
# A lightweight job execution control framework

Jakub T. Moscicki, CERN/IT


# DIANE tool

- Help smaller scientific communities using distributed (Grid) resources more efficiently
  - reduce the application execution time
  - reduce the manual work overhead by providing fully automatic execution and failure management,
  - efficiently integrate local and Grid resources
- R&D project started in 2000
  - **http://cern.ch/diane**
  - now part of EGEE Respect suite
  - part of Grid Application Support Services at CERN
    - together with Ganga (Job Management Interface)
    - **http://cern.ch/ganga**

# Grid Application Support @ CERN

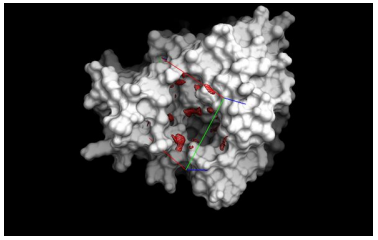
- Infrastructure for smaller communities/generic applications
  - a set of configured master hosts hosted at CERN
  - Grid resources: GEAR VO
  - tools: DIANE + Ganga
- Applications

Geant 4

COMPASS

HARP

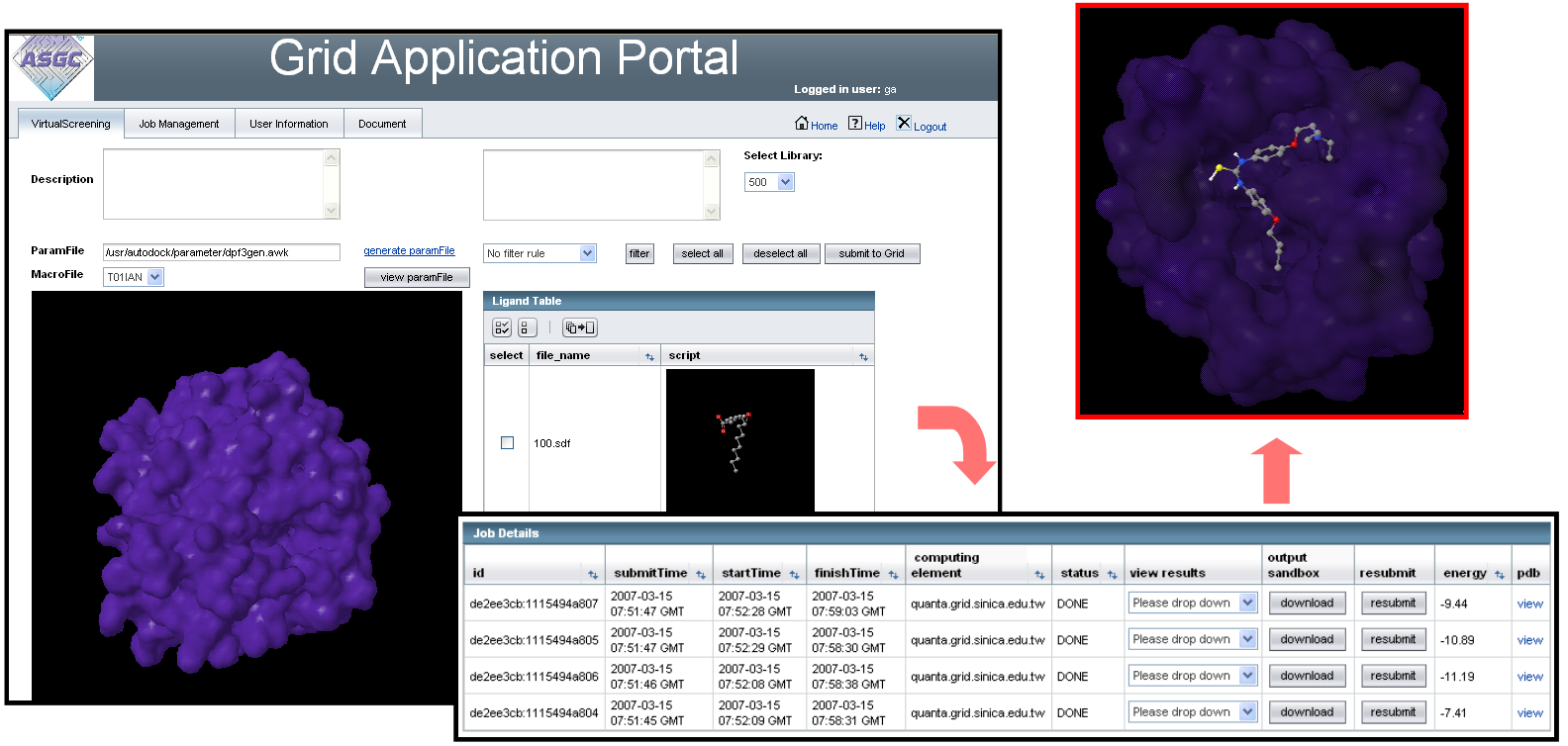
Garfield



Academia Sinica
Genomics Research Center

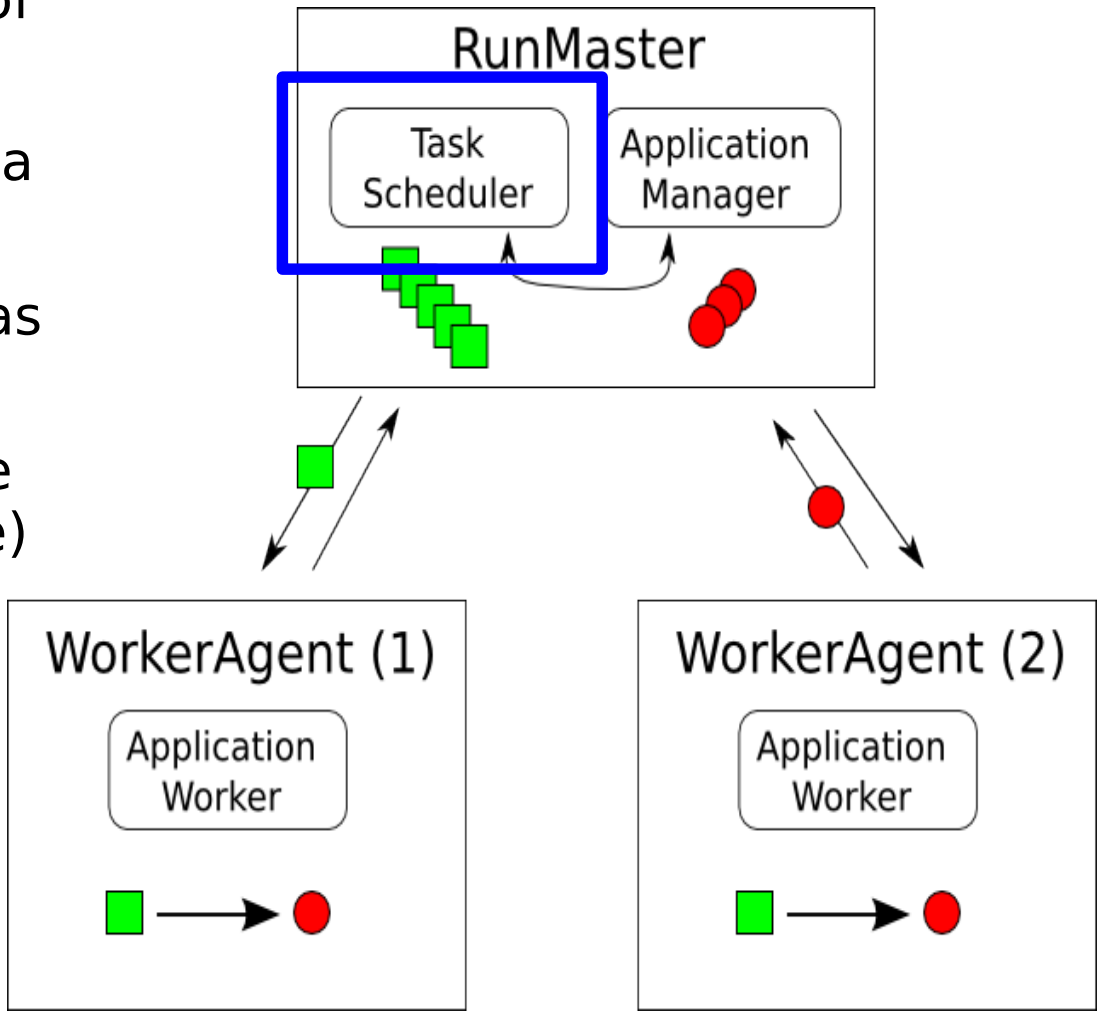
# Web-portal for biologists

- ASGC portal for virtual screening (docking)
  - AvianFlu Drug Search et al. (H.C.Lee 2006)
  - web-interface created by biologists
  - in the backend: DIANE and Ganga to handle Grid jobs

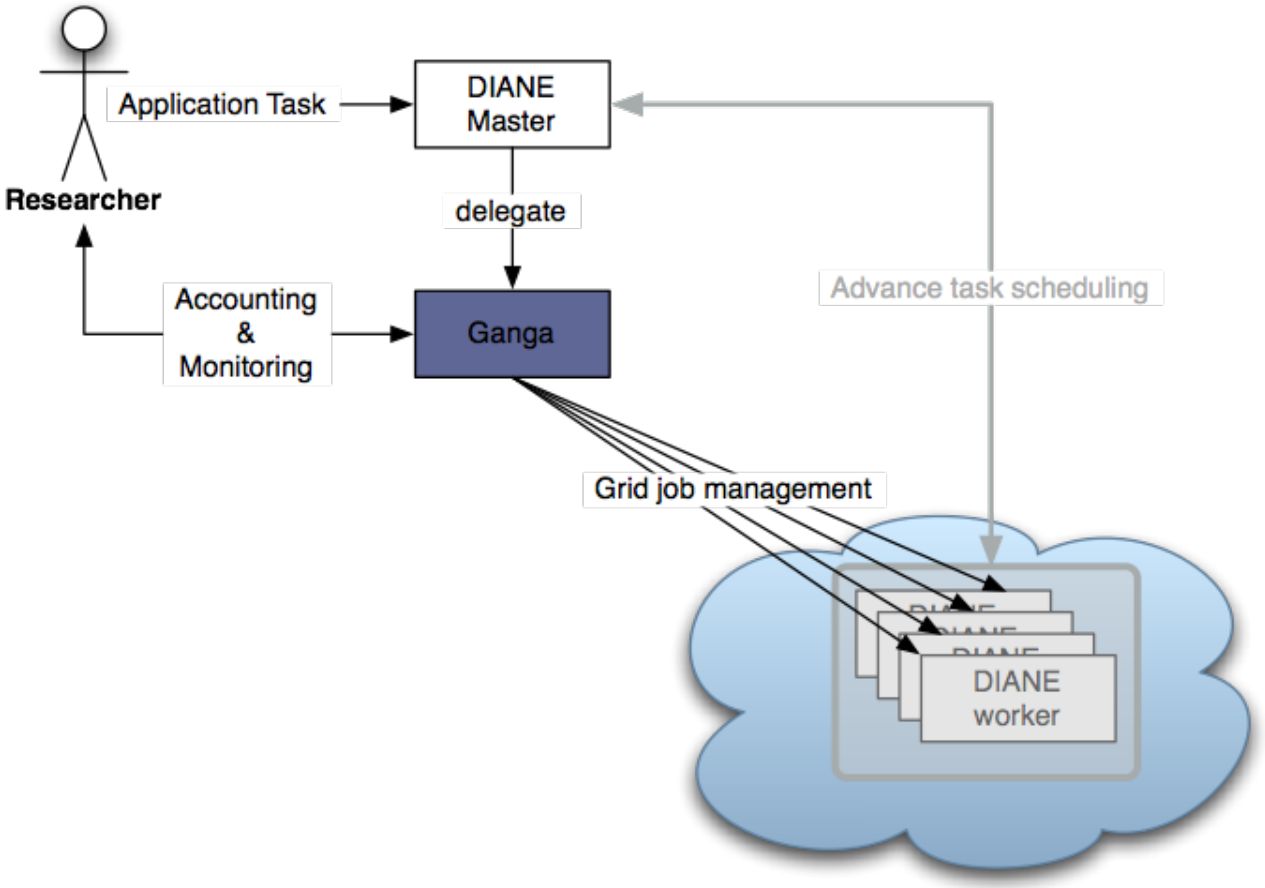


# Simple computation model

- Master/Worker processing of tasks
  - RunMaster executes on a local host
  - WorkerAgents execute as Grid jobs
- TaskScheduler is a software component (python module) which may be arbitrarily customized or replaced
- application “wrappers” (plugins):
  - ApplicationWorker
  - ApplicationManager



# Architecture

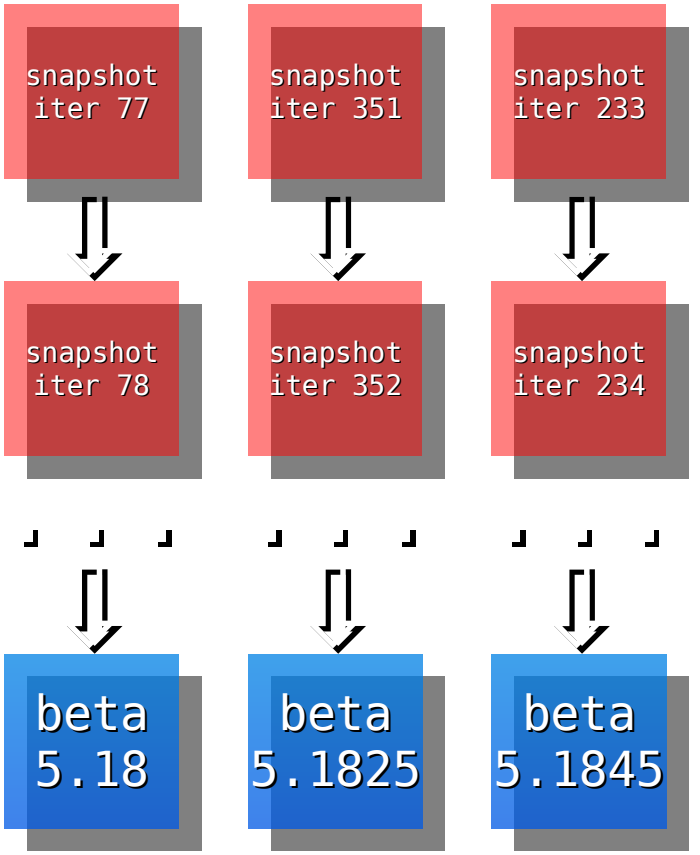


# Example: Lattice QCD @ Grid

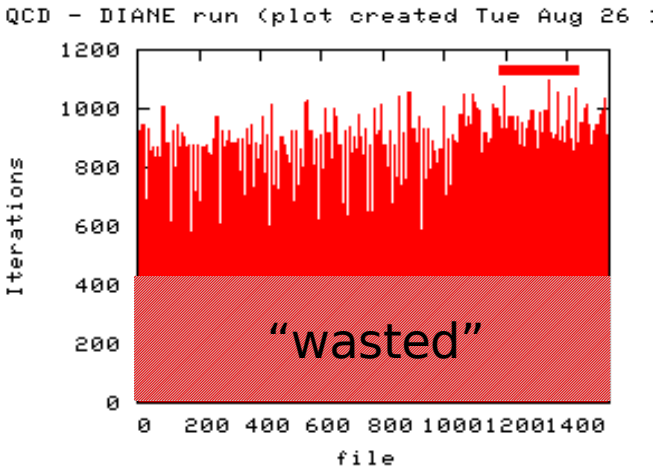
- Study the behaviour of the critical point of quark-gluon plasma
  - The scientific results obtained by the LQCD project were published in a paper P. de Forcrand et al.: "The chiral critical point of Nf = 3 QCD at finite density to the order (μ/T)4" and are available at http://arxiv.org/pdf/0808.1096
- Monte-Carlo simulation of discrete space-time lattice
  - need a lot of CPU
  - relatively small data (~Gbs)

# Structure of LQCD application

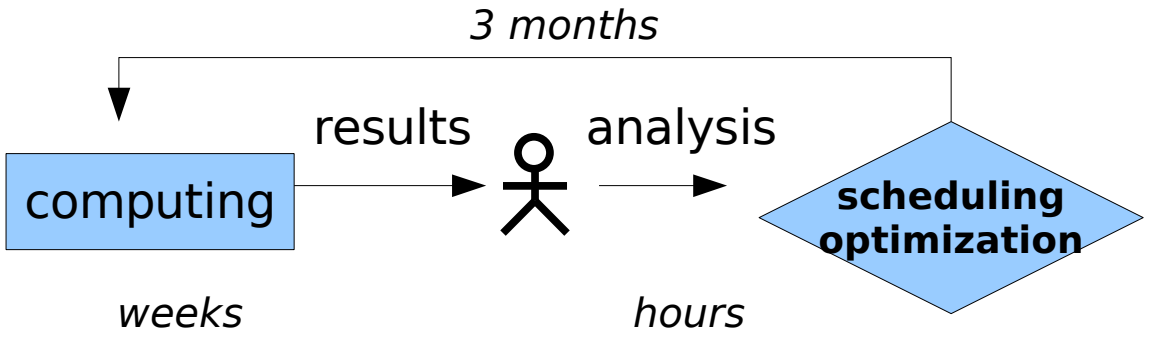
- Monte-Carlo simulation of discrete space-time lattice
  - snapshots of the lattice state are evolved iteratively
  - snapshot file is 10MB large
  - single iteration takes around 1.5 hours on a WN
  - task = one iteration
- 16 beta values x N random seeds = 400-1500 snapshots

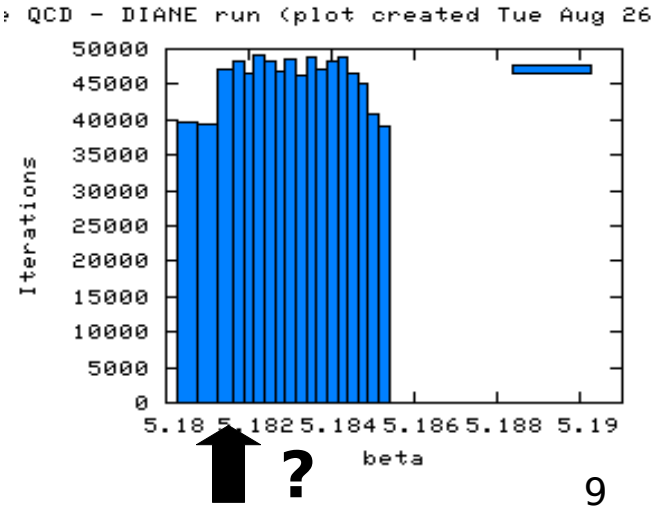


# Scheduling of snapshots

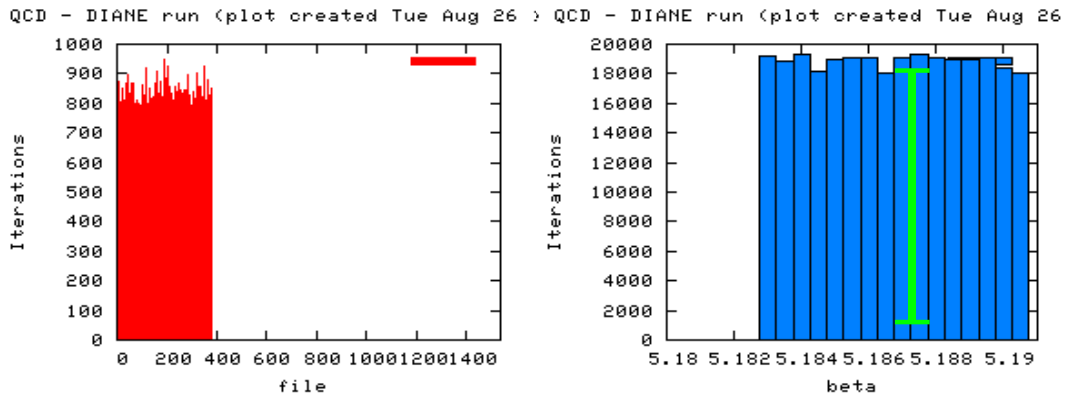



- constant overhead
  - each snapshot must diverge from the original state in order to be significant
    -> 300-500 iterations are "wasted"
    i.e. 20-30 CPU days
- how to best schedule the iterations?
  - make convergence of snapshots equally slow for all beta values
  - or prefer some beta values if more important scientifically?
- long-term optimizations with the user

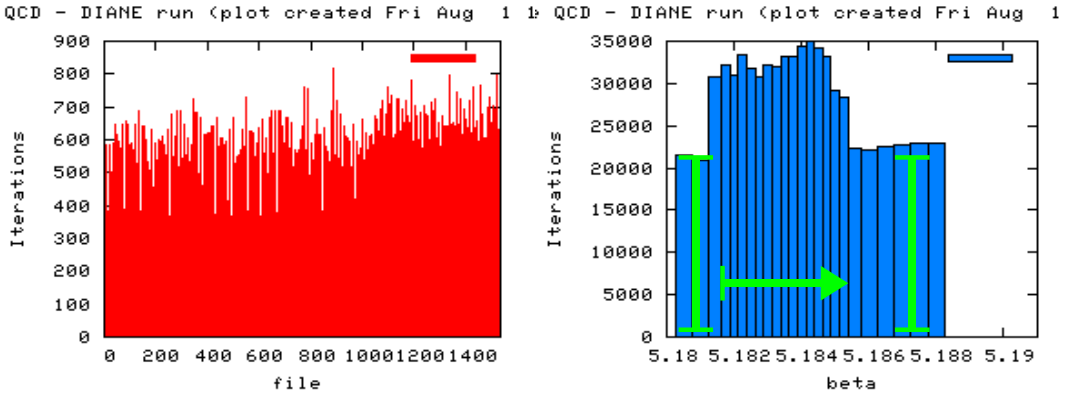


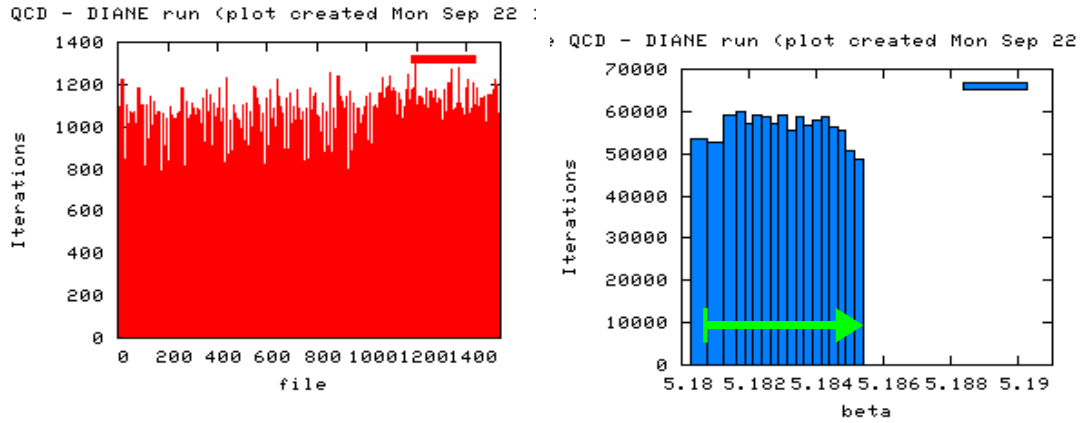
# Scheduling of snapshots




(1) 400 snapshots, 16 betas
all equal



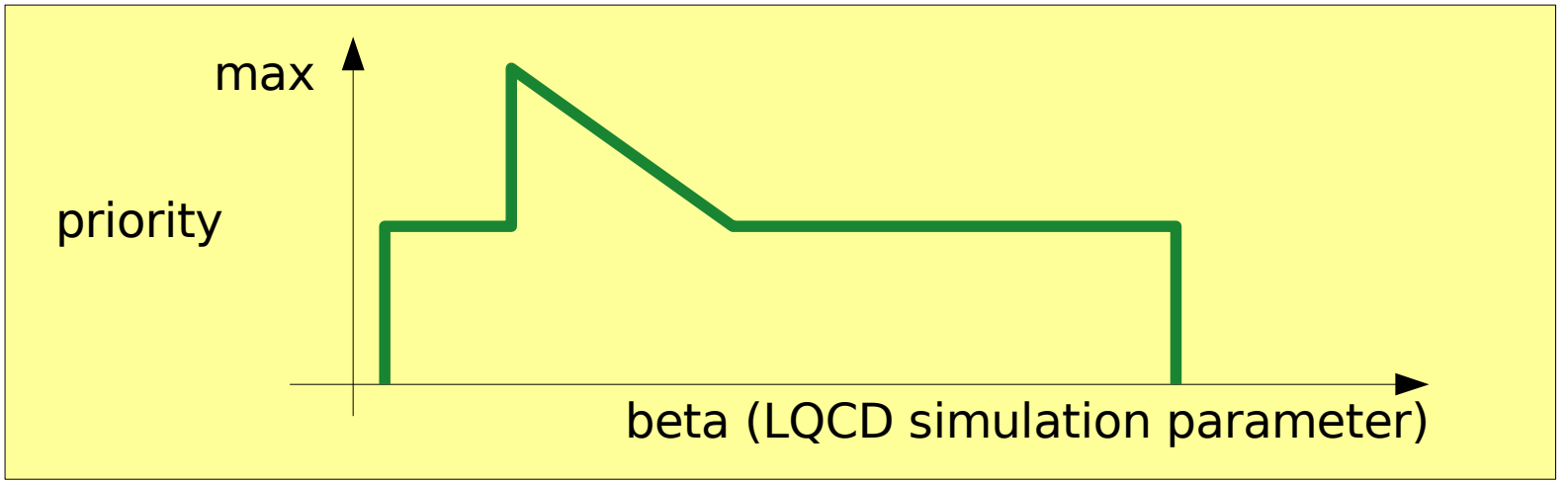

(2) 1500 snapshots, 16 betas
priority window left-to-right
equal outside of the window




(3) 1000 snapshots, 10 betas
priority left-to-right

# Task Scheduler plugin

- LQCD Task Scheduler plugin
  - simple python class (around 100 lines)




- standard DIANE scheduler
  - available out-of-the box
  - useful for most trivially parallel applications
  - simple scheduling policies

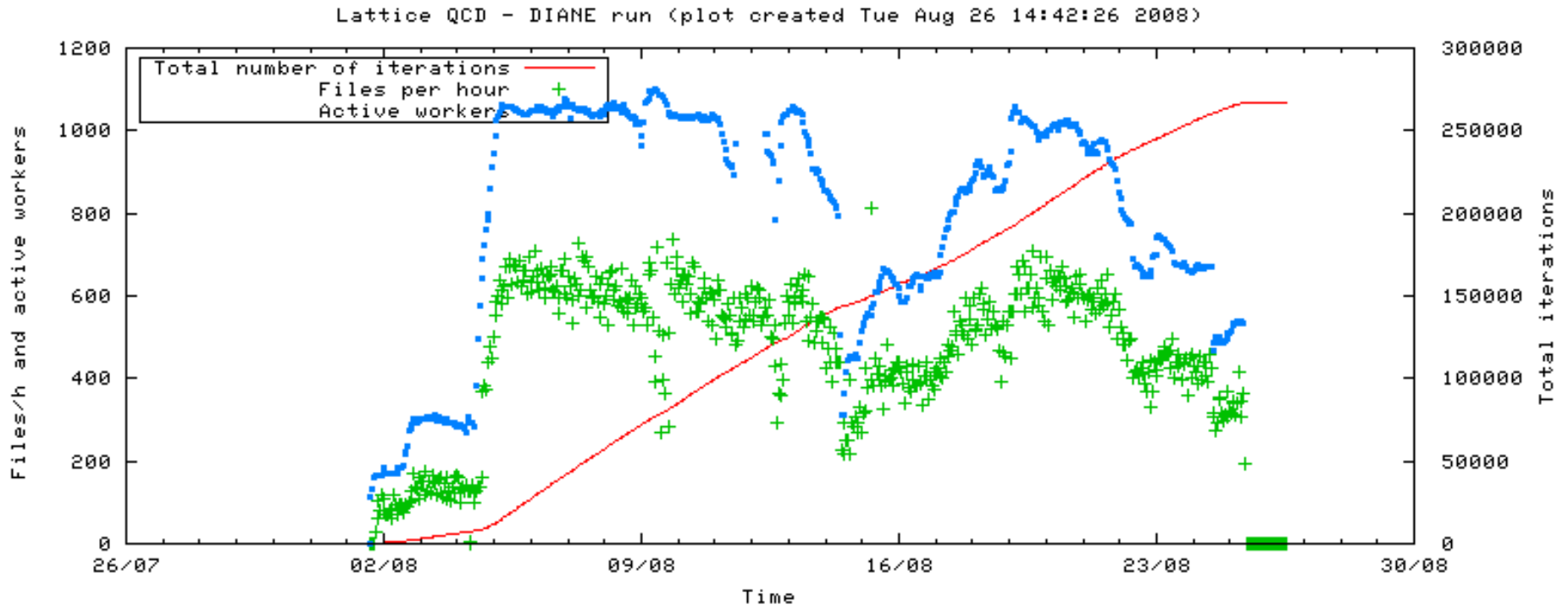
    similar to gLite WMS: retry counts, etc.

# LQCD execution history

- ongoing since May 2008
  - several phases (application and system upgrades, power-cuts, etc...)
  - long periods of uninterrupted computing
- little manual maintenance overhead
  - automatic task scheduling done by the master
  - automatic submission of pilot agents by an AgentFactory

# LQCD throughput

- ~2 million CPU hours / 3 months
  - i.e. 231 CPU years
- ~4.3 TB of data transferred
- 1000 simultaneous workers



# Summary

- DIANE main features
  - efficient application scheduling via python plugins
  - fully automatic job execution at large scale
  - 100% reliability from the user perspective
- Related events at EGEE 08:
  - ***(5) Grid-enabled Virtual Screening Service based on Grid Application Platform***
  - ***(24) Porting ThIS on the EGEE Grid***
  - ***(42) Ganga and DIANE: powerful job management and resource***
- ***http://cern.ch/diane***

# backups

# Other aspects

- One aspect was addressed: scheduling
- however DIANE may be used to control very short jobs too. Sometimes it may not be possible to statically optimize (cluster) short jobs (ITU) and then the only option is to use a “pull” scheduler.
- connection-oriented vs connectionless mode. possible to quasi-interactively control the worker nodes (cutting out the Grid scheduling overhead).
- a master may also serve as a synchronization point: e.g. one task requires output of other tasks (workflows) or task execution should be synchronized on some condition. such capabilities may be achieved by plugins.