00%)	10/0/3	10

page: [1](#) of 13
records/page: [10](#)
displaying records: 1 - 10
total records returned: 124

© 2000-2004 United Devices, Inc. All Rights Reserved. Mon, 15 Mar 2004 17:52:59 UTC

Web based – (after select job id)

Job
View details of Job #165

Quicklinks
[Job](#) : [Job Statistics](#) : [Aggregate Job Step Status](#) : [Job Steps](#)

Actions
→ [Create new Job Step](#)
→ [Manage Device Group Targeting](#)
→ [View Devices for Job](#)
→ [Create Data Set local to this Job](#)
→ [Wizard Create Data Set and Datas local to this Job](#)
→ [Edit](#)
→ [Delete](#)

Job

Job ID	165
Description	npllinac job
Application	NPLLINAC
Creator	sd1
State	Completed
Priority	10

Job Statistics

Scheduled Execution Start Time	Not Specified
Last Result Time	2004-03-14 23:04:12 UTC
Last Dispatch Time	2004-03-14 23:00:49 UTC
CPU Time	7 hours 18 minutes 46 seconds
Scheduled Execution End Time	Not Specified
Created	2004-03-14 22:46:01 UTC
Last Modified	2004-03-14 23:05:35 UTC

Web based – (after select job step)

Job Step - Microsoft Internet Explorer provided by NPL v1.5

Address: https://udserver.npl.co.uk/workload/jobs/job_step.ud?guid=D453AD04-1DF0-4030-9222-57DE

GRID MP PLATFORM

UNITED DEVICES™

Start [Home](#) > [Workload](#) > [Jobs](#) > [Job Step](#) User: [sd1](#) | [Help](#) | [Logout](#)

Job Step

View details of Job Step

Quicklinks
[Job Step](#) : [Job Step Status](#) : [Workunits](#)

Actions

- [Create Data Set local to this Job Step](#)
- [Wizard Create Data Set and Datas local to this Job Step](#)
- [View Data Sets used by Workunit generation](#)
- [Create Workunits From Data Sets](#)
- [Edit](#)
- [Delete](#)

Job Step

Job ID	165
Program	NPLLINAC
Scheduling Type	Shared Scheduled
State	Completed
Results per Workunit	1
Errors per Workunit	10
Concurrent Dispatches per Workunit	5
Workunit CPU Timeout	None
Workunit Wall Clock Timeout	1 hour
Workunit Generation Done	Yes

Job Step Status

Recorded Time	2004-03-15 17:57:47 UTC
---------------	---

Local intranet

Web based – browsing results

Job Step - Microsoft Internet Explorer provided by NPL v1.5

Address https://udserver.npl.co.uk/workload/jobs/job_step.ud?guid=D453AD04-1DF0-4030-9222-57DE

Error Count: 162
Created: 2004-03-14 22:46:02 UTC
Last Modified: 2004-03-14 23:05:17 UTC

Workunits
+ Filter/Search Options

Workunit Index	State	Successful Results	Unsuccessful Results	Errors
1	Completed	1	0	5
2	Completed	1	0	1
3	Completed	1	0	0
4	Completed	1	0	1
5	Completed	1	0	5
6	Completed	1	0	3
7	Completed	1	0	0
8	Completed	1	0	0
9	Completed	1	0	1
10	Completed	1	0	1
11	Completed	1	0	1
12	Completed	1	0	0
13	Completed	1	0	0
14	Completed	1	0	0
15	Completed	1	0	2
16	Completed	1	0	1
17	Completed	1	0	0
18	Completed	1	0	1
19	Completed	1	0	0
20	Completed	1	0	4

page: 1 of 5
records/page: 20
displaying records: 1 - 20
total records returned: 100

© 2000-2004 United Devices, Inc. All Rights Reserved. Mon, 15 Mar 2004 17:58:30 UTC

and so on

Or use scripts – e.g. perl

```
Command Prompt
D:\stuff\UD\np1linac\qu10mvm_h>dir *.pl
Volume in drive D is Local Disk
Volume Serial Number is 1C25-98B1

Directory of D:\stuff\UD\np1linac\qu10mvm_h

06/03/2004  22:46                2,788 delete_job.pl
06/03/2004  22:46                1,922 generate_seeds.pl
06/03/2004  22:47                2,229 get_job_complete.pl
06/03/2004  22:47                3,394 retrieve_job_results.pl
06/03/2004  22:47                8,558 send_work_units.pl
           5 File(s)              18,891 bytes
           0 Dir(s)  24,371,159,040 bytes free

D:\stuff\UD\np1linac\qu10mvm_h>
```

```
Command Prompt
D:\stuff\UD\np1linac\qu10mvm_h>get_job_complete.pl
Get Job Complete (1.4) - NPL Distributed Computing Utilities
Usage: perl get_job_complete.pl <user> <password> <job_id>

D:\stuff\UD\np1linac\qu10mvm_h>
```

Perl scripts at NPL adapted by Keith Lawrence from UD examples

scripts – e.g. python

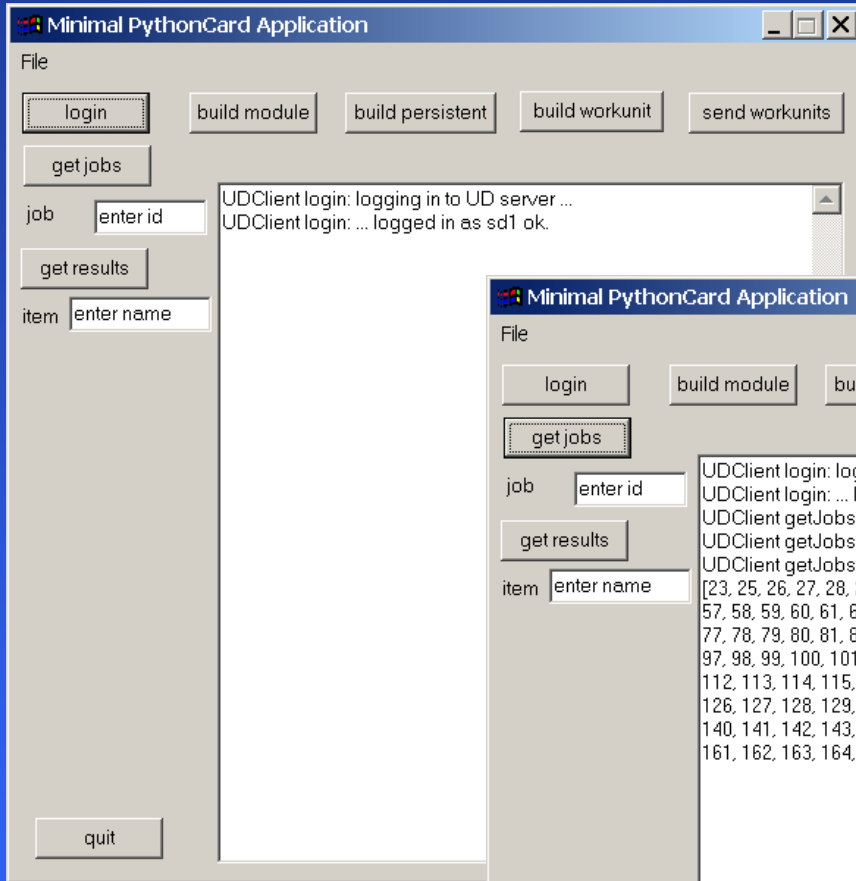
```
UDClient01.py - Code Editor PythonCard Application
File Edit View Format Shell Help
1 #!/usr/bin/python
2
3 """UDClient module - a console mode interface is built in"""
4
5 from SOAPpy import SOAPProxy
6 from urllib import urlencode, urlopen
7 from sets import Set
8 from getpass import getpass
9 from os import system
10 from time import ctime
11
12 class UDProgram:
13     def __init__(self, name):
14         self.name = name
15
16 class UDJob:
17     def __init__(self, appName):
18         self.description = appName + ' job'
19         self.annotation = self.description + ' submitted at ' + ctime()
20         self.application_gid = session.getApplicationGid(appName)
21         self.priority = 10 # lowest priority
22         self.state_id = 1 # enabled
23         self.job_gid = session.server.createJob(session.authKey, self)
24
25 class UDJobStep:
26     def __init__(self, job, progName):
27         self.iob_gid = iob.iob_gid
28
29 File: D:\stuff\UD\ynllnac\UDclient\UDClient01.py | Line: 1 | Column: 1
```

source code

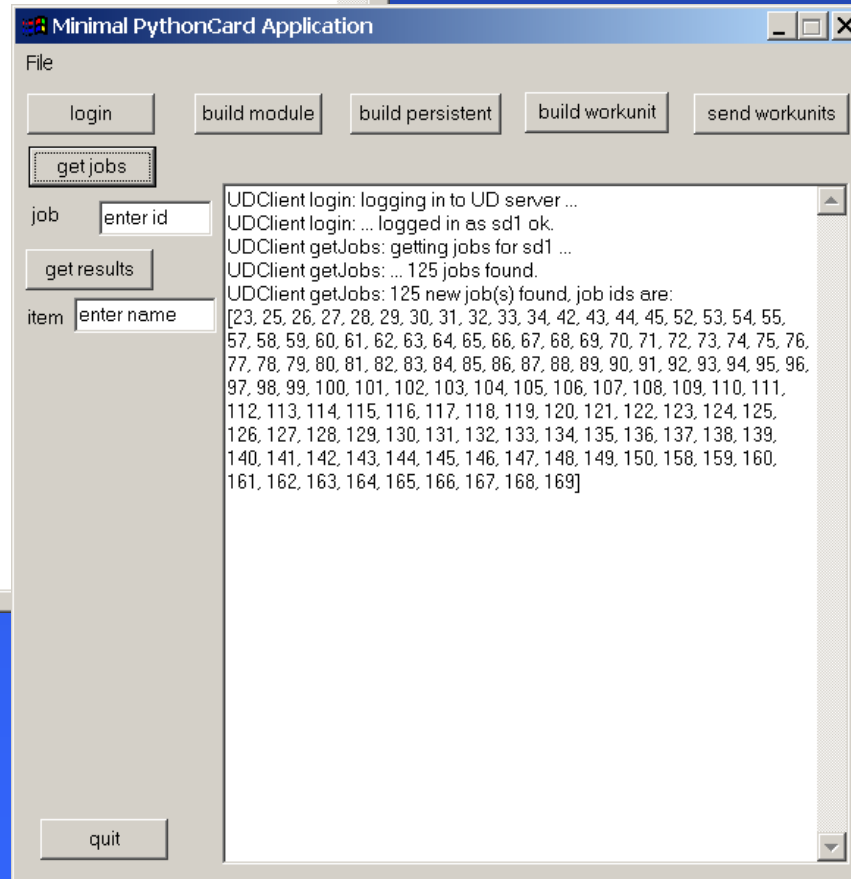
python in use – command line

```
Command Prompt - c:\python23\python
D:\stuff\UD\npllinac\UDClient>c:\python23\python
Python 2.3.2 (#49, Oct  2 2003, 20:02:00) [MSC v.1200 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from UDClient01 import *
>>> s = UDClient()
>>> s.login()
UDClient login: logging in to UD server ...
username: sd1
Password:
UDClient login: ... logged in as sd1 ok.
>>> s.getJobs()
UDClient getJobs: getting jobs for sd1 ...
UDClient getJobs: ... 125 jobs found.
UDClient getJobs: 125 new job(s) found, job ids are:
[23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 42, 43, 44, 45, 52, 53, 54, 55, 57,
 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97,
 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 1
14, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 1
30, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 1
46, 147, 148, 149, 150, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 1
69]
>>>
```

python in use – GUI



GUI



Early results

- PTRAN
 - Proton transport code (Hugo Palmans' talk...)
 - Physics (ion chambers in phantom in proton beam)
 - Jobs submitted Friday evening, ready Sunday lunchtime (equivalent to about 4 months on a desktop PC)
- EGS4/PRESTA – NPLLINAC usercode
 - Phase space generation
 - A benchmarking exercise (so far)
 - Aim to discover the (io) limits of the system

NPLLINAC – performance on a laptop

- 4, 6, 8, 10, 12, 16, 19 MeV electrons onto a
 - Tungsten target
 - With or without Al filter
 - With collimator (makes a beam 11cm diameter at 125cm)
 - Range rejection, brems splitting turned on ...

	4MV heavy filt	19 MV light filt
Histories /sec	4870	1025
Particle yield	0.005	0.23
Data rate	700 byte/sec	6.4 kbyte/sec

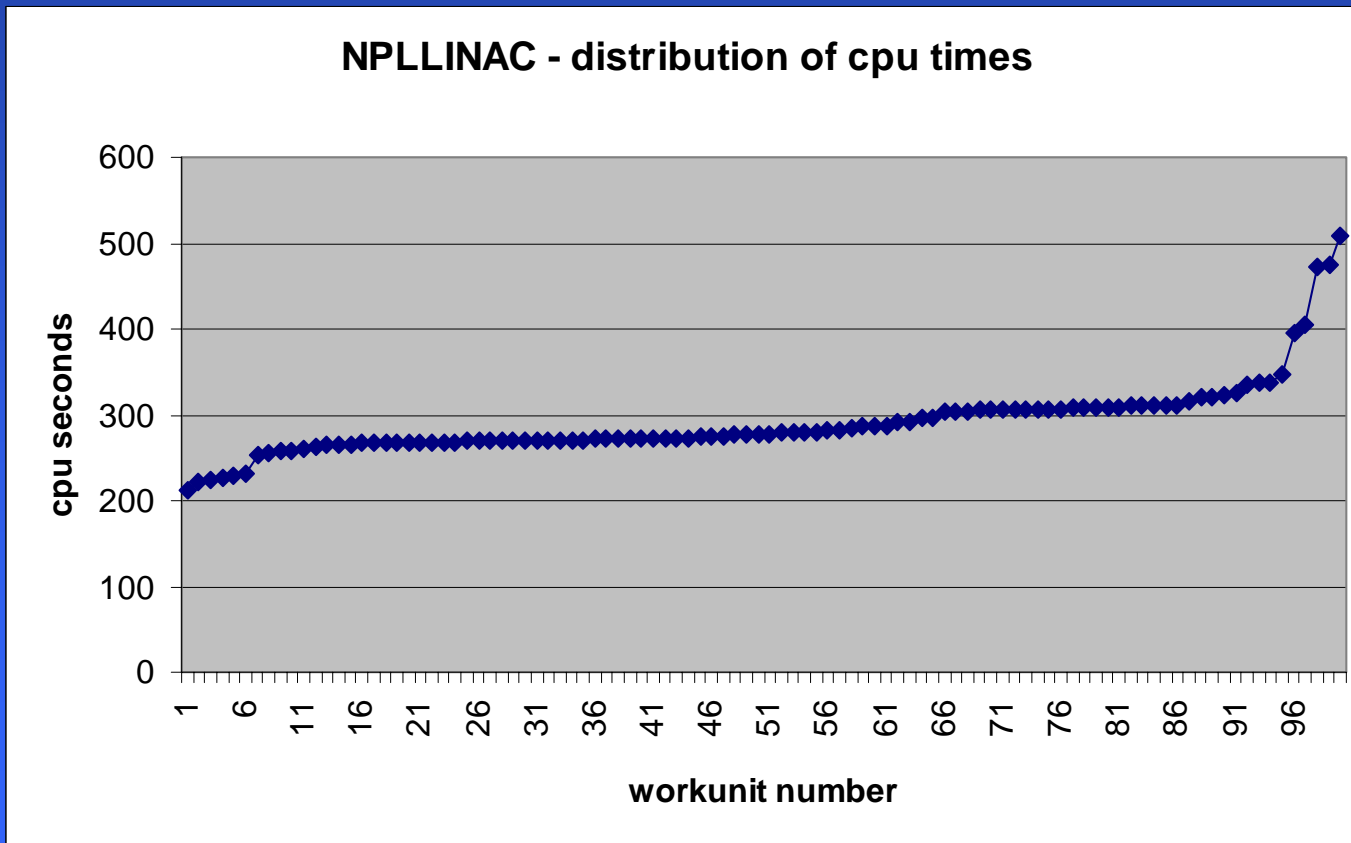
Run on a Pentium M 1.6GHz (g77 in CygWin)

Grid of 100 devices?

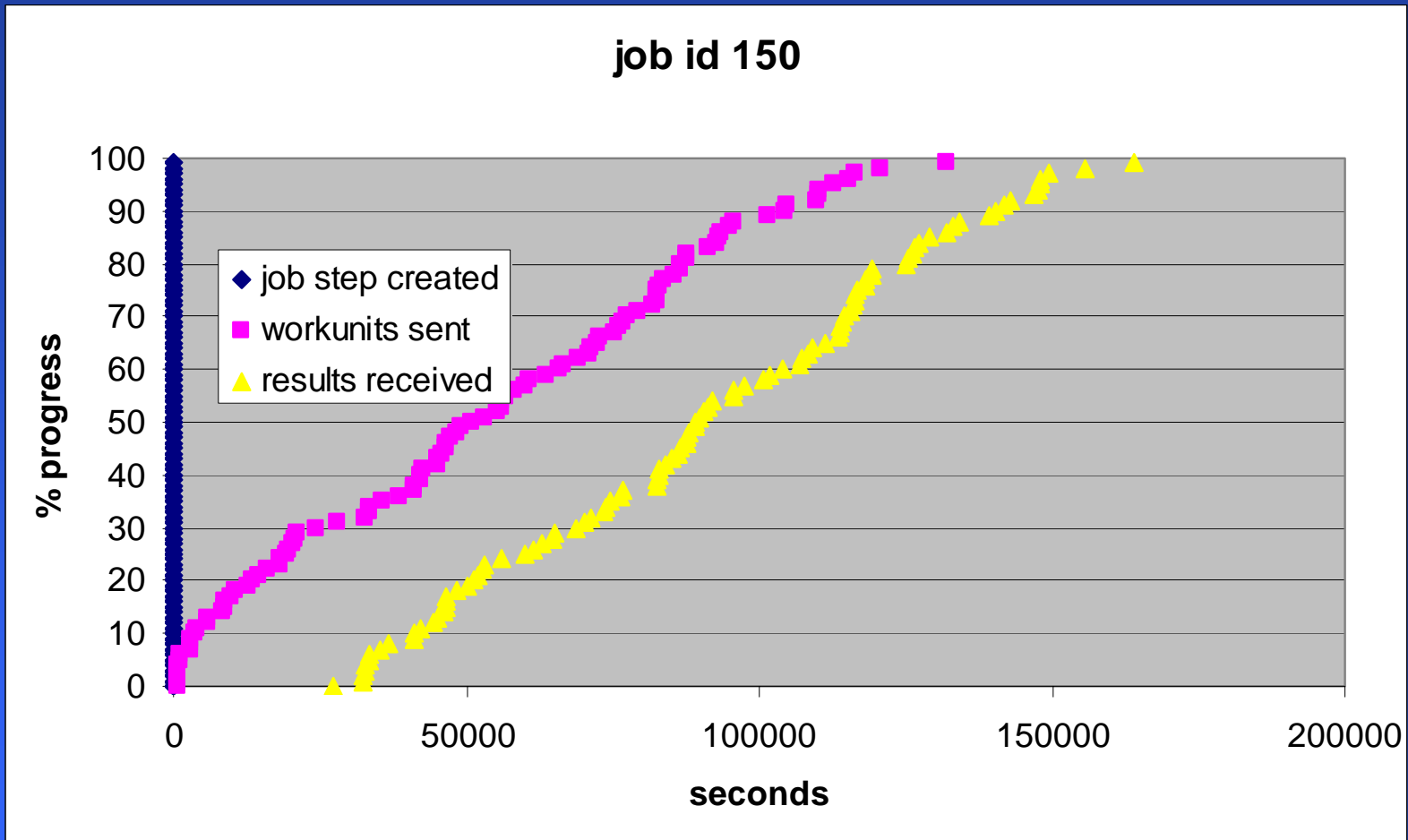
- 4MV heavy filtration
 - 100 x 700 byte/s = 70 kB/s
 - Should be manageable
- Even 19MV light filtration should be ok ...

Some results

- How long does each agent spend on a workunit?



What about progress?



Efficiency?

- Most realistic measure:
total cpu time / elapsed time / number of devices available
- Result (really preliminary, no tuning of server)
About 50% for jobs that last 4 minutes (NPLLINAC)
About 90% for jobs that last 10 hours (PTRAN)

Conclusions (i)

- Grid computing really *is* here now
- It is available on Windows as well as linux, etc.
 - (it's not that I am a Windows enthusiast, but that's what's 95% of the cpu on site run)
- It makes obvious economic sense:
 - Licence is ~ \$100 per device
 - Transferable when PC hardware is replaced
 - Transferable when switch from MS Windows to Linux
 - The business requires that most staff have PCs – the marginal cost of harnessing all those cpus is not much more than the electricity cost of leaving them on out of hours
 - United Devices have to compete with a lot of Free Software (on Unix if not on Windows) – their pricing is “flexible”.

Conclusions (ii)

- There are some unresolved issues in our system – maybe because many/most devices have the same speed (100 Mbs) network interface as the servers
 - Workunit result error rate is sometimes high
 - (recalculated automatically)
 - Devices may be trying to hit server with 30 Gbyte/min...
 - Devices can timeout during initial transfer from server
 - (resent automatically)
 - Server may try to send out 100 x 20MB simultaneously

Ever the optimist

- Our use of the Grid has only just begun
 - Already good for production runs with fixed data size code
- I expect that with a bit of learning / tuning
 - It will even be good for short “steering” runs, of a few minutes (though probably not a few seconds)

Post-script (i)

- What about coprocessors?
- In 1988, on hearing about the Weitek chip that would offer 5 Mflops, BLIF said
That could change the way I work
- In 2004, I googled for weitek and found out about
 - www.clearspeed.com
 - Their CS301 coprocessor:
 - 64 processors on a chip
 - 2 FPU per processor
 - 12800 Mips
 - 25.6 Gflops
 - 2W at 200MHz
 - 10 Gflops / Watt
- That could change the way / work...

Post-script (ii)

- Who are ClearSpeed?
 - I googled some more, and found an “ex-Inmos employees re-united” site, and found that
 - some had gone on to Meiko (late '80s)
 - Others had ended up at ClearSpeed
- The Transputer lives on, in spirit ...