High Performance Computing Systems and Computing for the ATLAS Experiment

Borut Kersevan and Richard Mount

(Computing Coordinator and Deputy Computing Coordinator of the ATLAS Experiment)

Draft 3

May 14, 2013

1 Introduction - High Energy Physics and High Performance Computing

The ATLAS Collaboration is actively seeking access to High Performance Computing Systems.

ATLAS scientists are convinced that their codes need fundamental re-engineering to address the realities of future commodity highly parallel processors and that they could emerge from this process with codes that exploit High Performance Computing facilities (HPC) quite well. Explicitly targeting HPC in the re-engineering efforts is a tactic that will trickle down to benefit the Ethernet cluster approach that is now used for most HEP computing.

Thought experiments and existing proof-of-concept prototypes indicate that speedups of 10 to 100 may be attainable in the long term by major re-engineering to bring out the latent vector-friendliness and potential GPU-friendliness in our simulation and reconstruction tasks.

2 Scale of Needs

The ATLAS experiment [ATLAS DETECTOR] at the CERN LHC uses a geographically distributed grid of approximately 130,000 cores continuously, (over 1000 million core-hours per year) to simulate, and analyze[1] its data. After the early success in discovering a new particle consistent with the long awaited Higgs boson [ATLAS 2012], ATLAS is preparing for the precision measurements and further discoveries that will be made possible by much higher LHC collision rates from early 2015. The need for simulation and analysis would overwhelm the expected capacity of ATLAS computing facilities unless the range and precision of physics studies were to be curtailed. Even today, important analyses such as measuring the Higgs decaying into a b-quark and a b-antiquark ($H \rightarrow bb$), that require large simulated samples, expect to have to wait for many months before ATLAS resources will be able to provide the simulated data they need. Moving forward, this type of backlog only stands to grow larger and must be addressed for continued good progress of the field and responsible exploitation of the accelerator and detector.

Storage of ATLAS data, currently at over 100 petabytes of active data worldwide, increasingly strains against the limits of manageable and analyzable data.

From the usage information above, it becomes clear that HPC contributions of the order of 10 Million or more core hours per year become important and valuable. HPC contributions delivering the equivalent of hundreds of millions of core-hours

[1] We use "analysis" to describe the entire process of extracting physics results from data. This process includes tasks such as "reconstruction", performed systematically on the acquired data to produce derived data, and physics analysis, performed, by many teams or by individual physicists, using derived real data in combination with simulated data.

would have a major positive impact on scientific productivity. ATLAS computing can also be a close to ideal "crack-filling" application. The ATLAS production management system is being upgraded to make it aware of dynamically changing resources, and thus able to exploit groups of processors that become available for relatively short times. Storage capacity on the petabyte scale will be very valuable when available, but simulation, the largest user of CPU power, can be effectively performed on facilities offering only transient storage.

3 Technical Issues and Computer Science Challenges

The major issues and challenges that must be addressed to attain excellent scientific productivity on advanced computer architectures are:

- **Execution Efficiency:** In common with most of HEP, ATLAS code typically executes around 0.8 instructions per clock cycle, even though the (Intel, AMD) hardware it executes on can perform more than 10 instructions/cycle and achieves between 2 and 3 instructions per cycle for much scientific code. The reasons for this "bad" performance are the many conditional branches and lack of repetitious activities, both due to the complex nature of the physics and devices being simulated or analyzed.
- **Parallelism:** In its current form, ATLAS simulation and analysis is highly amenable to trivial high-level parallelism: events (*i.e.* an initial collision and the subsequent response of the detector to all the collision products) can be independently simulated or analyzed on thousands of cores. Current ATLAS simulation and analysis makes heavy use of this parallelism. ATLAS codes also have the potential for massive low-level parallelism that may be a good match for vector and GPU hardware. To expose this parallelism, the execution architecture must be transformed to present many similar (but unrelated) operations for simultaneous execution. The level of parallelism that can present "close-to-SIMD work packages" for efficient execution on many-core CPUs or GPUs will benefit from the high bandwidth communications network, typical of an HPC facility, to assemble work and distribute the results.
- **Production Environment:** Major HEP simulation and analysis, such as that of ATLAS, requires sophisticated systems to control the flow of data and tasks to computing resources, to monitor workflow and to assure automated recovery from most "environmental" errors. To make productive use of HPC, interfacing to such production environments is essential. Work is already in progress on interfacing to some HPC systems [BigPanDA]. HPC facilities vary significantly in the interface and environment they present to external workload management systems. Cooperation and assistance from HPC facility experts is already important in solving interface issues such that HPC resources can be integrated effectively in ATLAS workflow and data management systems. Initiatives from HPC facilities in providing more

uniform interfaces are welcome and are directly beneficial in reducing and reusing porting efforts.

- **Living Code:** There are 6 million lines of active code in the ATLAS SVN repository. This is living code, being continuously improved as our understanding of the detector and of physics develops. The code must remain intelligible and maintainable. While small amounts of code can certainly be optimized for particular target architectures, the majority of the code must remain generic.

4 Summary: Importance of HPC

It is now accepted that ATLAS simulation and analysis must undergo a transformation to be able to make good use of current and future computer architectures. Only by using hardware efficiently will ATLAS be able to achieve the massive computing throughput that will be needed in the next decade. As an example of the approach, the decomposition of an event simulation into tens of thousands of small steps will provide an opportunity to map the code on to a wide range of architectures. HPC facilities are a particular target because of the massive computing power they can bring to bear, and because they provide the lowest possible communications overhead for massively decomposed tasks.

Fully optimized use of HPC facilities is a long-term goal, certainly requiring work of five or more years. Early access to HPC facilities will provide an important stimulus to this work. Access to facilities coupled with collaborative help in the transformation of ATLAS code would be a major scientific contribution to the physics discoveries of the next ten years.

5 References

[ATLAS 2012] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012) 1-29

[ATLAS DETECTOR] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation 3 (2008) S08003

[BIGPanDA] http://www.usatlas.bnl.gov/twiki/bin/view/PanDA/WebHome