HPC @ ATLAS: Blue Waters

US ATLAS Software Planning Meeting
LBNL
August 21, 2014

Mark Neubauer

University of Illinois

Blue Waters

- NSF-funded ($200M) Supercomputer @ U. Illinois & NCSA
- Cray XE6/XK7 hybrid machine composed of
  - AMD 6276 "Interlagos" processors
  - NVIDIA GK110 "Kepler" accelerators (XK nodes)
  - Cray Gemini 3D Torus interconnect
- Compute
  - 23k Cray XE nodes
    - 362k Bulldozer cores; 1.4 PB memory
  - 4.2k Cray XK nodes
    - 34k Bulldozer cores + 4.2k Kepler accelerators
    - 135 TB CPU / 25 TB GPU memory
- Storage, I/O
  - Online: 26 PB, aggregate I/O > 1 TB/sec
  - Near-line: 380 PB,1.2 GB cache front-end, 58 GB/s aggregate to tape

National Petascale Computing Facility

Blue Waters Allocation

- Policy: "*2% of the available time will be allocated to university projects:* ***(i)*** *faculty whose research and/or education programs would be greatly enhanced by access to Blue Waters and* ***(ii)*** *research and/or education proposals where a commitment of Blue Waters resources will significantly increase the competitiveness of the proposals*"
- Three types of allocations: **Exploratory**, **General**, **Education**
- Exploratory proposals (twice per year, next is due Sept 15!):
  - Evaluate/tune code for platform, demo application for General proposal
  - 20k-50k node-hours over 6 month period (non-renewable)
  - "*Project proposals are expected to demonstrate that no other resource would be suitable for a given problem, as Blue Waters is not merely a large source of compute cycles*"
    - my interpretation ➔ "don't view us simply as a bag of rocks to pour sand into"

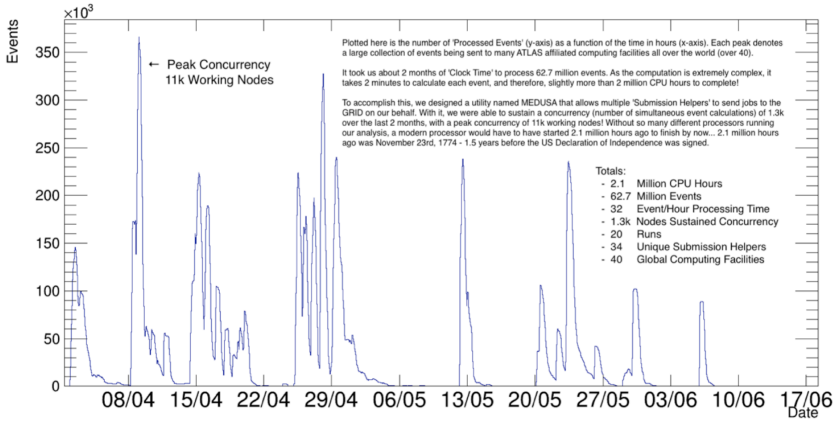
Possible Applications on Blue Waters

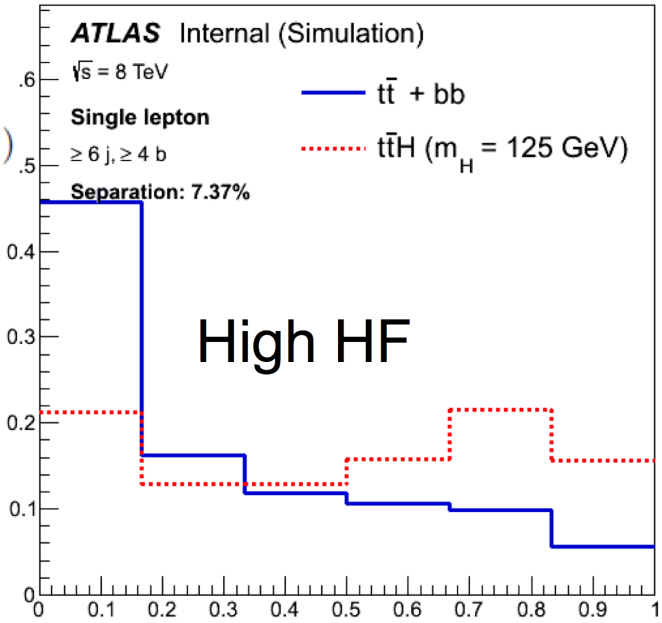
- Matrix-element (ME) calculations
  - Calculate likelihood function for signal and background discrimination
  - Numerous applications of ME technique in ATLAS. At U. Illinois:
    - Search for ggF/VBF H→WW→lvlv (P. Chang). Based on MadWeight
    - Search for ttH(bb) search (A. Bayse). MG-based stand-alone code

$$w_i(\mathbf{y}) = \frac{1}{\sigma_i} \sum_{\text{perm}} \int dx \int dx_a dx_b \frac{2g(x_a;Q_i)g(x_b;Q_i)}{x_a x_b s} \times$$

$$\times \delta\{(x_a P_a + x_b P_b) - \sum_{k=1}^{8} p_k\} \mathcal{R}\left(\vec{\rho}_T | \vec{P}_T\right) \left|\mathcal{M}_i^{LO}(\mathbf{x})\right|^2 W(\mathbf{y}|\mathbf{x})$$

  - A. Bayse: “Epic Grid use” (Ms of CPU-hours)

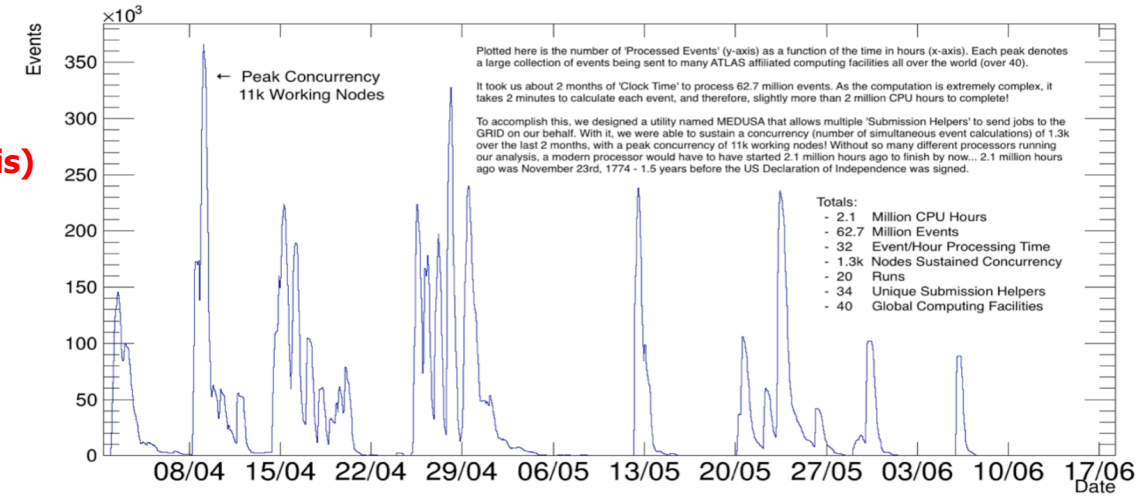




Possible Applications on Blue Waters (cont.)

- ME calculations (cont.)

A. Bayse (Illinois)
ttH analysis



  - Been thinking of better ways to do ME computing for ATLAS analysis
    - P.Chang (Illinois) worked on GPU acceleration of phase space integration and demonstrate 500-fold speed-up on Fermi Tesla GPU
      - New Illinois student will start working on this; collaboration with Chicago
    - We've been thinking through increasing longevity and applicability of ME calculation
      - Large-scale phase space sampling (i.e. create a look-up table)
      - NLO ME with phase-space transformation (like in MadWeight) a holy-grail-like
  - BW, as a large GPU resource, could help with the ME calculations scale

Possible Applications on Blue Waters (cont.)

- Fast Hardware Tracker (FTK) simulation?
  - Having a mixed hardware (CPU + co-processor) solution can improve the simulation performance
    - FTK algorithm (based on pattern recognition) are designed to be parallelized
  - ATLAS-wide FTK simulation strategy has been discussed (see https://indico.cern.ch/event/309997) where several approaches were proposed
  - Strategy document for b/τ trigger simulation (https://cdf.cern.ch/record/1747057)
  - ATLAS Simulation load is large and needed @ high pile-up
- High-pile simulation? (high-memory needs)
- Event server demonstrator?

Plans

- Listen carefully to HPC discussing in this workshop (ongoing)
- Develop Exploratory Proposal and decide if its worth submitting in Sept
  - If not, rinse and repeat next year

Bonus Material

University of Illinois

XE Compute Node

AMD 6276 Interlagos Processors	2
Bulldozer Cores*	16
Integer Scheduling Units**	32
Memory / Bulldozer Core	4 GB
Total Node Memory	64 GB
Peak Performance	313.6 GF
Memory Bandwidth	102.4 GB/s

XK Compute Node

AMD 6276 Interlagos Processors	1
Bulldozer Cores*	8
Integer Scheduling Units**	16
Memory / Bulldozer Core	4 GB
Node System Memory	32 GB
GPU Memory	6 GB
Peak CPU Performance	156.8 GF
CPU Memory Bandwidth	51.2 GB/s
CUDA cores	2688
Peak GPU Performance (DP)	1.32 TF
GPU Memory Bandwidth	200 GB/s

Interconnect

Architecture	3D Torus
Topology	24x24x24
Compute nodes per Gemini	2
Peak Node Injection Bandwidth	9.6 GB/s

Online Storage

Total Usable Storage	26.4 PB
Aggregate I/O Bandwidth	> 1 TB/s

File System	Size (PB)	# of OSTs
home	2.2	144
projects	2.2	144
scratch	22	1440

See: https://bluewaters.ncsa.illinois.edu

Slide from CMS version of ttH(bb) ME analysis

- Calculate signal and BG likelihood function

$$w_i(\mathbf{y}) = \frac{1}{\sigma_i} \sum_{\text{perm}} \int dx \int dx_a dx_b \frac{2g(x_a;Q_i)g(x_b;Q_i)}{x_a x_b s} \times$$

$$\times \delta\{(x_a P_a + x_b P_b) - \sum_{k=1}^{8} p_k\} \mathcal{R}\left(\vec{\rho}_T | \vec{P}_T\right) \left|\mathcal{M}_i^{LO}(\mathbf{x})\right|^2 W(\mathbf{y}|\mathbf{x})$$

- Weight with b-tagging likelihood for bbbb or bbll hypothesis

$$\mathcal{P}_S(\mathbf{y},\xi) \equiv w_S(\mathbf{y})\, \mathcal{L}_{bbbb}(\xi)$$
$$\mathcal{P}_{B_1}(\mathbf{y},\xi) \equiv w_B(\mathbf{y})\, \mathcal{L}_{bbbb}(\xi)$$
$$\mathcal{P}_{B_2}(\mathbf{y},\xi) \equiv w_B(\mathbf{y})\, \mathcal{L}_{bbjj}(\xi)$$

- Calculate ME likelihood ratio
  - For ttbb vs ttll discrimination: $P_{b/j} = \frac{\mathcal{P}_{B_1}}{\mathcal{P}_{B_1} + \mathcal{P}_{B_2}}$
  - For ttH vs tt+bb/ll discrimination: $P_{s/b} = \frac{\mathcal{P}_S}{\mathcal{P}_S + \lambda_{b/j}\mathcal{P}_{B_1} + (1-\lambda_{b/j})\mathcal{P}_{B_2}}$
  - $\lambda_{b/j}$ is ttbb/ttll fraction from MC (not varied in systematic unc.!!)