

# Abstract

Most event data analysis tasks in the ATLAS project require both intensive data access and processing, where some tasks are typically I/O bound while others are compute bound. This dissertation work mainly focus improving the code efficiency of the compute bound stages of the ATLAS detector data analysis, complementing a parallel dissertation work that addresses the I/O bound issues.

The main goal of the work was to design, implement, validate and evaluate an improved and more robust data analysis task, originally developed by the LIP research group at the University of Minho. This involved tuning the performance of both Top Quark and Higgs Boson reconstruction of events, within the ATLAS framework, to run on homogeneous systems with multiple CPUs and on heterogeneous computing platforms. The latter are based on multi-core CPU devices coupled to PCI-E boards with many-core devices, such as the Intel Xeon Phi or the NVidia Fermi GPU devices.

Once the critical areas of the event analysis were identified and restructured, two parallelization approaches for homogeneous systems and two for heterogeneous systems were developed and evaluated to identify their limitations and the restrictions imposed by the LipMiniAnalysis library, an integral part of every application developed at LIP. To efficiently use multiple CPU resources, an application scheduler was also developed to extract parallelism from simultaneously execution of both sequential and parallel applications when processing large sets of input data files.

A key achieved outcome of this work is a set of guidelines for LIP researchers to efficiently use the available computing resources in current and future complex parallel environments, taking advantage of the acquired expertise during this dissertation work. Further improvements on LIP libraries can be achieved by developing a tool to automatically extract parallelism of LIP applications, complemented by the application scheduler and additional suggested approaches.

# Resumo

## Processamento Eficiente de análise de eventos do ATLAS em plataformas homogéneas e heterogéneas

A maior parte das tarefas de análise de dados de eventos no projeto ATLAS requerem grandes capacidades de acesso a dados e processamento, em que a performance de algumas das tarefas são limitadas pela capacidade de I/O e outras pela capacidade de computação. Esta dissertação irá focar-se principalmente em melhorar a eficiência do código nos problemas limitados computacionalmente nas últimas fases de análise de dados do detector do ATLAS, complementando uma dissertação paralela que irá lidar com as tarefas limitadas pelo I/O.

O principal objectivo deste trabalho será desenhar, implementar, validar e avaliar uma tarefa de análise mais robusta e melhorada, desenvolvida pelo grupo de investigação do LIP na Universidade do Minho. Isto envolve aperfeiçoar a performance das reconstruções do Top Quark e bóson de Higgs de eventos dentro da framework do ATLAS, a ser executada em plataformas homogéneas com vários CPUs e em plataformas de computação heterogénea. A última é baseada em CPUs multicore acoplados a placas PCI-E com dispositivos many-core, tais como o Intel Xeon Phi ou os dispositivos GPU NVidia Fermi.

Depois de identificar e restructurar as regiões críticas da análise de eventos, duas abordagens de paralelização para plataformas homogéneas e duas para plataformas heterogéneas foram desenvolvidas e avaliadas, com o objectivo de identificar as suas limitações e as restrições impostas pela biblioteca LipMiniAnalysis, uma parte integrante de todas as aplicações desenvolvidas no LIP. Um escalonador de aplicações foi desenvolvido para usar eficientemente os recursos de múltiplos CPUs, através da extracção de paralelismo de execução em simultâneo de tanto aplicações sequenciais como paralelas para processamento de grandes conjuntos de ficheiros de dados.

Um resultado obtido neste trabalho foi um conjunto de directivas para os investigadores do LIP para o uso eficiente de recursos em ambientes paralelos complexos. É possível melhorar as bibliotecas do LIP através do desenvolvimento de uma ferramenta para extrair automaticamente paralelismo das aplicações do LIP, complementado pelo escalonador de aplicações e outras alternativas sugeridas.