Overview of CompF1 Report

Snowmass Community Summer Study
July 20, 2022
Topical Group Conveners: Giuseppe Cerati, Katrin Heitmann, Walter Hopkins

Introduction

- Who are we? CompF1: Experimental Algorithm Parallelization
- Our mission: Experimental algorithms across the frontiers; processing resources; technology evolution; coordination with Simulation (CompF2) and Machine Learning (CompF3) working groups
- Draft report is available on the Snowmass wiki
- Comments and suggestions are very welcome and can be send to us either directly (contact information) or via slack: #compf01-expalgos
- Live notes are in a google doc

Experimental Algorithm Development on HPC Architectures

- **Computational requirements** have grown rapidly in recent years, in particular resource needs from the experiments for analysis, processing
- HPC architectures provide **excellent opportunities** to satisfy some of the demands, however, for experimental algorithm development they also pose **key technology challenges**
- We discuss several approaches in the report to address these challenges:
  - Parallelization and optimization of **specific (time consuming) algorithms** that take full advantage of specific architectures
  - **Portability solutions**
  - Development and optimization of **common tools**
- Approaches need to be supported by the **Software frameworks** of the experiment

Experimental Algorithms and their Connection to AI/ML

- AI/ML is expected to play an increasingly significant role in experimental algorithm
- AI/ML can improve algorithms, both in terms of computational speed and physics performance
- AI/ML frameworks (e.g., TensorFlow, PyTorch) can serve as a portability solution
- Development of fast inference could expand the use of AI/ML within experimental algorithms
- Challenges with AI/ML: Interpretability, robustness, and incorporating domain knowledge

Experimental Algorithms in the Energy Frontier

- Challenging environment require new algorithms to maximize experimental reach
  - High resolution objects (jets, leptons) to ensure precision measurement.
- Increasingly complex detectors and more diverse architectures in computing environment
  - Heterogeneous computing resources, highly-granular detectors
- Opportunity: moving online and offline algorithms closer together
  - Using heterogeneous resources: FPGAs, GPUs
- Developing and maintaining expertise is essential
  - Detector AND software expertise are required
- Portability is required

Experimental Algorithms in the Cosmic Frontier

- Many different experiments and surveys in Cosmic Frontier; data challenges more recent
- Different challenges in different areas are discussed in the report:
  - **Optical surveys:** Both “static” science and “transient” science, algorithmic requirements very different; high data throughput for static science; analysis and processing algorithms
  - **CMB experiments:** Reduction of raw detector data to maps will be major challenge in the future
  - **Dark matter experiments:** Event processing and reconstruction algorithms; not a major focus in the past

Experimental Algorithms in the Neutrino Frontier

- Experimental algorithms are mainly signal processing, calibration, and higher-level reconstruction. Most stringent (but not only) limitations come from processing of supernovae data stream, requiring large computing resource both in terms of disk and CPU
- Experimental neutrino data is **suitable for parallelization** at multiple levels due to modular structure of neutrino detectors, and can lead to significant speedups and reduction in memory usage. However, problems arise of **competing resource usage**. Interplay with framework for support of parallelization paradigm is crucial. See WireCell white paper.
- Examples of optimization of **specific algorithms** for parallel execution: hit finder
- HPC resources are being exploited for data processing for neutrino experiments: need to move from demonstration and R&D to routine work
- **LArSoft** is the common repository for LArTPC algorithms: continued support is vital for NF

Experimental Algorithms in the Rare Events Frontier

- Report development and usage of **Allen** for LHCb trigger
- We also mention an LOI looking to GPU trigger for mu2e-II
- Actively looking into **feedback and input** on how to complement the information for the RE frontier

Recommendations (Draft!)

- Experimental programs require workable solutions to exploit **heterogeneous computing platforms** efficiently.
- Support for continuing operation and evolution of **software frameworks** to adapt to new computing landscape is critical for enabling the usage of parallel algorithms in production environments of the experiments.
- **Interdisciplinary collaborations and programs** are critical to develop the most impactful algorithmic approaches. The support of such programs needs to be strengthened considerably and will be extremely valuable for the HEP community.
- **Training opportunities** for early-career researchers are essential to ensure that the experiments implement optimal algorithms that can take full advantage of new hardware and software developments.
- **Career opportunities** need to be provided for researchers who focus on experimental algorithm development. The field has become very complex and requires deep knowledge of diverse computational approaches.
- **Resource allocations** need to be sufficient to go beyond the R&D phase of algorithm and software development to production ready software.
- **Project support** has to be provided to enable long-term sustainability of software developments and to offer job security for junior researchers. In a similar way that detector development is supported by projects, software development has to be treated in a similar way.