

# Future Heterogeneous Platforms: High Throughput Data Analysis

D. Piparo, A. Naumann for the **ROOT** Team



- ▶ HL-LHC needs 10x faster software
  - And things won't get any easier after that...
- ▶ Parallelism must be granular, nested, optimising memory access, on all possible platforms
  - **Not only fat multicore, general purpose CPUs but a variety of co-processors**
  - GPUs, thin/fat cores, specialised devices - use all the silicon that's there!
- ▶ Need additional flexibility, especially for data analysis
  - Strategies for scheduling work effectively on coprocessors
  - **Decide at runtime where our analysis code (and data) should land**
- ▶ Today we produce just in time (JIT) machine code for our analysis code
  - Accomplished with ROOT's interpreter

**Can this paradigm be applied to heterogeneous environments?**

- ▶ Squeeze performance out of CPU and non CPU compute devices
  - On servers but also scientists laptops
- ▶ Automatic discovery of available resources
- ▶ Make ROOT capable to generate source and machine code for non-CPU's (CUDA / OpenCL / ...)
  - Rely on underlying industry tested compiler technology, LLVM
  - Same code adapted to different devices
- ▶ Parallelisation and scheduling: hidden by high-level interfaces
  - “Implicit parallelism” approach



# Project Context and Potential Milestones

---

- ▶ Offer a simple programming model for heterogeneous computing
  - Let scientists focus only on the algorithm, not the backend code
  - ML, minimization, etc, auto-adapt to available processing power
- ▶ Tentative intermediate milestones
  - Demonstrate support for today's GPUs, one vendor only
  - Provide simple scheduler handling standard data transfer patterns to/from devices
  - Runtime hot-spot automatic identification and analysis: real-world applications
    - Contacts to CS research on hot-spot based accelerator use in science code!