GRID-based off-line infrastructure for the PADME experiment at the DAFNE BTF

Alessandro De Salvo¹, Georgi Stefanov Georgiev², Emanuele Leonardi³, Francesco Safai Tehrani³, Elisabetta Vilucchi³
¹INFN Roma, ²University of Sofia, ³INFN LNF

The Dark Photon

A possible solution to the Dark Matter problem is the hypothesis that it interacts with the SM through fields only via "portals" which link our world to the "dark" world. The simplest model postulates the existence of a U(1) symmetry with its corresponding A' vector boson: SM particles are neutral under this symmetry, so that this field, thanks to the possible mixing with the photon, would couple to the SM only with an effective charge e'. For this reason it is often called «dark photon».

The PADME Experiment

PADME (Positron Annihilation into Dark Matter Experiment), approved by INFN in 2015, will look for invisible A' decays using the Beam Test Facility (BTF) line of the Linac Accelerator at the INFN Frascati National Laboratory (LNF).

In the first year of run PADME will collect a total of $10^{15}$ e' on target corresponding to 0.9PB s⁻¹. DAQ consists in a total of 900 ADC channels, for a total DAQ rate of ~80MB/s, and will produce ~30TB of raw data which will include all ADC samples (after applying zero suppression). Data will be written to a 4TB local RAID disk buffer, then moved to the WLCG Tier in LNF, and copied to the tape library at INFN-CNAF (Bologna). An emergency copy of the data will also be kept on the KLOE2 tape library at INFN. The local disk buffer allows data taking even in absence of a network connection, up to a maximum of ~2 weeks.

Event Reconstruction

Both DAQ and MC events will be fully reconstructed. The reconstruction procedure is still at a preliminary stage, so it is not possible to have a precise estimate of the CPU needs and the total amount of storage generated. Given the relatively simple structure of the experiment and the fact that reconstructed events will not include the full set of ADC samples, we expect the CPU needs to be a fraction of that for MC production, and storage to be roughly 1/100 of the raw data.

GRID infrastructure

To support all the processing and storage needs of the experiment we developed a GRID-based infrastructure. This allowed access to distributed resources with a homogeneous interface. Ressources used included:

- CPU farms in INFN-LNF, INFN-CNAF, and Sofia University
- Disk storage systems in INFN-LNF and INFN-CNAF
- Tape libraries in INFN-CNAF and INFN-LNF

All data (Raw, MC, Recco) will be saved to tape. All Recco data will be kept on disk for data analysis, while MC and Raw data will reside on disk only for the time strictly needed to reconstruct them.

PADME software is distributed over the GRID using CVMFS (hosted at INFN-CNAF).

Conclusions

- In terms of available man power, PADME is a small experiment, but the amount of data produced and the computing needs are not negligible and will benefit from a GRID-based infrastructure.
- The initial learning curve of the GRID is quite steep: information is somewhat scattered around and one has to rely on help from GRID site managers and knowledgeable people from other (mainly LHCC) experiments.
- A "GRID 101" (aka "GRID for Dummies") manual would greatly improve the situation.
- A few basic GRID-related services (file manager, job submission manager, data transfer manager) were missing or were too complex to be used in a low manpower environment. A simple version of these services was therefore implemented within the collaboration. As all these services are very likely needed by any experiment, with very small variations, creating easily customizable standard versions of these packages should not be difficult and would greatly reduce the time for the initial GRID-ification of an experiment.

REFERENCES