

# Exascale Computing at the LHC : Narrative

## **Principle Investigators :**

Peter Elmer, Matthew Jones, Steven Ko,  
Salvatore Rappoccio, Lukasz Ziarek

**Mentors :** Reneta Barneva, Valentin Brimkov

May 14, 2013

# 1 Funding Strategy

This research proposal outlines the necessity to extend the computing capacity of the experiments at the Large Hadron Collider (LHC) in Geneva, Switzerland, to the exascale of high-throughput processing. This is an enormously challenging task, but a necessary one to ensure the long-term success of the LHC experiments.

Exascale computing (in this case, in high throughput) is an enormously growth-oriented area. Even during lean economic times, the US Federal Government is pledging to support this area of research, for instance in the Department of Energy Exascale Computing Initiative [1]. Indeed, the DOE Office of Science quotes :

The Exascale initiative will be significant and transformative for Department of Energy missions.

The current level of funding for the DOE EIC and related activities is \$21M. [2]. This “transformative” strategy is one that is envisioned to continue well into the future. In addition to DOE programs, the National Science Foundation (NSF) has several programs to address the problems of high-throughput exascale computing [3, 4].

In addition to governmental programs such as above, private sector funding sources are also available, such as the Google Faculty Research Awards [5], which has an interest in exascale computing projects also.

## 2 Project Organization

The principle investigators (PIs) of this proposal have a widely-varied and applicable skill set to accomplish the goals of extending LHC computing to the exascale.

- Salvatore Rappoccio has 15 years of experience programming in a high-energy physics environment, as well as other numerical software design for the private sector. He is an expert in critical areas of event reconstruction at CMS which can be optimized for multicore usage.
- Lukasz Ziarek has 9 years of experience in language, compiler, and runtime design targeted at improving multicore performance. He has worked on 5 compilers and 3 Java VMs. He is an expert at speculative and transactional computation focusing on the extraction of parallelism and lightweight concurrency.
- Steven Ko has 10 years of experience in distributed systems. His recent focus has been large-scale data processing in the cloud using MapReduce and other technologies built on top of it. He also has 5 years of experience in large-scale storage and data management in data centers.
- Peter Elmer is the deputy offline software coordinator of the CMS experiment at the LHC. He was responsible for the design and implementation of the data and workflow management system used by CMS, as well as its core event processing software and software development environment. He has made important contributions to the software and computing of several high-energy physics experiments over the past 20 years and is currently focused on planning the computing R&D needed for the next decade in CMS.
- Matthew Jones has more than 12 years experience in scientific software development, with a particular emphasis on parallel programming. As the Associate Director of the Center for Computational Research at the University at Buffalo, he has also been responsible for designing and administering high-performance computing facilities for large-scale scientific computing.

In addition to the PIs for the project, there are also two mentors, senior collaborators whose invaluable expertise and guidance will be used to complete the project.

- Reneta Barneva
- Valentin Brimkov

## 2.1 Introduction

With the discovery of the Higgs boson (the  $H$ ) with mass around  $m = 125$  GeV by the Large Hadron Collider (LHC) experiments “ATLAS” and “CMS” [6, 7], the standard model (SM) of particle physics is now complete. This model unifies the electromagnetic force (carried by the *photon*) with the weak force, responsible for radioactive decay (carried by the  $W$  and  $Z$  bosons). At long last, physicists now understand that via interactions with the  $H$  field, the  $W$  and  $Z$  bosons acquire a mass, but the photon does not. This is referred to as “electroweak symmetry breaking”.

A new phase of particle physics has therefore begun. The questions have shifted from the cause of electroweak symmetry breaking, to the study of the  $H$ , and its interactions, in detail. Two major questions arise. The first is the exact nature of the  $H$ . The second is how the  $H$ , with a relatively low mass around  $m = 125$  GeV can be responsible for electroweak symmetry breaking, when naively, quantum mechanics tells us that the mass should be much larger. To understand the larger picture of the fundamental forces in nature, it is imperative to continue the study of these questions. The past excellence of the LHC experiments must therefore continue unabated in the face of new technical challenges.

One of the major technical challenges that lies ahead is the continuation of the scaling of computational power year by year, known colloquially as “Moore’s Law”. With  $25 \text{ fb}^{-1}$  of 7 and 8 TeV data delivered by the LHC in 2011-2012, and instantaneous luminosities reaching  $7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ , the processing time to reconstruct each event collected by CMS was approximately 20 seconds per event. However, as the instantaneous luminosity is increased, the computational time currently scales quadratically. As the upgraded LHC is expected to deliver  $> 12 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$  in the upcoming run, the processing time per event is expected to reach several minutes per event as shown in Figure 1. Furthermore, in future runs of the LHC in the next 15 years, the luminosity is expected to reach as high as  $> 1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , which would correspond (naively) to several hours of computational time per event! Clearly, it is necessary for the computing power to scale in order to compensate for this dramatic increase in CPU time with instantaneous luminosity.

However, with the expected end of the historic scaling of single-core processing capability [8], it is imperative to utilize a parallel processing strategy in order to maintain the levels of computational speed of LHC data in the

immediate future in experiments such as CMS. As a starting point we first note that during 2012, the CPU usage for offline computing activities was very roughly divided up as:

- $\sim 40\%$  event simulation
- $\sim 20\%$  prompt event reconstruction (within 48 hours of data-taking)
- $\sim 40\%$  mixed user analysis applications

Oftentimes, the codes used by CMS (and experimental HEP in general) tend to lack clear numerical “kernels” where optimization efforts can be focused. Given these characteristics they are generally more properly classified as “high throughput computing” (HTC) rather than “high performance computing” (HPC). In terms of their detailed behavior on the CPU many of these codes resemble more general enterprise or “cloud” applications [9, 10].

However, there are several numerical algorithms where parallelization could be exploited more directly. One of these is the so-called “jet clustering” algorithms used at CMS. At CMS, this computation is done very often by individuals accounted in the 40% of CPU usage from “mixed user analysis applications” as described above. It is likely, therefore, that improvements observed in jet clustering will primarily benefit this portion of the CPU usage. We now discuss the prospects for utilizing parallelization in jet clustering in detail.

## 2.2 Jet Clustering:

The energetic deposits of charged and neutral hadrons in the hadronic calorimeter, as well as the deposits of electrons and photons in the electromagnetic calorimeter and the deposits of charged hadrons described above, (which are combined into a single “particle flow candidate” at CMS) need to be clustered to obtain the complete response. This is because the process inherently involves a “shower” of particles that spreads in both the lateral and radial directions, called a “jet”. This “jet clustering” is a well-established technique employed at many different particle physics experiments worldwide, and is implemented in a common software framework called **fastjet** [11]. The single-core optimization of the mathematical implementation of the nearest-neighbor (NN) algorithm chosen is outlined in Ref. [12], and depends on the number of “candidates” that are input to the jet clustering algorithm. The

single-core optimization yields  $O(N^2)$  or  $O(N \ln N)$  operational times. This is analogous to the “K-nearest neighbors algorithm” [13] (kNN).

Even though the single-core computational strategy is sufficient for many applications, there are two key components of jet clustering that are still inefficient for future LHC data processing, which can be solved with appropriate development of a parallelization strategy. The first is the parallelization of the NN algorithm itself to make optimum use of the more advanced vectorization capabilities of modern multicore CPUs. The second is to use a “divide and conquer” strategy to reconstruct a single collision event in several disjoint sections of the CMS detector in parallel.

There is existing work and literature on the topic of the parallelization of the kNN algorithm, for instance, in Refs. [14, 15, 16], where improvements  $O(100)$  in CPU performance are observed over standard algorithms. Since the proposed use case is very similar to the kNN algorithm, similar improvements to the processing time by parallelization strategies are expected. Furthermore, a critical piece of information in the jet algorithm is the “area” of each individual jet (in the NN-related metric). Currently, the procedure to estimate this area is to add a large number of infinitesimally small candidates called “ghosts” that are uniformly distributed over the total area. These are fed into the jet clustering algorithm, and the positions of these ghosts are used to estimate the size. The resolution of this area is inversely related to the number of ghosts per unit area, so in practice many ghosts ( $O(1000)$ ) are added in the neighborhoods of the jets. This severely limits the computational speed with which these can be processed. Judicious usage of parallelization of the algorithm may be able to solve this problem, and drastically reduce the computational time needed to compute the area.

In addition to the usage of GPU and other parallelization strategies to estimate the jet area, it is also possible to parallelize the computation by dividing the event into disjoint sections, and computing the jet clustering in parallel over multiple cores. This optimization is factorized from the previous, in that it can be done with current single-core computational algorithms, but simply divides the problem into several smaller ones. There are both computational and physics-related challenges that need to be overcome in this strategy, and using a synergistic approach is absolutely critical.

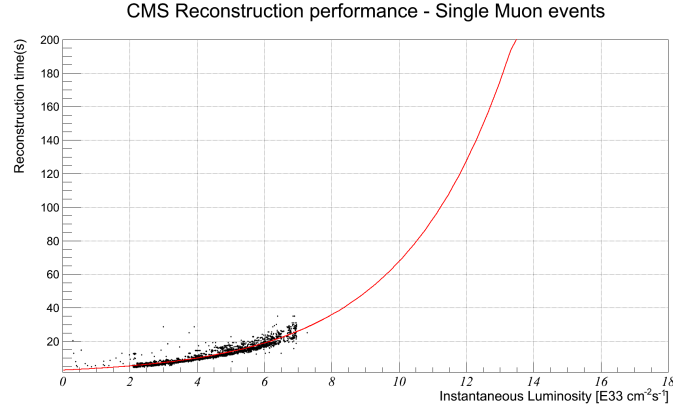


Figure 1: Event processing time versus instantaneous luminosity.

## References

- [1] US Department of Energy Office of Science, (2010), The Opportunities and Challenges of Exascale Computing.
- [2] US Department of Energy Office of Science, (2013), ASCR Budget.
- [3] National Science Foundation, (2013), Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA).
- [4] National Science Foundation, (2013), High Performance System Acquisition: Building a More Inclusive Computing Environment for Science and Engineering.
- [5] Google, (2013), Faculty Research Awards.
- [6] CMS Collaboration, S. Chatrchyan et al., Phys. Lett. B (2012), 1207.7235.
- [7] ATLAS Collaboration, G. Aad et al., Phys. Lett. B (2012), 1207.7214.
- [8] S.H. Fuller and E.C.o.S.G.i.C.P.N.R.C. Lynette I. Millett, The Future of Computing Performance: Game Over or Next Level? (The National Academies Press, 2011).



- [9] M. Ferdman et al., Proceedings of the seventeenth international conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS '12, 2012.
- [10] P. Calafiura et al., J.Phys.Conf.Ser. 396 (2012) 052072.
- [11] M. Cacciari, G.P. Salam and G. Soyez, Eur.Phys.J. C72 (2012) 1896, 1111.6097.
- [12] M. Cacciari and G.P. Salam, Phys.Lett. B641 (2006) 57, hep-ph/0512210.
- [13] T. Cover and P. Hart, Information Theory, IEEE Transactions on 13 (1967) 21.
- [14] V. Garcia et al., IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 2010.
- [15] V. Garcia, E. Debreuve and M. Barlaud, CVPR Workshop on Computer Vision on GPU, Anchorage, Alaska, USA, 2008.
- [16] V. Garcia, Suivi d'objets d'intrt dans une squence d'images : des points saillants aux mesures statistiques, PhD thesis, Universit de Nice - Sophia Antipolis, Sophia Antipolis, France, 2008.