WP5
High-speed real-time data processing

Status Report

Mauro Morandin
on behalf of the LHC and SKA groups

23 September 2014

WP5 report

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- I'm reporting the outcome of the work performed within the group originally associated to the WP5
  - i.e.: “Real time High-speed real-time data processing”
- our group has actually served as a **forum** for discussing and organizing all the activities proposed by the SKA and LHC communities
- currently only a few of the proposed activities would indeed fit within the scope of WP5
- others will have to be “relocated” to other WPs

The group

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Currently the group include collaborators from:
  - the LHC community:
    - INFN (13)
    - CERN (8)
    - Univ. of Edimburgh (3)
  - the SKA community:
    - UCL (2)
    - Univ. of Cambridge (1)
    - STFD SCD (3)
    - Univ. of Edimburgh (1)
- a representative from Nvidia has been included in the mailing list, but he has never showed up

Group activities

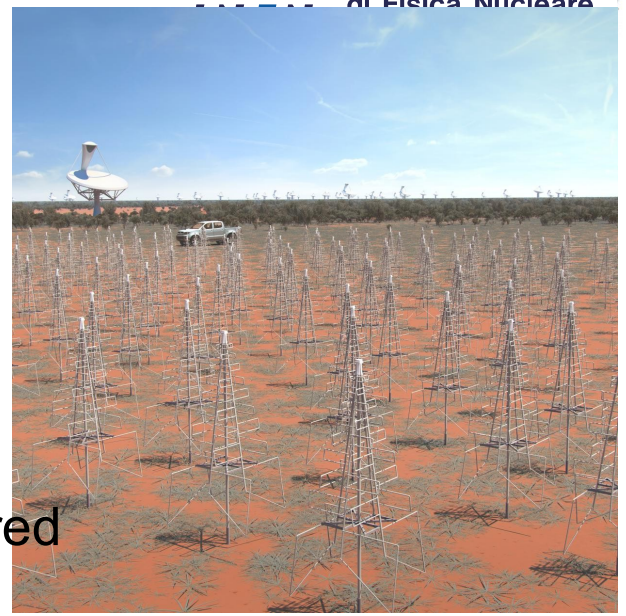
- the group has met so far three times in phone conferences
- the proposed activities have been described in documents made available on Google drive

Activities taken into consideration so far

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- LHC:
  - A1 Maximum Likelihood Fits on Heterogeneous Many Core Architectures
  - A2 Exploitation of HPC clusters for LHC data intensive applications
  - A3 Development of a parallel framework for LHC real time filtering applications
  - A4 Performance Optimization tools
  - A5 Extending Performant Detector Simulation to new architectures and physics
- SKA:
  - A6 Real time Data ingestion
  - A7 Pseudo-real time data calibration – Eigen solving 1Mx1M dense Matrices

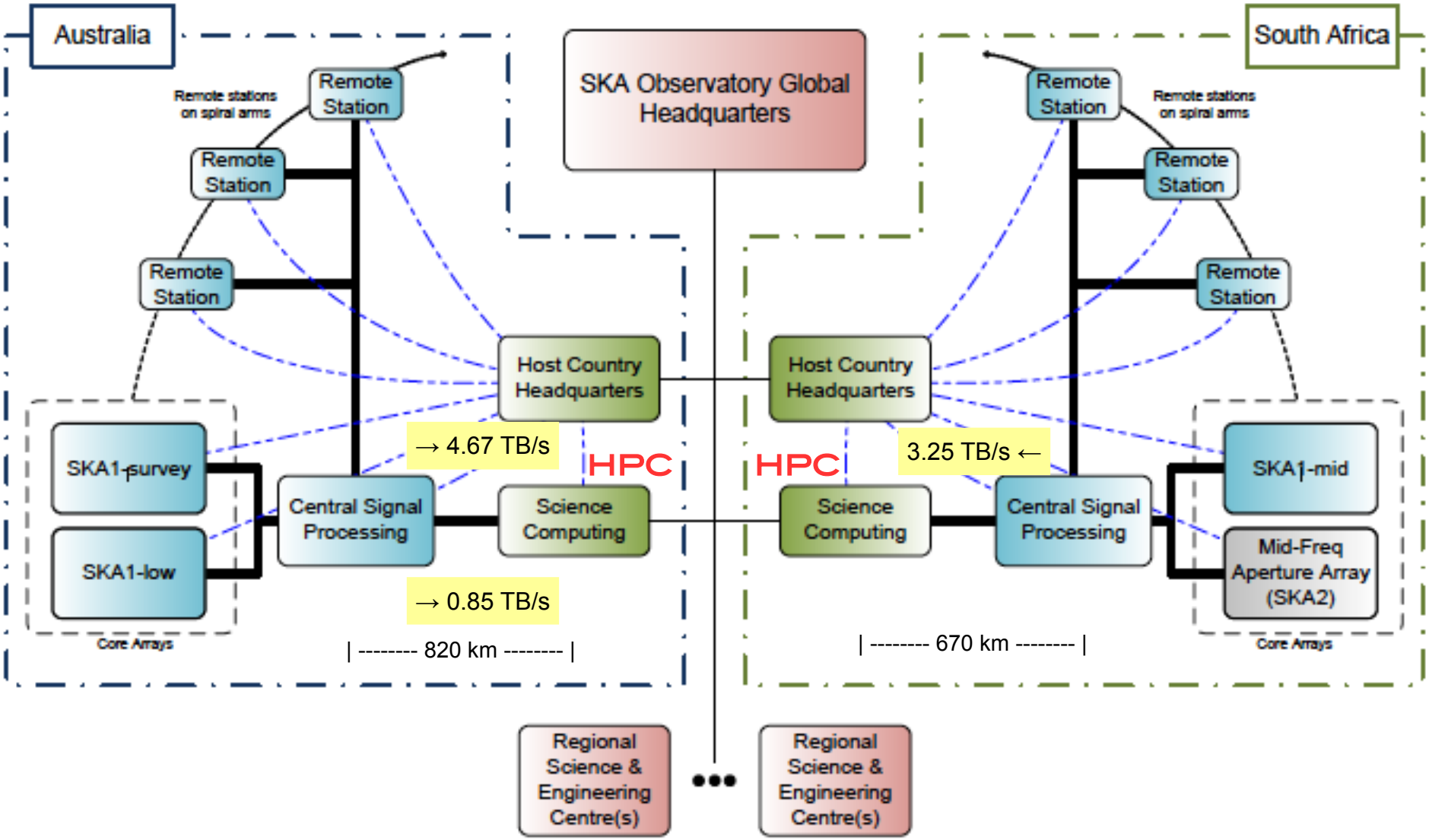
SKA (phase I)



- SKA1-low (Australia)
  - over a quarter of a million wide bandwidth dipole antennas that cover the lowest frequency band, 50-350 MHz
  - grouped in 911 stations (35 m diameter), configured in a central core (866 st. < 3km radius) and three equally spaced spiral arms within a 45 km radius
- SKA1-mid (South Africa)
  - 64 13.5-m diameter dishes (MeerKAT array) and 190 15-m SKA dishes, arranged in a central core (3 km radius) and along three spiral arms within a
  - 350-3050 MHz band
- SKA1-survey (Australia)
  - 36 12-m diameter dishes (ASKAP array) and 60 15-m SKA1 dishes
  - core (2 km radius) + spiral arms (25 km radius)

SKA Data paths

Istituto Nazionale di Fisica Nucleare
Sezione di Padova



The SKA – a hard IT problem

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- The global system parameters for the SKA visibility data buffer and Archive/Scratch (off line work) systems are listed below.
- For Phase 1 (650M euro) in Australia for 2020 deployment
  - Science and Data Processor FLOPS: 300 PetaFLOPS
    - 10,000 nodes, with lots of inherent parallelism
  - Required Data Ingest rate for visibility Buffer: 5-7 TByte/s
  - 430 PB per day – similar internet business usage
  - Scratch storage throughput: 20 TByte/s
  - Scratch storage capacity: 100 PetaBytes
    - Both disks and solid state disk possibility
  - FDR Infiniband switches could deliver the bandwidth needed
    - Latency is an issue
    - Organising the data is an issue
  - Commoditisation and general drive to low Capital & Operational costs
    - Pushes us away from proprietary solutions

The SKA proposed activities

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- The CoE allows us to investigate novel and innovative ways of solving two key areas.
- A6 Real time Data ingestion
  - Develop software and hardware solutions that can ingest and organise data that is coming into a file system at 5-7 TB/s. These are requirements on the proposed SKA Data Processor system. This whole IT system has to fit within in a 3MW energy budget .
  - **Produce solutions for large-scale energy efficient data movement and meta-data organization onto disks**
- A7 Pseudo-real time data calibration
  - Develop novel software and hardware solutions for Bayesian co-variance matrices which involve eigensolving 1M x 1M dense matrices in 5 minutes . This would allow a complete statistical calibration solution to be applied to SKA visibilities and so improve data quality.
  - **Produce solutions that will have widespread application in academy and industry**

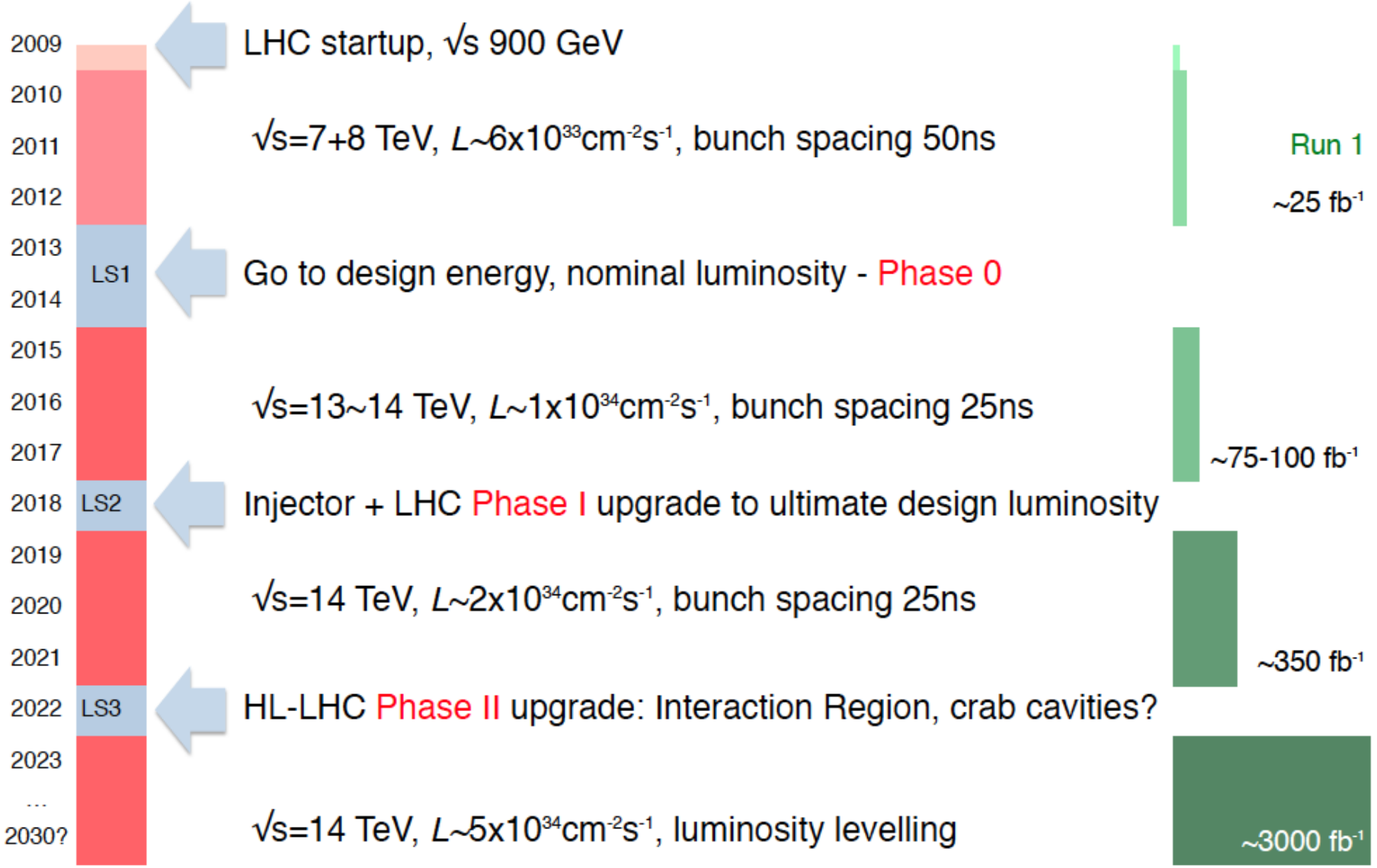
Impact (I)

- Access to cheap ways of implementing low power data transfer and ingestion will mean much more information can be carried and digested than at present.
  - This will enable business and public sector, such as Health, to move data around in much greater volumes than hitherto, so **improving business outcomes** (better model outputs/analysis) and **Health outcomes** (complex imaging and records data can be moved around much more easily).
- Eigensolving large dense matrices quickly and efficiently will have profound consequences as it allows Bayesian statistical methods to be applied to many more areas than hitherto – any diagnosis activity for instance.
- In addition areas in engineering and nanoscience will be able to model much more complex systems with new solvers and hardware solutions.

Impact (II)

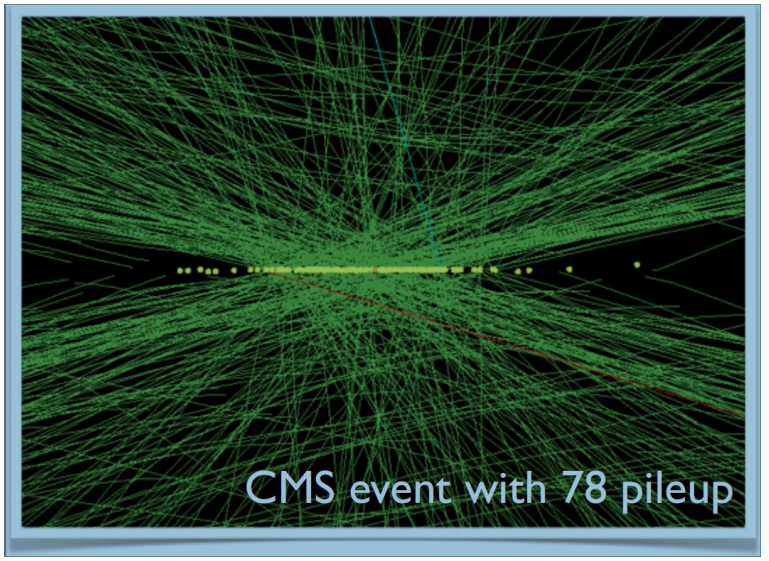
- These will address the following needs:
  - improved competitiveness for companies and SMEs
  - European leadership in applications that address societal challenges
  - benefits for industrial applications through better code performance, maintenance and availability
- The CoE will produce more scientists and engineers trained in the use of computational methods and optimization of applications that can use these high data transfer rate technologies and eigensolving very large matrices.

LHC roadmap to achieve full potential



LHC outlook

- in the coming years:
  - LHC inst. lumin. will increase 10x
  - cross sections for interesting channels will grow significantly with CM Energy
  - experiment event pile-up will go up (up to 5x)

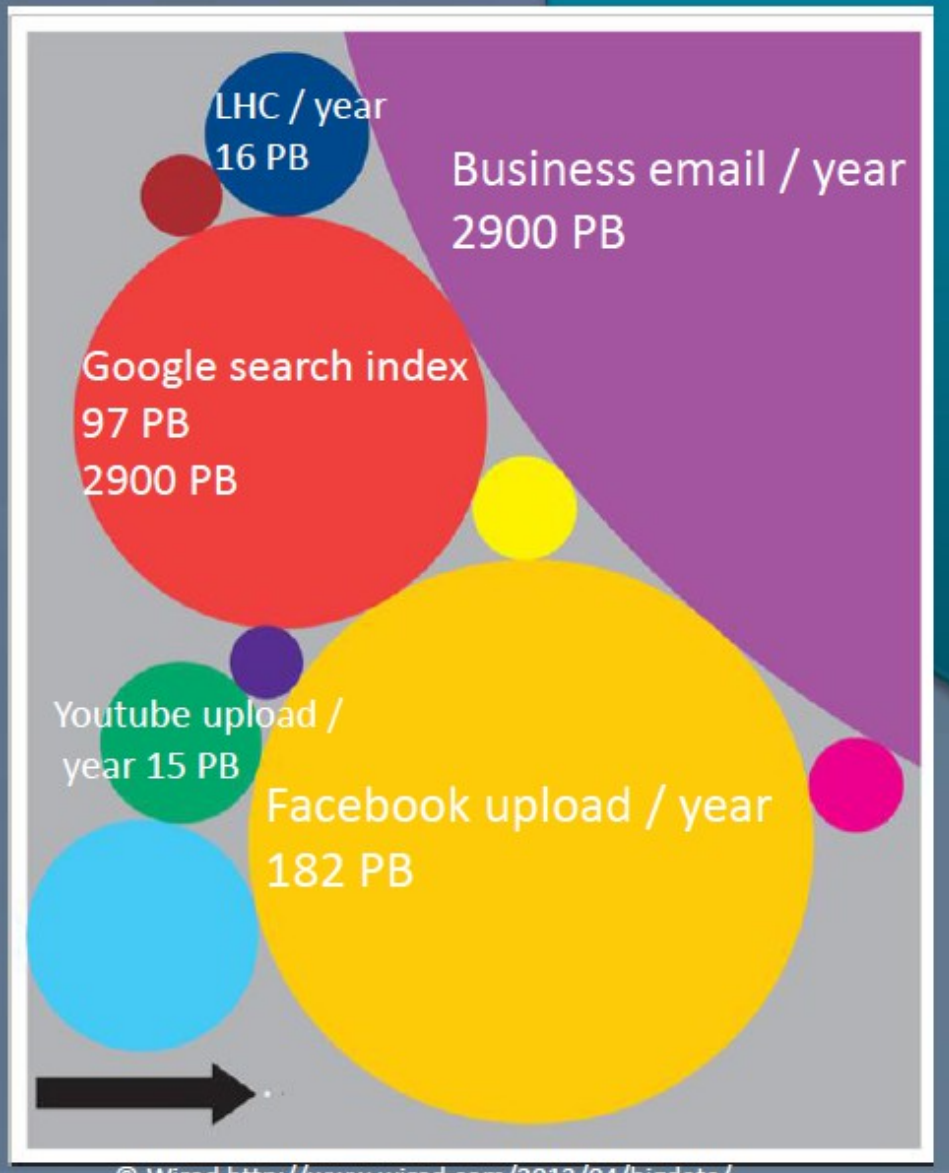


- detector will be upgraded so as to maintain or improve on the present performance
  - a particular challenge is to refine the hardware (level-1) and software real time filters to maintain sensitivity with many interactions per bunch crossing – “pileup”
  - offline algorithms also need to be developed to maintain performance with pileup

Data growth expected

- LHC detector
  - readout rates:
    - now: 0.1 TB/s
    - from 2020: up to 4 TB/s
  - raw data storage rates:
    - now: 1-2 GB/s
    - from 2020: O(10 GB/s)
    - from 2023: O(20-40 GB/s)
- LHC global storage:
    - 2014: 200 PB (disk) + 300 PB (tape)
    - by 2020: 10x increase

Talking about BIG DATA

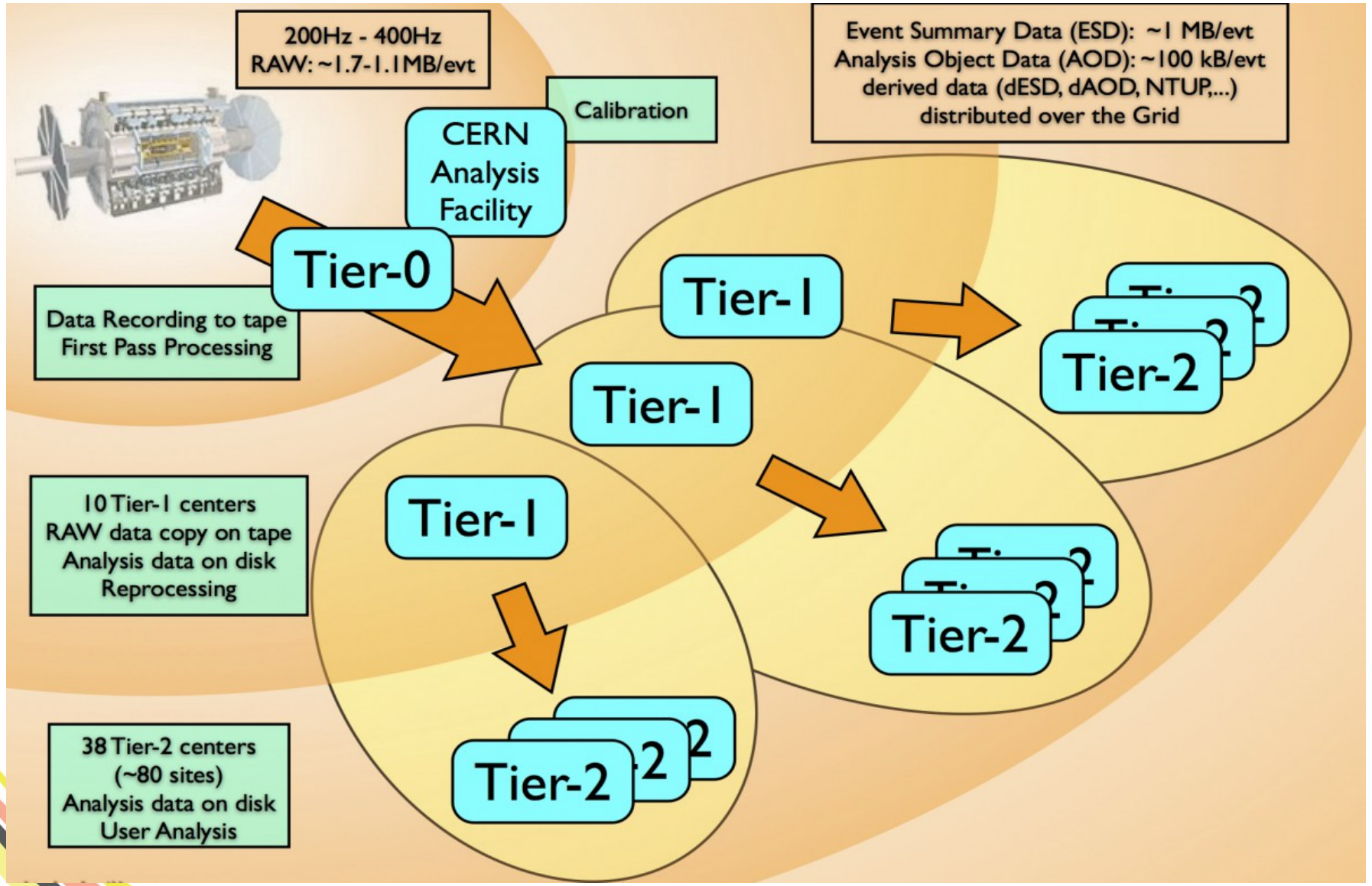


Data processed by the
LHCb software
trigger per year from 2021

19000 PB

N. Neufeld – HL-LHC Workshop Sep. 14

The LHC distributed computing model



LHC and HPC

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- LHC simulation and reconstruction programs require **huge amount of CPU cycles**, at the level of the most demanding HPC applications.
- So far, parallel execution of event processing jobs over the **distributed Grid infrastructure** and the **exponential growth of CPU scalar** performance, have made it possible to exploit very efficiently the evolution of commodity hardware, keeping investments in LHC dedicated data centers under control;
- but for the coming years, the demand for computing resources is expected to grow **beyond the scalar performance increase** delivered at constant costs by the technology evolution .
- **Applications will have to be deeply revised** (if not completely rewritten) for exploiting parallel architectures like the ones employed in HPC centers, more extensively, qualitatively and quantitatively;
- in US the funding agencies are also starting to pushing the LHC collaboration to **use large HPC centers** for running the most CPU demanding LHC codes; exercises of this kind are under way also in Europe.

Data management

- the hierarchical model of distributing and processing the LHC data, originally adopted by the experiments, is now superseded by a more flexible approach whereby the functions are getting more and more interchangeable among the Tier levels
- in addition the relentless growth of network link speed has made it possible to efficiently process remotely located massive amounts of data, using lightweight data access protocols
- this opens the way for the exploitation of HPC centers also for large scale data processing, without the need of locally deploying complex data management systems

LHC selected activities

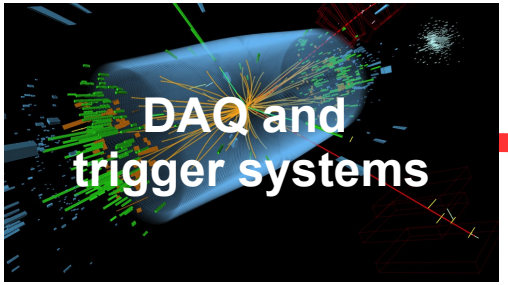
- Activities selected
  - will address important aspects that are:
    - relevant for the future exploitation of the LHC physics potential
    - potentially useful to other communities beyond HEP
    - targeting general aspects, not specific LHC experiment
  - will produce significant contributions exploiting the limited amount of funds that may become available if the CoE proposal is approved

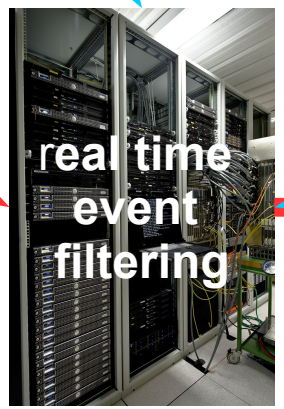
LHC activities

Istituto Nazionale di Fisica Nucleare

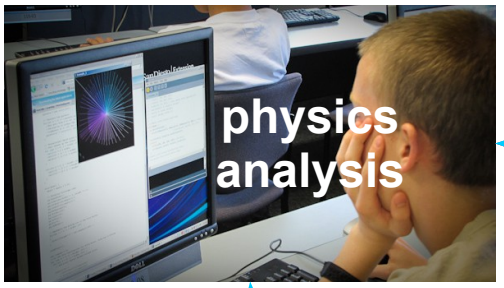
A5 - Extending Performant Detector Simulation to new architectures and physics

A3 - Development of a parallel framework for LHC real time filtering applications





A4 - Performance optimization tools



A1 - Maximum Likelihood Fits on heterogeneous architectures

A2 - Exploitation of HPC clusters for LHC data intensive applications

A1 - Maximum Likelihood Fits on Heterogeneous Many Core Architectures

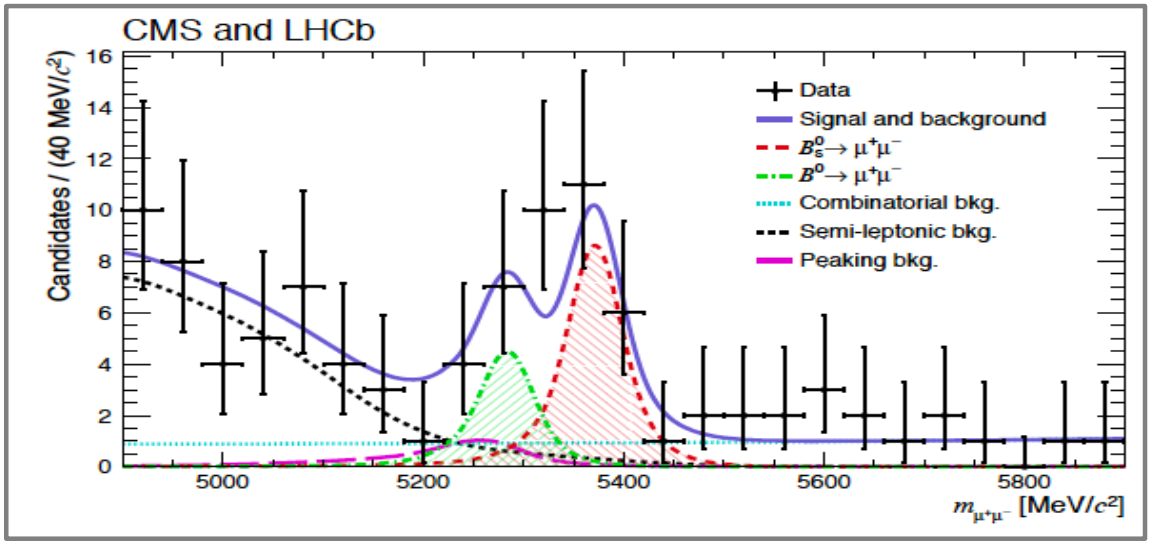
The problem

- The final step of the LHC data analysis consists, in most cases, in extracting the measured physical parameters by a Maximum Likelihood (ML) fit

$$L = \prod_{i=1}^{N} P(X_i \mid \alpha)$$

- computing the Likelihood function is a very CPU-intensive application since:
  - the likelihood function is computed via a loop over a number of events that, as in the case of many recently published LHC analyses, can be of the order of millions;
  - the description of the observed data requires complex Probability Density Functions (PDFs) which are described by a number of parameters to be estimated simultaneously that can be as large as hundreds
-

A ML fit example



PDF of signal and various backgrounds are superimposed and must be disintangled

- The speed depends on
  - Number of events
  - Number of free parameters
  - Complexity of the model

Evolution

- In principle, parallelization of the ML fits is quite straightforward, in particular through event data partitioning;
  - several approaches that help the users by providing interfaces to describe the PDFs in an abstract, target-independent way, have been developed so far;
  - however no general solution addresses the global exploitation of the various level of parallelism available in an HPC cluster of heterogeneous processors;
- the main goal of the proposed activity is thus to develop a modular platform-independent toolkit that fills the current gap, taking into account the current experience and the most common user interfaces;
- this will enable the LHC community, as well as data modeling scientists at large, to dramatically shorten the long analysis turn-around time that characterize the most demanding ML fits.

Timeline

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- study and design of the optimal parallelization patterns and their integration with the most used statistics toolkits (Root, R, ...) (2 FTE x 4 months)
- development and extensive tests of a prototype, implementing the basic functionalities (2 FTE x 3 months)
- integration of the parallelization patterns with the final user interfaces (2 FTE x 5 months)
- tests with real use cases (1 FTE x 3 months)
- dissemination of the results in the LHC community and beyond (1 FTE x 3 months)

A2 - Exploitation of HPC clusters for LHC data intensive applications

HPC for data-intensive computing

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- E-infrastructure in Europe (and worldwide) is converging bringing together HPC and data-intensive computing
  - Already exploratory initiatives for HEP use of HPC* that focus on simulation workflows (minimal I/O)
  - Also considerable current improvements in data and storage in WLCG which add flexibility.
- We will bring this activity together to take the next steps:
  - **Bring workflow tools and approaches to production-level**
    - **Adding high-performance data management solutions**
- To provide a working data/compute solution at HPC centres for LHC and other data intensive (“big data”) communities.

* See for example https://agenda.infn.it/conferenceDisplay.py?confId=8521

Challenges

- HPC: tightly coupled highly-parallel workloads
- HEP workflows can be highly data intensive*
- Differences in computing environment including:
  - Limited network access from computing nodes
  - Limited storage local to compute node, data and software held on shared filesystem
  - Local identity management that may not fit with grid workloads, restrictions on user submit rate, etc.
- These vary depending on HPC centre: some are policy choices – some have technical motivations.

*For example, ATLAS analysis ingested around an exabyte of data in 2013

Goals (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

HEP on HPC workflow development:

- LHC software compatibility: Ensure that simulation software regularly used by the LHC experiments is compatible with execution in an HPC environment.
- Incorporate HPC resources seamlessly into LHC distributed computing systems: Ensure high service availability and reliability, and develop new job scheduling mechanisms to maximise the use of allocated HPC resources.
- Exploit non-x86 based HPC architectures: unlock the significant resource opportunities available on non-x86 compute architectures (e.g. Blue Gene).

Goals (II)

- Data access and management
  - Improvement in the I/O layer of HEP software to make optimal use of limited bandwidth or high latency scenarios.
  - Use of caching, and data-delivery, solutions going beyond current HEP approaches to enable high-performance data access and make effective use of HPC storage.
  - Development of interoperable solutions using standard protocols to enable the use of existing high-volume data storage whether at local or remote sites.

Deliverable and timeline (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Phase 1 (6 months) : LHC workflow development (1 FTE)
  - Validation of LHC software for immediate deployment at participating HPC facilities
  - Development of WLCG middleware to connect HPC resources and collaborate with LHC distributed computing software teams on task allocation algorithms for HPC resources
- Data Access and Management (1 FTE)
  - Enabling key HPC centres to access data from existing LHC data stores. Developing benchmark applications for data access performance and scale testing.

Deliverable and timeline (II)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Phase 2 (1 year):
  - LHC workflow development (1 FTE)
    - Inclusion of HPC sites as a persistent resource for LHC simulation. Design LHC software validation processes to cover all possible HPC site restrictions . Optimization of resource reservation for HPC job scheduling
    - Developments to enable HPC for data-intensive analysis workflows
    - Provide demonstrator of suitable LHC workflow on non-x86 architectures
  - Data Access and Management (1 FTE)
    - Optimising I/O layer of the benchmark applications
    - Deploying data-delivery and caching services at HPC centres to improve data throughput.

Deliverable and timeline (III)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Phase 3 (6 months)
  - LHC workflow development (2 FTE)
    - Production-level service for data-intensive LHC workflows in HPC
    - Enabling interoperation with non-LHC data storage and multi-discipline data-stores

A3 - Development of a parallel framework for LHC real time filtering applications

LHCb HLT

- LHCb will be the first proton collider experiment to implement, at the end of this decade, a completely trigger-less readout system
- Substantial increase in physics reach is only possible with massive increase in read-out rate
  - from ~ 1 MHz to 30 MHz
  - a total of 4 TB/s processed in real time → 0,35 EB/day
  - O(1000) event filtering ratio

Instantaneous luminosity	$2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Pile-up	7.6
Input rate	30 MHz
Maximum processing time per event	13 ms
Output bandwidth	$20\,\mathrm{kHz} \times 100\,\mathrm{kB} = 2\,\mathrm{GByte/s}$

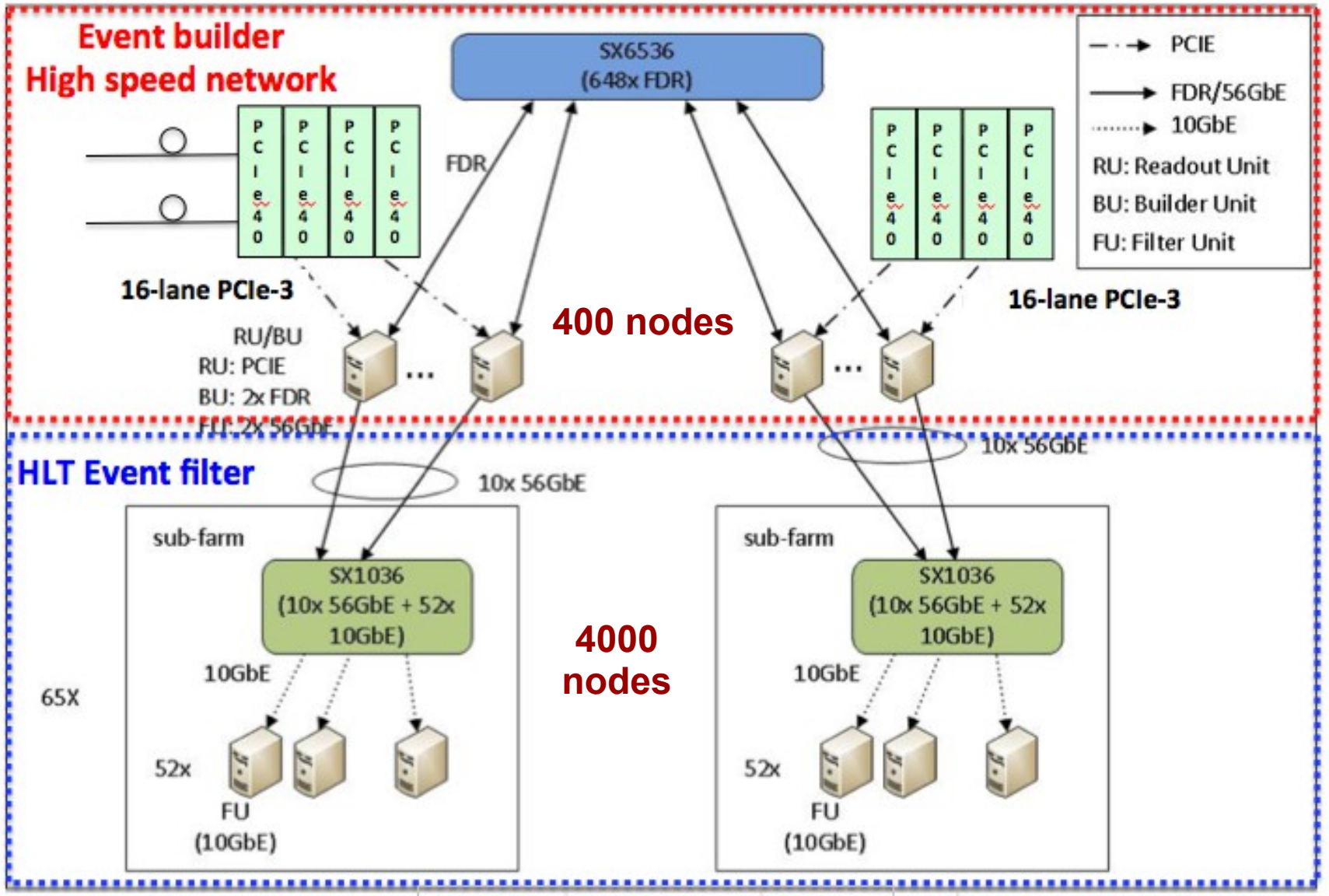
Motivations

Istituto Nazionale
di Fisica Nucleare
Sezione di Padova

- the need to cope with much increased event rates will require in the near future higher bandwidth event building and tighter event filtering using the most sophisticated reconstruction and selection codes, previously employed only offline.
- both functionality will be implemented by very large clusters interconnected by low-latency networks that can be in fact assimilated to large HPC facilities
- since performance per unit of cost is a very critical aspect in these large systems, exploitation of accelerators and co-processors in the filtering systems is expected to grow very rapidly.

The event building and filtering clusters

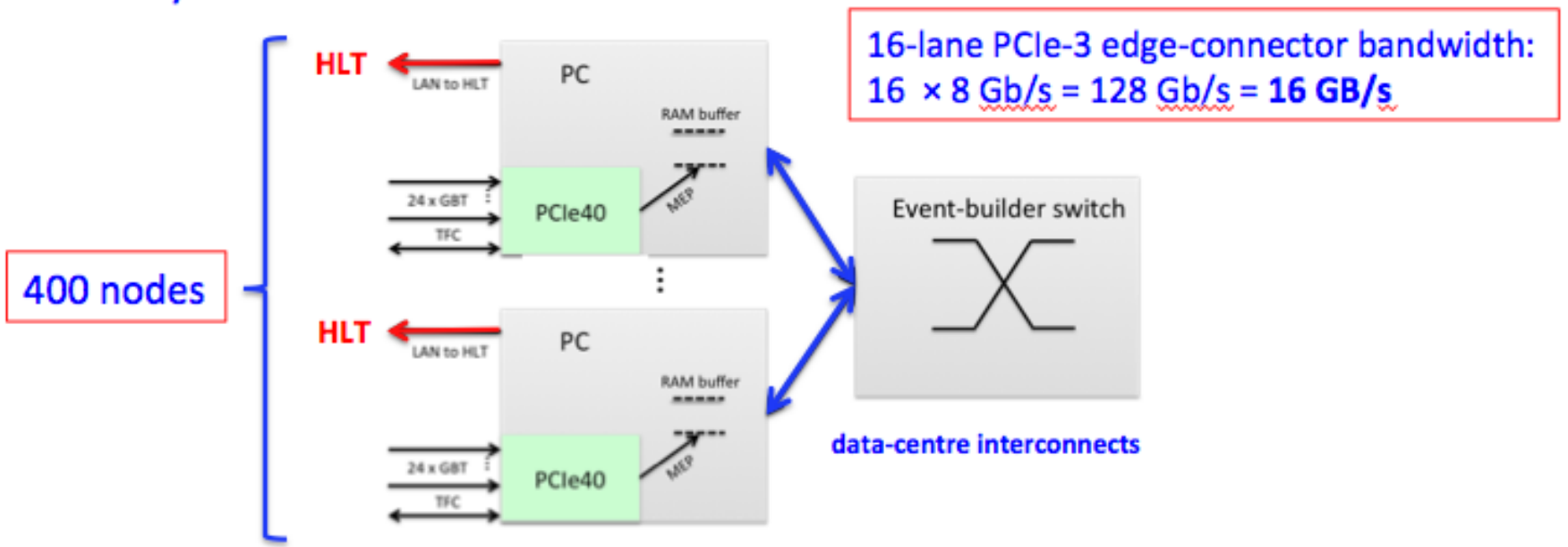
HPC clusters with x86 nodes and Infiniband LAN will be used to simulate the Event Builder



The EB nodes

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Use PCIe Generation 3 as communication protocol to inject data from the FEE directly into the event-builder PC …



- It is a **much cheaper event-builder network** because **data-centre interconnects** can be used on the PC, which are not realistically implementable on an FPGA (large software stack, lack of soft IP cores,…)
- Moreover PC provides: huge memory for buffering, OS and libraries.

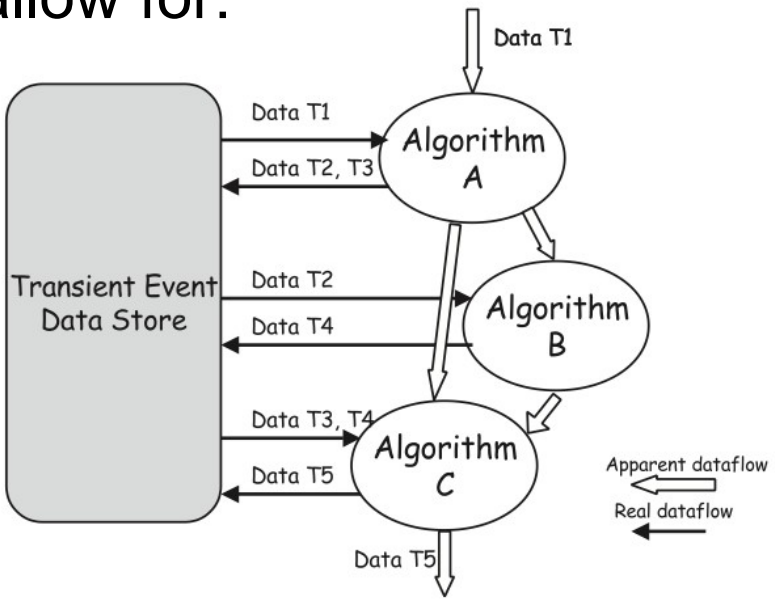
  Up to date network adapter cards and drivers available as pluggable modules.

CoE proposal

- two main aspects of the evolution of the filtering software that are related to parallelism must be addressed:
  - **redesign the algorithms to exploit accelerators, aiming at significantly reducing the global cost of the cluster**
    - the bulk of the work, mostly experiment-specific
  - **modify the current framework that orchestrates the event processing so that data from multiple events can be fed to accelerators and multiple algorithms can be processed in parallel**
    - this part of the work might be re-used also for other applications that have happily enjoyed so far a naïve event data parallelism in multi-core CPU

HEP framework evolution

- experiment event processing applications are based on C++ frameworks that originally did not support concurrency
  - they have been recently modified to manage independent execution of events processing threads on multi-core processors
    - main benefit: much reduced overall memory footprint
- to exploit finer grade parallelism, e.g. via accelerators, more flexible approach is needed to allow for:
  - concurrent execution of algorithms
  - simultaneous processing of multiple events by the accelerators
  - parallelism within the algorithms



A3 – main steps (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- step 1 (1 FTE x 3 months)
  - the parallelization patterns of the real time filtering application will be re-examined, taking into account the specific features of low power heterogeneous processors ;
  - the additional services to be provided by the new parallel framework will be defined;
- step 2 (1 FTE x 6 months)
  - a prototype of the new framework will be developed and iteratively refined by measuring its overall performance on a small scale realistic HPC test-bed, where various accelerators technologies will be employed and compared;
  - the outcome of this study will be crucial for **co-designing the hardware implementation of the final filtering clusters**;

A3 – main steps (II)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- step 3 (2 FTE x 12 months)
  - the full framework will be developed and tested with the new filtering algorithms being rewritten to exploit the fine-grained parallelism of the accelerators
- step 4 (1 FTE x 3 months)
  - the filtering application will be optimized on the final hardware architecture

A4 - Performance Optimization tools

Motivations

- LHC software codes consist of many millions of lines of code and are very CPU intensive.
- Understanding and improving the performance is a crucial enabling component of the evolution of the software and require dedicated and sophisticated tools.
- Experience in LHC with optimization of current applications have led to the development of a profiling tool (IgProf) designed to provide a unique set of desirable features:
  - open source project, extendible to multiple architectures;
  - designed for the bulk of the users who do not have specific computing background and expertise and therefore:
  - compatible with the experiment automated central build and test environments.

IgProf evolution

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- IgProf is being continuously improved and it will be ported soon to the ARM architecture (32bit and 64bit) and Intel MIC.
- What is not foreseen at the moment, and would be most interesting for the employment in HPC clusters, is the possibility to include the capability of profiling applications running on heterogeneous architectures.
- This can be done using the binary instrumentation capabilities of IgProf. There are hooks in common parallel programming frameworks like CUDA, OpenCL and Intel's TBB which can be used to obtain profiling information.
- The results from both the standard IgProf profiling and the parallel programming frameworks can be combined into a single visualization of the profile.

IgProf evolution (II)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- other important developments that could be carried out in the second year are:
  - to combine the information from IgProf (and in principle other profiling tools) with information obtained from a static code analyzer to identify basic problems and opportunities for performance improvements and to provide additional code-context information for profile reports.
  - extend IgProf to support profiling of codes that are written in a combination of Python and Fortran/C/C++
    - a common pattern in scientific software frameworks to use a high-level language (typically Python) as the user application and algorithm building framework and implement the computationally-intensive portions of the problems in underlying Fortran/C/C++ libraries.

Timeline (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- We expect that this activity requires about 24 PM of effort in total. It will proceed in five steps.
- First year:
  - Initial investigations of support profiling capabilities and API's of CUDA, OpenCL and TBB: understanding issues related to implementing the instrumentation and timing issues related to minimizing the impact of the profiling activity itself [1 FTE x 3 months]
  - Based on phase 1, the IgProf design will be updated to include not only profiling on the processor, but also profiling information obtain from the CUDA, OpenCL and TBB. [1 FTE x 6 months]
  - The IgProf analysis and visualization will be updated to permit profile visualization from combined profiles from both processors and CUDA, OpenCL and TBB [1 FTE x 3 months]

Timeline (II)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Second year:
  - IgProf will be instrumented with hooks into the Python interpreter internals to provide sufficient information for construction of a complete and seamless source level profile of combined Python and Fortran/C/C++ applications [1 FTE x 4 months]
  - a small framework (including analysis and web-based visualization tool), based on the Clang/LLVM static analyzer toolkit, will be developed, capable of reading IgProf as well as other profiling tools output; it will be configurable to run one or more static analyses and two types of information will be output: code performance suggestions and code context information [1 FTE x 8 months].
- For this activity collaboration with other communities present in the CoE collaboration would be very welcome

A5 - Extending Performant Detector Simulation to new architectures and physics

The Geant application

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- The Geant series of Monte Carlo simulation tools has represented a de-facto worldwide standard for HENP research.
  - The current generation, the Geant4 toolkit, is in use by nearly all leading HENP experiments
  - Geant4 is capable of transporting particles from sub-thermal energy neutrons to the highest energies of current and planned particle accelerators; the geometrical models of the detector setups can be composed by up to several million volume elements.
- Beyond HENP, the software has been applied in several domains where a detailed particle transport simulation is needed, including:
  - health: radiation treatment planning, medical instrument design and calibration,
  - safety: effect of ionization on living tissues and on DNA, calculation of dose from ionizing particles,
  - radiation protection in space environments for humans and equipment.
  - several industrial processes, e.g.: sterilization of food and medical instruments, industrial material tomography

Geant-related activities

- To exploit the different novel features of existing and emerging processors requires an extensive redesign, reengineering and rewrite of the simulation transport code and all its constituent parts, using novel algorithmic and computing techniques and models.
  - exploratory work with a specific prototype (GEANT-V) has already started and will define the basic architectural evolution of the toolkit
- the employment of new technologies will also bring important benefits not only for the HENP community but also for extending the usage of Geant in related fields, including industries

Geant-related activities

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- within the CoE we propose to address two topics that are not only important for the HENP community,making it feasible to bring large simulations up to the HL-LHC scale, but that could also enhance the usability of the Geant toolkit in embedded and mobile HPC systems.
- the two proposed developments are:
  - 1. porting Geant-V foundation libraries to the ARM architecture; this activity will be aimed at addressing the quest for achieving higher energy efficiency
  - 2. developing a new module for low-energy hadronic physics simulations that can effectively exploit accelerators and heterogeneous architectures; this will make it possible to deploy applications in the industrial and medical sectors capable of providing the requires outcome in real-time.

Timeline (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- A. Expanding the underlying libraries to new architectures and testing with existing kernels (1 FTE x 2 years)
  - (T1+3 months) Test, validate, and refine/extend port of Vc library to ARM architecture
  - (T1+6 months) Enable ARM architecture for optimised geometrical calculations
  - (T1+12 months) Enable ARM architecture for optimised electromagnetic physics processes
  - (T1+15 months) Create port of Vc library to AltiVec instruction set, and first evaluation with small geometry kernels
  - (T1+21 months) Report on extensive validation and benchmarking of the Vc ports (ARM & AltiVec) and of the physics routines using these ports
  - (T1+22/23 months) Disseminate results

Timeline (II)

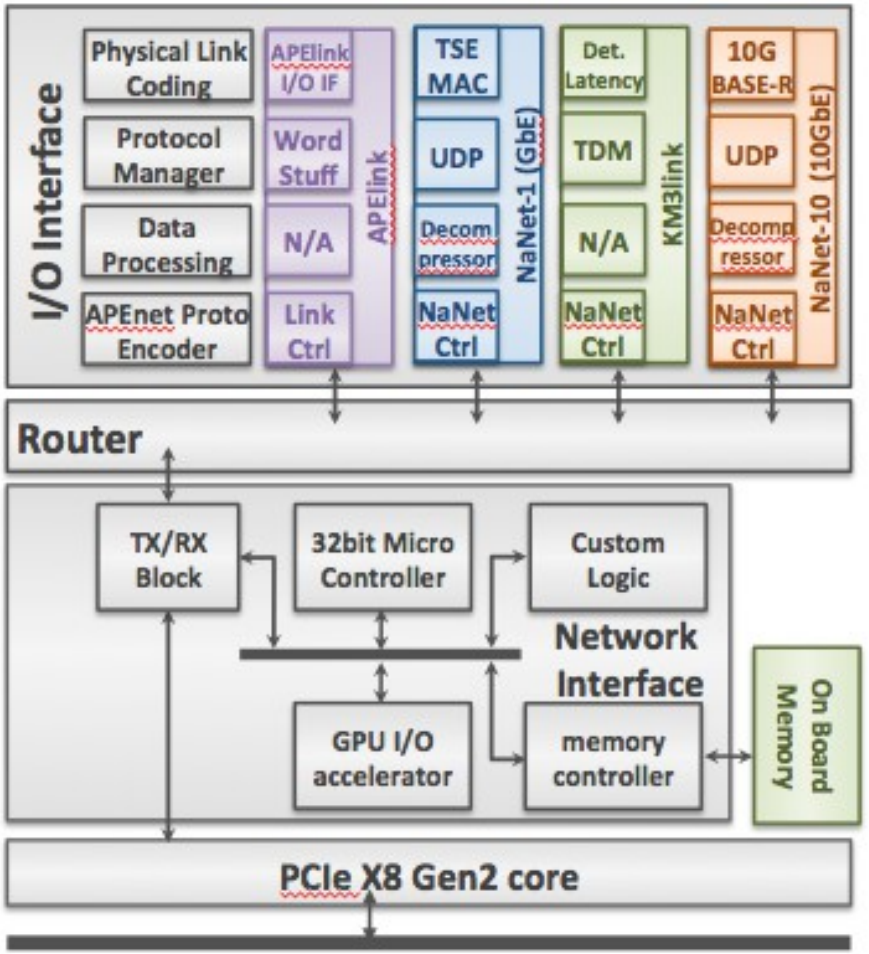
Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- B. Development of low-energy hadronic processes with generic programming approach (scalar/vector/ ….) including low-energy interactions of neutrons, protons and pions (1.5 FTE x 2 years)
  - (T2+6 months) Port pre-compound physics model to generic vectorised kernel approach
  - (T2+9 months) Port evaporation physics models to generic kernel approach
  - (T2+15 months) Integrate optimised code into existing Geant-V framework
  - (T2+18 months) Report on extensive benchmarks and comparison on different architectures (Intel Xeon, Intel Phi, Nvidia GPUs)
  - (T2+21 months) Report on additional benchmarking with ARM and AltiVec port of Vc
  - (T2+22/23 months) Disseminate results

Additional EoI to participate in WP5

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

NaNet: a FPGA-based NIC design for Real-Time GPGPU Systems



- Multiple link technologies
  - Standard: **1GbE** (1000Base-T), working on **10GbE** (10Base-R).
  - Custom: **APElink** (>20 Gb/s QSFP), **KM3link Det. Lat.** (2.5 Gb/s optical).
- Network protocol offloading
  - **UDP**, **Time Division Multiplexing**.
- Processing on data stream
  - e.g. rearrange event data (size and alignment) in a GPU-friendly style.
- Trigger and timing distribution system with HSMC daughtercard.
- GPUDirect P2P/RDMA
  - Direct data transfer between network and GPU memory (no bounce buffers on CPU memory).
- PCI Express x8 Gen2
  - Working on **PCIe X8 Gen3.**

Opportunities for collaborations

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- Co-design approach to real-time data transport for GPU-based systems proved to be effective:
  - Achieved low and stable data transfer latency.
- Modular FPGA-based design allows for a fast and straightforward integration in new experimental contexts:
  - Implemented in short time KM3Net-IT on-shore NIC with four optical deterministic latency channels (NaNet^3).
  - HSMC expansion cards add functionalities needed by the experiment (e.g. TTC card)
  - FPGA resources available to perform processing on data streams before they are transferred to GPU memory.
- Eager to proof our design in other contexts and wish to collaborate (maybe a commodity network is OK for you, maybe not…).

For more info: http://apegate.roma1.infn.it

Manpower estimate

- first order estimate of the required efforts in Person-Month (PM)
  - LHC (total = PM)
    - A1: 36 PM
    - A2: 50 PM
    - A3: 36 PM
    - A4: 24 PM
    - A5: 60 PM
  - SKA (total = PM)
    - A6: 48 PM
    - A7: 40 PM

Expected impact of CoE LHC activities

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

- **improved access** to computing applications and expertise:
  - all activities will bring **significant benefits** to the user communities involved
- **improved competitiveness** for companies and SMEs:
  - developments will be carried out **in partnership with industries** involved both in computing hardware as well as in application developments in various fields, from “big data” analytics to industrial material tomography
- **European leadership** in applications that address **societal challenges**:
  - "Big data" has the potential to revolutionize the approach to many societal challenges; specific application domain important for the society (e.g.: health) will benefit from developments of improved algorithms and applications
- **benefits for industrial applications** through better code performance, maintenance and availability
  - Obtaining excellent code performance through exploitation of new processor architectures will represent the most important goal in most activities
- more **scientists and engineers trained** in the use of computational methods and optimization of applications
  - HEP scientists will increase their **expertise in parallelization** that will be critical to application performance. Other communities will be exposed to **data management expertise**, helping to address the growing data challenges of their applications

Activities mapping (I)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

A1	**Maximum Likelihood Fits on Heterogeneous Many Core Architectures**	**WP1**
A2	**Exploitation of HPC clusters for LHC data intensive applications**	**WP1**
A3	**Development of a parallel framework for LHC real time filtering applications**	**WP5**
A4	**Performance Optimization tools**	**WP3**
A5	**Extending Performant Detector Simulation to new architectures and physics**	**WP1**
A6	**Real time Data ingestion**	**WP5**
A7	**Pseudo-real time data calibration using the SKA as an exemplar**	**WP1/WP2**
WP1	Community software development and data management	
WP2	New algorithms for large-scale simulation and data analytics	
WP3	Languages, programming models and tools for heterogeneous/HPC/Exascale architectures	
WP5	High-speed real-time data processing	
WP6	Dissemination and training	

Activities mapping (II)

Istituto Nazionale di Fisica Nucleare
Sezione di Padova

A1	**Maximum Likelihood Fits on Heterogeneous Many Core Architectures**		**WP1**
A2	**Exploitation of HPC clusters for LHC data intensive applications**		**WP5**
A3	**Development of a parallel framework for LHC real time filtering applications**		**WP5**
A4	**Performance Optimization tools**		**WP3**
A5	**Extending Performant Detector Simulation to new architectures and physics**		**WP1**
A6	**Real time Data ingestion**		**WP5**
A7	**Pseudo-real time data calibration using the SKA as an exemplar**		**WP2**
WP1	Community software development ~~and data management~~		
WP2	New algorithms for large-scale simulation and data analytics		
WP3	Languages, programming models and tools for heterogeneous/HPC/Exascale architectures		
WP5	Data intensive processing and management		
WP6	Dissemination and training		