

# Creating US PhenoGrid

## A particle physics resource based on OSG

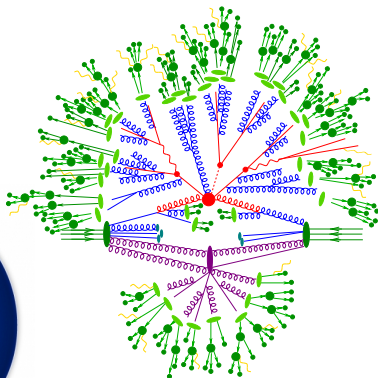
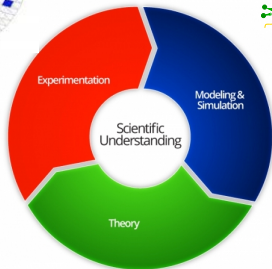
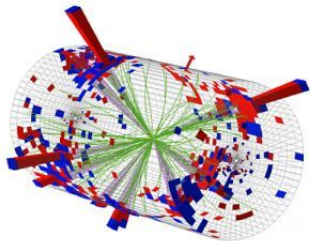
Stefan Höche, Frank Würthwein

SLAC & UCSD

OSG Council Meeting

Chicago, 10/08/15

# Monte-Carlo simulations in hadron collider physics



$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + h.c.$$

# The MCnet collaboration



**Members:** CERN, Durham, Göttingen, Karlsruhe, Louvain, Lund, Manchester, UCL

**Associates:** Dresden, Monash, SLAC, Zürich

# General-purpose MCs developed by MCnet

[<http://www.montecarlonet.org>]

**Herwig** CERN, Durham, Karlsruhe, Manchester

- ▶ Originated in studies of coherent QCD evolution
- ▶ Front-runner in matching of NLO QCD ME and PS
- ▶ Original framework for cluster fragmentation

**Pythia** CERN, (FNAL), Lund, Monash

- ▶ Originated in hadronization studies → Lund string
- ▶ Leading in development of models for non-perturbative physics
- ▶ Extensive PS development and earliest ME+PS matching

**Sherpa** CERN, Dresden, Durham, Göttingen, SLAC, Zürich

- ▶ Started with matching of LO ME and PS
- ▶ First automated framework for NLO calculations
- ▶ First automated merging of ME and PS at NLO
- ▶ First NNLO+PS matching in general-purpose MC

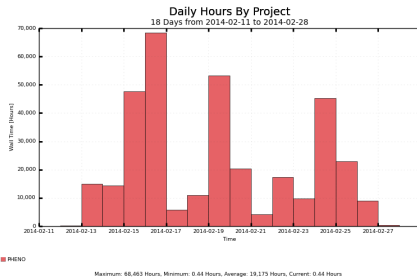
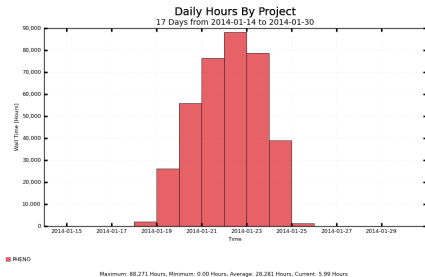


# Why OSG?

- ▶ LHC final states typically very complex (lots of jets!)  
interesting signals small and in tricky corners of phase space  
→ Need many events to simulate backgrounds accurately  $\mathcal{O}(10M)$
- ▶ Precise theoretical predictions needed to extract physics parameters  
→ Simulation of one event can take anywhere from 10ms to 100s
- ▶ Theory uncertainty estimates mandatory to control systematics  
→ Multiplies computational effort by factor of  $\mathcal{O}(1 - 10)$

**A single high-precision MC prediction costs  $\mathcal{O}(250k)$  CPU hours**

# Examples: $H/t\bar{t}$ +jets simulation at NLO



- ▶  $H$ +jets:  $\sim 481$ k hours worth of CPU time in  $\sim 6$  days
- ▶  $t\bar{t}$ +jets:  $\sim 345$ k hours worth of CPU time in  $\sim 14$  days

# PhenoGrid Proposal

## **Aim**

- ▶ Make output of simulations useful beyond a single analysis
- ▶ Provide reference event samples for both theory and experiment

## **Realized by integrated system of**

- ▶ Generator setups managed by experts
- ▶ Coherent runtime environment
- ▶ Storage and management of results

## **Requirements**

- ▶ Storage:  $\sim$  2-20 TB per event sample
- ▶ CPU: 16 core parallelism or better (initialization stage only)
- ▶ Memory: 2-4 GB / core