All grown-up; 18 years of LHC@home

David Cameron[1,*], *Laurence* Field[2,**], *Frederik* Van der Veken[2,***], *Nils* Høimyr[2,****], *Davide* Di Croce[3], *Melissa* Gaillard[2], *Germano* Masullo[2], *Cath* Noble[2], *Ignacio* Reguero[2], *Ivan* Reid[4], and *Ben* Segal[2]

[1]Department of Physics, University of Oslo, Oslo; Norway
[2]CERN, Geneva, Switzerland
[3]Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
[4]Brunei University, UK

Abstract. LHC@home was launched as a BOINC project in 2004 as an outreach project for CERN's 50 years anniversary. Initially focused on the accelerator physics simulation code SixTrack, the project was expanded in 2011 to run other physics simulation codes on Linux thanks to virtualisation. Later on the experiment and theory applications running on the LHC@home platform have evolved to use containers and take advantage of the CVMFS file system as well as content delivery networks. Furthermore, a substantial part of the contributed computing capacity nowadays is provided as opportunistic back-fill from data centers with spare capacity, in addition to enthusiastic volunteers. The paper will address the challenges with this distributed computing model, new applications to exploit GPUs and the future outlook for volunteer computing.

1 Introduction

Physicists and other scientists have always been hungry for computing power. In addition to the processing and analysis of experimental data from the LHC detectors as well as other experiments at CERN, the computing facilities at CERN and the WLCG also need to provide capacity for simulations and upgrade studies. Given the high load on the batch computing services at CERN, an effort started in 2003 to take advantage of idle capacity of office desktops with the "CERN Physics Screen Saver" (CPSS)[1]. CPSS ran the accelerator beam simulation application SixTrack[2] on CERN desktop computers when idle, and hence provided extra simulation capacity for accelerator physics studies.

At the time, there had been several similar efforts to harness idle capacity from home and office computers, notably Seti@home[3] and ClimatePrediction.net[4]. These research projects harvest computing capacity from volunteers, e.g. donors among the general public. Volunteer computing is suitable for handling massive simulations on small data sets, and these developments led to a common framework; the Berkeley Open Infrastructure for Network Computing, popularly known as BOINC [5, 6].

---

[*]e-mail: david.cameron@cern.ch
[**]e-mail: laurence.field@cern.ch
[***]e-mail: frederik.van.der.veken@cern.ch
[****]e-mail: nils.hoimyr@cern.ch

In 2004, for CERN's 50 year anniversary, SixTrack was ported to BOINC. Now also the general public could contribute CPU, and LHC@home[7, 8] was born.

BOINC is designed as a distributed client/server application, with one or several Web servers to handle scheduling, task and user management. One BOINC server is setup as a BOINC project. For outreach and communication with volunteers, the web front-end includes a set of message boards. The server implementation is written in a mixture of C and C++ for the scheduler, clients and validator, while the web user pages are PHP-based. Python scripts are also used for some components, and the underlying data for BOINC is stored in a MySQL[9] database. BOINC clients are available for Windows, Mac, Linux and Android. The BOINC server distributes applications and tasks to the computer hosts belonging to volunteers who subscribe to the projects. With BOINC, volunteers are awarded (virtual) credit for their contributions.

The outreach events and publicity around the 2004 anniversary attracted many volunteers, and hence LHC@home soon provided serious computing capacity.

Since the launch of LHC@home, the beam dynamics simulation application SixTrack[8] has accumulated thousands of CPU-years thanks to the contributions of volunteers running on Windows, Linux and Macs.

Due to lack of resources in CERN IT, LHC@home was operated from Queen Mary University from 2008 to 2011, in collaboration with EPFL and the accelerator physics team at CERN.

The original LHC@home was in 2011 joined by Test4Theory, that pioneered the use of virtualization for volunteer computing, thanks to the use of CernVM[10] to launch a Linux Virtual Machine[11] on the volunteers computers. To run virtualized applications, a local hypervisor, the free Oracle VirtualBox[12] can be installed with BOINC. When contacting the BOINC server, BOINC downloads a wrapper application "vboxwrapper" [13] that takes care of launching the Virtual Machine (VM). This advance led CERN IT to launch a renewed effort on volunteer computing [14], and soon more BOINC pilot projects based on the experience with Test4Theory followed with applications from the LHC experiments. LHCb, ATLAS and later CMS.

Following the successful pilot projects with LHC experiment applications running under BOINC, it was decided to consolidate all of these separate BOINC projects as different applications under a single BOINC project, LHC@home. Hence in 2016, the forums and databases with volunteer credit from Test4theory and the other pilot projects were merged with the orginal LHC@home that has hosted SixTrack since 2004. Volunteers were able to choose whether they wanted to run multiple applications, where e.g. participants in LHC experiment collaborations could choose the application from their experiment. This approach was favoured by volunteers, and gave a more coherent brand for LHC@home.

2 BOINC applications at CERN

At CERN, several teams provide applications that run under LHC@home, a brief description of the applications and their infrastructure is provided here.

2.1 Sixtrack

SixTrack is a particle tracking code for simulation of particles in a high energy particle accelerator. Sixtrack has been running on LHC@home under BOINC since the start of the project in 2005, and is the only application on LHC@home that is compiled with BOINC for

Windows, Mac and Linux as a native BOINC application. There has been continuous development of SixTrack[15], both for HL-LHC, and future accelerator studies like the FCC. Furthemore a new build system to produce executables for more computer architectures was set up in 2016. Hence an Android test version of SixTrack was made available on LHC@home in 2017. During major simulation campaigns with SixTrack, there have been peaks of 500k tasks in progress.

2.2 Garfield

The Garfield[16] computer program for simulation of gaseous detectors was available as a native BOINC application for Linux and Windows (Cygwin) on LHC@home from 2006 to 2008.

2.3 Theory

The Theory simulations application on LHC@home runs Monte Carlo simulations of physics events in a virtual machine under CernVM. The results are gathered on MCPlots[17], a repository of Monte Carlo data. The jobs running in the VMs dispatched with BOINC are handled via HTCondor[18] and the application pulled from the CERN VM file system (CVMFS)[19].

2.4 ATLAS

The ATLAS application [20] runs Monte-Carlo simulation of events in the ATLAS detector. Jobs are submitted to the BOINC server using the ARC CE [21] grid middleware. In 2017 a native Linux version of the ATLAS application was developed which removes the need for virtualisation on Linux platforms, instead using Singularity [22] containers. (Singularity later was renamed to Apptainer.) This opened the door to ATLAS jobs running on data centre infrastructures such as new machines not yet commissioned or even grid worker nodes where BOINC can run as a "backfill" when the grid jobs are not fully using CPU. These new resources have hugely increased the amount of work done by the ATLAS application, and in 2022 it processed roughly 5% of all ATLAS simulation events, of which 85% were processed by the native Linux app.

2.5 CMS

Initially a proof of concept[23] in 2014, the CMS@home application was developed into a production quality platform[24]. Like other high energy physics applications, the CMS software stack runs on Linux and hence Virtualization is required on volunteer computers running Windows and MacOS. BOINC is used to distribute a wrapper CMS application, that will run a virtual machine for a fixed period of time. From the CMS perspective, the VM is similar to another VM in a compute cloud resource pool. Once running, the virtual machine starts an agent (pilot) to obtain (pull) work from a central queue of CMS tasks. The agent and central task queue has been implemented using HTCondor and the output data is uploaded to a databridge, which acts as a sandbox between the volunteer and grid storage. The authentication mechanism used by both HTCondor and the databridge is based on x509[25] credentials from a certificate authority set up for the purpose. In 2023, this mechanism has been extended to support token based authentication.

2.6 LHCb

A first prototype of the Beauty@Home project was created in 2013[26] to help address LHCb's growing need for computing power by profiting from volunteer computing resources. Access to Beauty@home was initially restricted to members of the LHCb Virtual Organisation as a trusted x509 credential was required on the resource to contact the DIRAC system[27] for operations such as job matching, status updates and data management operations. BOINC used to dispatch applications to volunteers, and the application wrapper started a CernVM Virtual Machine. Within the VM, the application triggered a standard LHCb pilot job which interacted with the DIRAC system to obtain the real job and upload the output. To allow volunteers who did not have LHCb affiliation and credentials, a DIRAC gateway service called Workload Management System Secure Gateway (WMSSecureGW) was developed to act as a sandbox for untrusted resources to the DIRAC System, and hence allowed VMs on running volunteer computers to execute LHCb jobs. Following this development, the LHCb application was integrated with LHC@home and volunteers were able to run LHCb simulations until the experiment decided to pause submissions in 2018 due to lack of resources to maintain the DIRAC gateway.

2.7 Xtrack

Xtrack/Xsuite is a new generation of modular beam tracking code replacing SixTrack. It can also run on GPUs, and is currently being tested on BOINC. With Xtrack a wider range of beam simulations can be submitted to LHC@home, and hence take advantage of volunteer capacity, also on GPUs.

3 Outreach

Outreach to attract and retain volunteers is an essential part of volunteer computing. For this purpose, BOINC provides message boards that are complemented by a project web site. To improve communication among scientists and volunteers, the LHC@home team organised the first "Pan-Galactic BOINC workshop" at CERN in 2005. 2 more workshops were organised in a joint effort with the University of Geneva in 2006 and 2007.

Later on, to improve interaction with volunteers, the LHC@home portal was extended with a Drupal web site. An improved LHC@home web site with Frequently Asked Questions (FAQ) and education material was launched in 2014. The web site has undergone several upgrades, notably for the LHC@home consolidation in 2017[28] and upgrade to Drupal 8 in 2019.

Furthermore video tutorials and other publicity material was provided, notably for CERN visitors.

A screen with a promotion video as well as a wallside banner were added outside the CERN Data Centre Visit Point in building 513, so that visitors to CERN IT are made aware of LHC@home.

Major contributors of CPU and other helpful volunteers who have invested time to test applications and help others on the BOINC forums were awarded with polo shirts and CERN "Thank You" cards from the LHC@home project team.
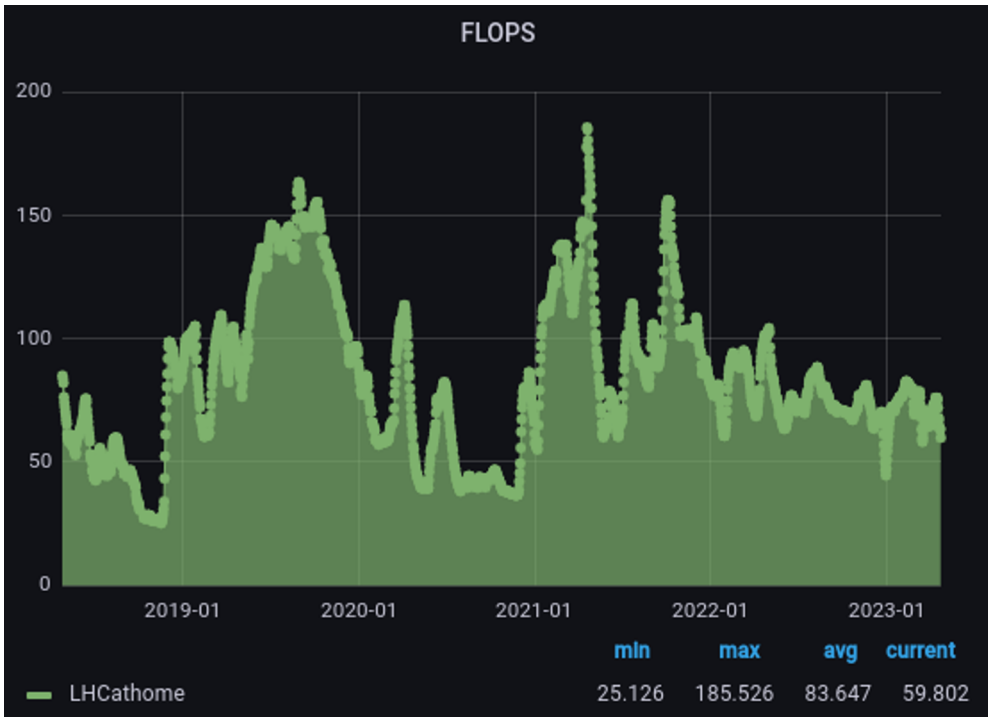
4 Statistics and computing infrastructure

4.1 Jobs and statistics

Over the last years, the number of jobs running on LHC@home has fluctuated depending notably on SixTrack simulation campaigns, that tend to come in batches. (See Figure 1.)
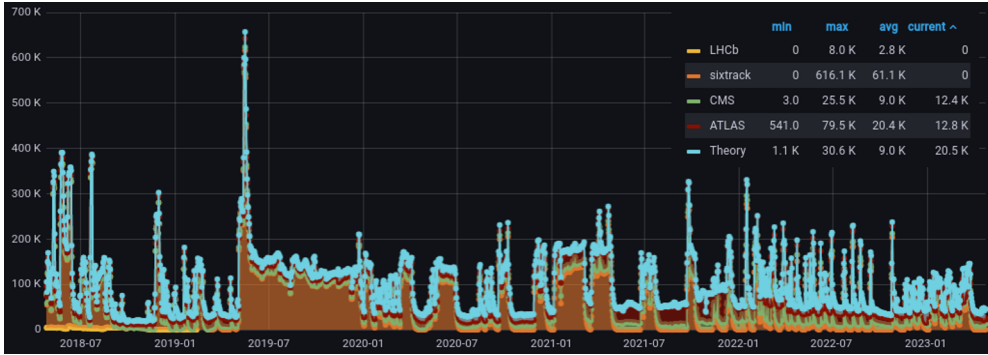
Figure 1. Running tasks per application over time. (Peak in May 2019 corresponds to the BOINC Pentathlon.)


Figure 2. Compute capacity in teraflops.

Overall, the average computing capacity contributed to LHC@home averages more than 80 TFlops, similar to a Tier-2 Grid site. (See Figure 2.) There are peaks well above, notably during volunteer competitions like the yearly "BOINC Pentathlon" that included LHC@home in 2015, 2017 and 2019. (See Figure 1.)

4.2 Servers, compute capacity and backfill

The BOINC server infrastructure for LHC@home has evolved from a single web server with use of AFS as a shared environment for job submission in the early days to a fully distributed server environment running on VMs in CERN's OpenStack private cloud.[28] The distributed environment with multiple file servers has proven to scale well under high load, and also the Cephfs [29] storage backend could handle this I/O well. 2 web schedulers and 3 file servers have been used most of the time, scaled up to 3 schedulers and 6 file servers during high load periods.

Such high load periods were typically the BOINC Pentathlons mentioned earlier, but more recently also when ATLAS added BOINC as a source for temporary backfill job submission to HPC centers and Tier-2 sites. (See Figure 3.) The use case for BOINC for backfill has advantages with the minimal client setup required on the remote site. Hence for periods with spare capacity on local compute clusters, adding these as BOINC resources can provide significant increase of capacity beyond what is contributed by volunteers.
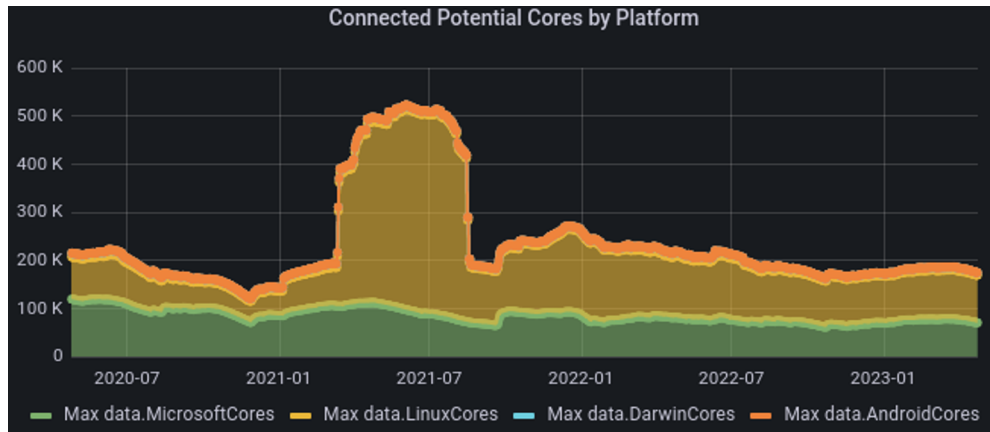
Figure 3. Evolution of number of cores connected to BOINC versus time. The large number of Linux cores in 2021 corresponds to temporary ATLAS job backfill from the running-in period of the VEGA HPC supercomputer in Maribor, Slovenia.

5 Activities with the BOINC community

An important effort has been done to strengthen the community of BOINC developers and project administrators. As part of this, the LHC@home maintainers at CERN have contributed towards the establishment of a new BOINC governance model, and helped to organise the BOINC workshop in 2017 in Paris[30] dedicated to this topic.

Furthermore, a member of the LHC@home team has taken responsibility for server releases and done a major effort to organize the developer community around the BOINC Github project[31].

6 Conclusions and outlook

During more than 18 years, volunteers have provided significant computing capacity for simulations of particle beams with SixTrack as well as other simulations for experiments and

theory in High Energy Physics. This additional computing capacity has allowed teams of researchers to advance further with their work than what would have been possible with available batch computing capacity at CERN. LHC@home has also been an outreach channel to the general public. The evolution from desktop to portable computing devices has led to stagnation in the number of active volunteers. However GPUs on game PCs and game consoles have a huge potential for applications that can run on GPUs, and we expect that the GPU-enabled Xtrack application will give a boost to LHC@home for the future years.

6.1 Acknowledgments

References

[1] CERN Physics Screen Saver: Help LHC tracking studies on your PC. https://cds.cern.ch/record/788161?ln=en

[2] SixTrack: http://sixtrack.web.cern.ch/SixTrack/ (Accessed September 2023.)

[3] Seti@home URL: http://setiathome.ssl.berkeley.edu/ (Accessed September 2023.)

[4] Climateprediction.net URL: http://www.climateprediction.net/ (Accessed September 2023.)

[5] D. Anderson, 2004. Proceedings of Fifth IEEE/ACM International Workshop on Grid Computing, pp. 4–10

[6] BOINC URL: http://boinc.berkeley.edu/ (Accessed September 2023.)

[7] LHC@home portal at CERN. URL: http://lhcathome.cern.ch (Accessed September 2023.)

[8] M. Giovannozzi et al. "LHC@home" 2012 International Particle Accelerator Conference.

[9] MySQL URL: http://www.mysql.com/ (Accessed September 2023.)

[10] P. Buncic et al. 2010 Journal of Physics: Conference Series, 219, 042003

[11] J.E. Smith and R. Nair, 2005 Virtual Machines, Morgan Kaufmann, Elsevier

[12] Oracle VM Virtual Box URL: https://www.virtualbox.org/ (Accessed September 2023.)

[13] Vboxwrapper: URL: http://boinc.berkeley.edu/trac/wiki/VboxApps (Accessed September 2023.)

[14] N. Høimyr et al. 2012 Journal of Physics: Conference Series, 396, 03205

[15] R. De Maria et al. "Recent developments and future plans for Sixtrack" 2013 International Particle Accelerator Conference.

[16] R. Veenhof et al "Garfield, a drift chamber simulation program" Conf. Proc. C, vol. 9306149 p 66-71 1993

[17] A. Karneyeu, L. Mijovic, S. Prestel et al. Eur. Phys. J. C (2014) 74: 2714. https://doi.org/10.1140/epjc/s10052-014-2714-9

[18] D. Thain, T. Tannenbaum, and M. Livny, "Distributed Computing in Practice: The Condor Experience" Concurrency and Computation: Practice and Experience, Vol. 17, No. 2-4, pages 323-356, February-April, 2005.

[19] J. Blomer et al. "Distributing LHC application software and conditions databases using the CernVM file system" 2011 J. Phys.: Conf. Ser. 331 042003

[20] D. Cameron et al. ATLAS@Home: Harnessing Volunteer Computing for HEP (CHEP-2015)

[21] A. Filipčič et al. 2012 Journal of Physics: Conference Series, 396 032039

[22] G.M. Kurtzer, V. Sochat, M.W. Bauer, PLoS ONE 12, e0177459 (2017)

[23] L Field and H Borras and D Spiga and H Riahi, CMS@home: Enabling Volunteer Computing Usage for CMS 2015 Journal of Physics: Conference Series, 664, 022017

[24] L. Field, D. Spiga, I. Reid, H. Riahi and L. Cristella, CMS@home: Integrating the Volunteer Cloud and High-Throughput Computing Computing and Software for Big Science 2018 2 2

[25] Internet X.509 Public Key Infrastructure https://tools.ietf.org/html/rfc4158 (Accessed September 2023.)

[26] J. Barranco, Y. Cai, D. Cameron et al. (2017). LHC@Home: a BOINC-based volunteer computing infrastructure for physics studies at CERN. Open Engineering, 7(1), pp. 379-393

[27] V. Garonne, A. Tsaregorodtsev, I. Yu Stokes-Rees, DIRAC : Workload Management System Computing in High Energy Physics and Nuclear Physics 2004 1041-1044

[28] D. Cameron, L. Field, N. Giannakis, N. Høimyr Extending CERN computing to volunteers https://doi.org/10.1051/epjconf/201921403016

[29] S. Weil et al. 2006 In Proceedings of the 7th symposium on Operating systems design and implementation (OSDI '06). USENIX Association, Berkeley, CA, USA, 307320.

[30] BOINC workshop 2017: https://indico.cern.ch/event/648533/ (Accessed September 2023.)

[31] BOINC GitHub repository: https://github.com/BOINC/boinc (Accessed September 2023.)