Available on CMS information server **CMS CR -2025/059**

The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

23 February 2025 (v2, 26 February 2025)

R and D Adoption and Progress in Full Simulation of the CMS experiment

N. Srimanobhas, S. Banerjee, F. Cossutti, V. Ivantchenko, N. Krammer, M. S. Muzaffar, K. Pedro, and L. Pezzotti
for the CMS Collaboration

Abstract

In this work we report on evolution of usage of Geant4 within CMSSW and adaptation of the newest Geant4 11.2.1, which is expected to be used for CMS simulation production in 2025. Physics validation results and results on CPU performance are reported. For the Phase-2 simulation several R&D are carried out. A significant update for CMS geometry description is performed. Different aspects of geometry description and physics simulation for the new detectors will be discussed. Progress on R&D efforts for the Phase-2 simulation will be presented, which includes reports on experience of application of G4HepEm external library.

Presented at *CHEP2024 27th International Conference on Computing in High Energy and Nuclear Physics*

R&D Adoption and Progress in Full Simulation of the CMS experiment

Norraphat Srimanobhas[1,*], *Sunanda* Banerjee[2], *Fabio* Cossutti[3], *Vladimir* Ivantchenko[4,5], *Natascha* Krammer[6], *Malik Shahzad* Muzaffar[4], *Kevin* Pedro[7], and *Lorenzo* Pezzotti[8] for the CMS Collaboration

[1]High Energy Particle Physics Research Unit, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
[2]University of Wisconsin -Madison, Madison, WI, USA
[3]Universita e INFN Trieste, Italy
[4]CERN, Geneva, Switzerland
[5]Joint Institute for Nuclear Research (JINR)
[6]Institut für Hochenergiephysik, Vienna, Austria
[7]Fermi National Accelerator Laboratory, Batavia, IL, USA
[8]Universita e INFN Bologna, Italy

Abstract. In this work we report on evolution of usage of Geant4 within CMSSW and adaptation of the newest Geant4 11.2.1, which is expected to be used for CMS simulation production in 2025. Physics validation results and results on CPU performance are reported. For the Phase-2 simulation several R&D are carried out. A significant update for CMS geometry description is performed. Different aspects of geometry description and physics simulation for the new detectors will be discussed. Progress on R&D efforts for the Phase-2 simulation will be presented, which includes reports on experience of application of G4HepEm external library.

1 Introduction

The Compact Muon Solenoid (CMS) experiment is one of the major experiments operating at the Large Hadron Collider (LHC) at CERN [1],[2]. CMSSW is the CMS data processing software used by CMS Collaboration. It is built around four key components:

1. Framework – Manages the execution of different software modules in a structured and efficient manner.
2. Event Data Model (EDM) – Defines how event data is stored, accessed, and processed.
3. Condition system - Non-event data, valid over a specific time period called the interval of validity (IOV), which spans multiple events. Examples include calibrations, alignments, geometry, magnetic field settings, or run conditions.
4. Services – Provide essential utilities for simulation, calibration, alignment, and reconstruction of event data. These utilities must not impact physics output.

---

[*] e-mail: Norraphat.S@chula.ac.th

In the context of full simulation, cmssw incorporates various modeling aspects, including (i) the interaction region, (ii) particle propagation through the hierarchical structure (a.k.a. geometry) of the CMS detector, along with relevant physical interactions, (iii) multiple interactions occurring per beam crossing and event overlays (commonly known as pile-up simulation), and (iv) the detector's electronics response (referred to as digitization). These simulation components are managed using the CMS detector simulation software, which is built upon Geant4 [3],[4],[5],[6]. As with any sophisticated software system, CMS simulation tools, including Geant4, undergo continuous refinement to enhance accuracy, computational efficiency, and overall compatibility [7]. CMS maintains a rigorous approach to quality assessment, ensuring that every update is thoroughly validated. These developments extend beyond modifications to the core software and also encompass updates to Geant4 versions, underlying components, and essential aspects such as geometry descriptions, all of which are critical for achieving precise simulations.

This proceeding discusses the progress in Full Simulation within cmssw. We begin with the evolution of Geant4 and cmssw. Next, we outline the steps required before migrating Geant4 under cmssw, including software validation. Additionally, this proceeding presents new validation studies, including the 2018 HGCal test beam and the 2006 test beam using G4HepEM. Finally, we compare the performance of Full Simulation across various cmssw versions.

2 Evolution of Geant4 versions with cmssw

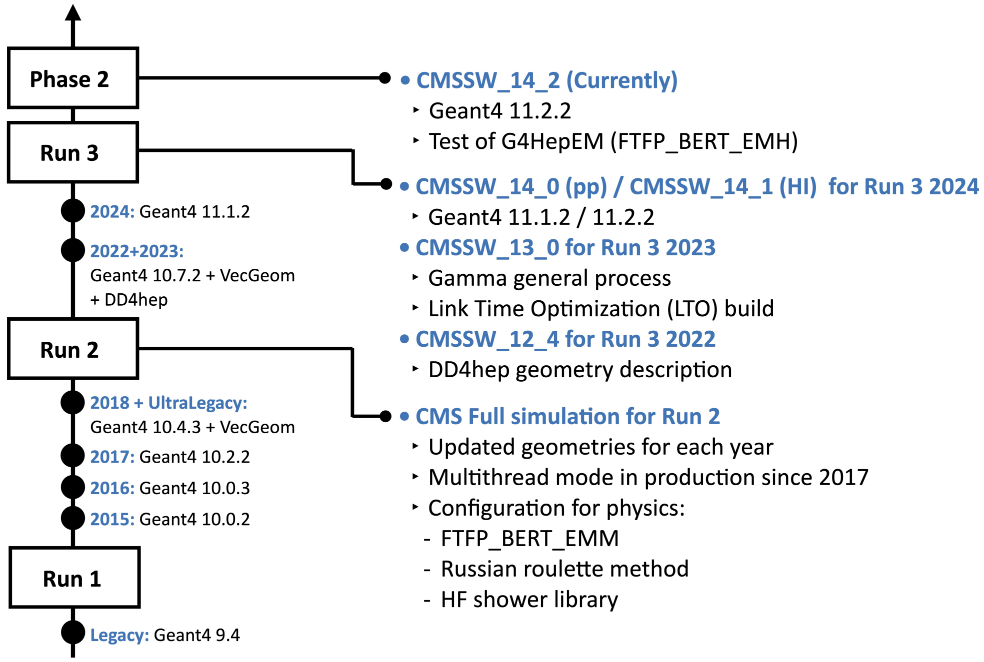

Figure 1. Evolution of CMS software, Geant4, computing environment from LHC Run 1, Run 2, and current Run 3 [9], and also Phase-2 development.

CMS Full Simulation migrated to Geant4 approximately fifteen years ago. Initially, cmssw used the QGSP_FTFP_BERT_EML physics list for the first comparisons with collision data, running on Geant4 version 9.4.2, before transitioning to 9.6.2. During the 13 TeV center-of-mass collisions in 2015, cmssw switched to a multi-threaded framework [8] and adopted Geant4 version 10.0.2, continuing with the QGSP_FTFP_BERT_EML physics list.

In the 2017 production plan, CMS migrated to a new physics list, FTFP_BERT_EMM, alongside Geant4 version 10.2.2. Over recent years, the FTFP physics model has undergone substantial improvements to enhance the accuracy of physics predictions. As a result, FTFP_BERT has become the recommended physics list from the Geant4 collaboration. Regarding EML vs. EMM, the key difference lies in the treatment of multiple scattering. EML applies a simplified multiple scattering model across all detectors, while EMM employs a more detailed multiple scattering model for the hadron sampling calorimeter (HCAL) while using the simplified model for other detectors. This distinction allows cmssw to better describe the response of HCAL, improving the overall accuracy of the simulation. In 2024, with both Run 3 and Phase-2 simulations, CMS uses Geant4 version 11.2.2 with cmssw Series 14. The FTFP_BERT_EMM physics list is still be used by default. Fig.1 shows an evolution of cmssw, Geant4 and its components.

3 Migration of Geant4 under cmssw

3.1 2006 test beam with CMS calorimeter prototypes

Regarding 2006 test beam [10], CMS conducted test beam studies using the H2 beamline at CERN's Super Proton Synchrotron (SPS). As depicted in Fig. 2, the setup included a barrel electromagnetic calorimeter (EB) placed in front of two production wedges from the hadron barrel calorimeter (HB), with the outer hadron calorimeter (HO) mounted on an adjustable platform. This arrangement was designed to replicate the geometric configuration of the CMS detector. The platform's mobility in both the ϕ and η directions allowed precise targeting of calorimeter towers, simulating the trajectory of particles originating from the CMS interaction point. To trigger data collection, four scintillation counters were positioned approximately three meters upstream of the calorimeters, operating in a subset coincidence mode. The experiment utilized monochromatic secondary and tertiary beams with momenta ranging from 2 to 350 GeV/c, supplemented by additional beam counters to ensure the selection of pure beam interactions.

In [11], there is a report comparing test beam data with Monte Carlo simulations using FTFP_BERT_EMM. The comparison is based on the mean energy response, defined as the ratio of the calorimeter's total energy to the beam momentum. This ratio is evaluated as a function of beam momentum across different beam types to identify discrepancies between data and simulation.

This report presents a comparison between test beam data and Monte Carlo simulations using Geant4 11.2.2 with the FTFP_BERT_EMM and FTFP_BERT_EMH physics models. The FTFP_BERT_EMH model serves as an alternative electromagnetic physics configuration in Geant4, implemented through the G4HepEm standalone library. In cmssw, G4HepEm is utilized as an external package, replacing the standard electromagnetgic physics for gamma and electron/positron interactions [12]. For this study, the G4HepEm sub-library is tested exclusively in a CPU-based environment. Figs 3 and 4 show the mean response for pion and proton, respectively [13]. A good agreement between the EMM and EMH physics lists has been observed. Interest in EMH has emerged due to its potential compatibility with GPUs, offering an expected speedup. Additionally, with the introduction of the new CMS HGCal, which functions as a large hadron sampling calorimeter, the simulation step may require more runtime due to the detailed multiple scattering model used in the EMM physics list.

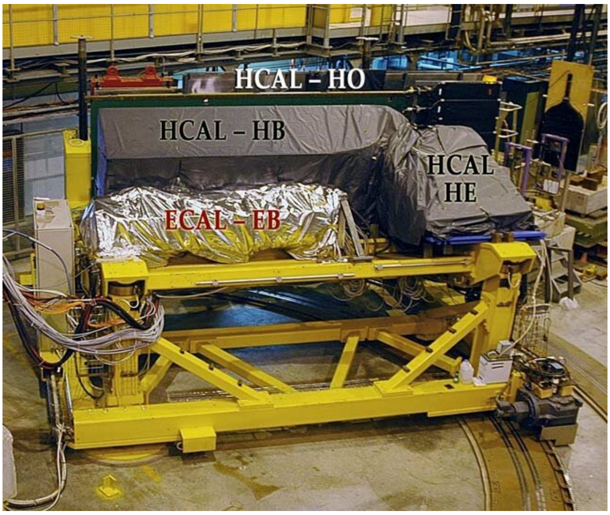

Figure 2. An overview of the H2 test beam area showcasing the HB, HE, HO, and ECAL EB components, all meticulously mounted on a versatile moving platform [10].

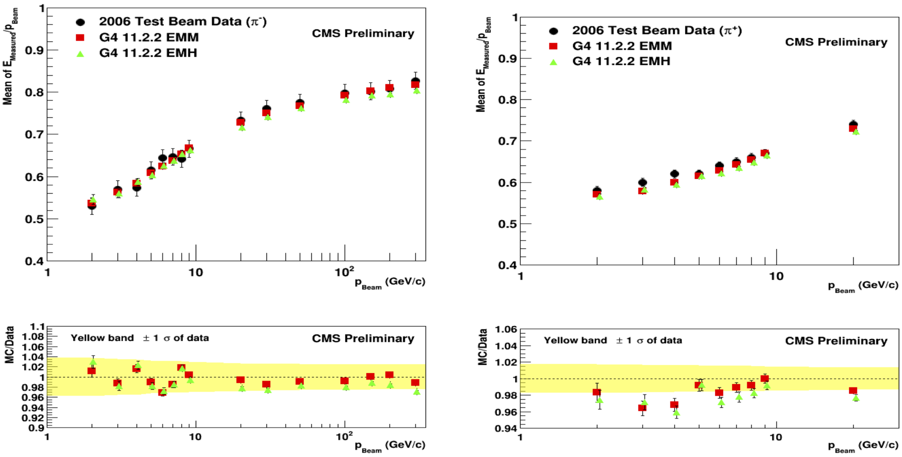

Figure 3. (Top) The mean response for negative (left) or positive (right) pions as a function of momentum compared to MC predictions; (bottom) Ratio of MC to data for negative (left) or positive (right) pions as a function of momentum. The yellow band shows one standard deviation of the data [13].

3.2 2018 HGCAL test beam

The CMS HGCAL is a sampling calorimeter consisting of an electromagnetic section (CE-E) followed by a hadronic section (CE-H), both of which are longitudinally segmented into 50 layers of the silicon-based modules.

The HGCAL prototype tested in the 2018 beam experiment [14] consists of two primary sections the CE-E prototype and the CE-H prototype followed by an Analogue Hadron Calorimeter (AHCAL) segment. The entire detector setup, including the scintillators, is positioned on a concrete platform within CERN's H2 beamline area.

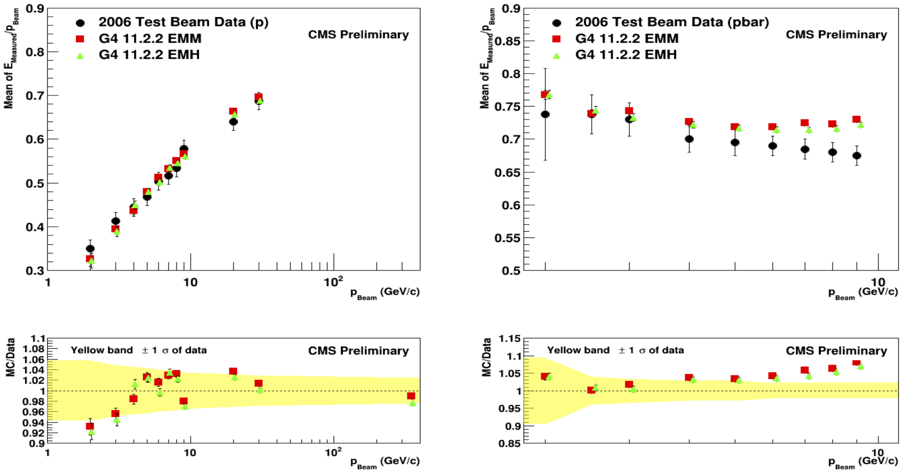

Figure 4. (Top) The mean response for protons (left) or anti-protons (right) as a function of momentum compared to MC predictions; (bottom) Ratio of MC to data for protons (left) or anti-protons (right) as a function of momentum. The yellow band shows one standard deviation of the data [13].

In terms of the beamline, protons are accelerated to a momentum of 400 GeV/c by the CERN Super Proton Synchrotron (SPS) before being directed onto a 500 mm thick beryllium target. This interaction produces secondary particles, such as muons, electrons, and pions, which are subsequently selected and transported to the HGCAL prototype, situated approximately 600 meters downstream. The beamline facilitating this transport is equipped with dipole and quadrupole magnets, as well as collimators, to ensure precise beam shaping and delivery. The schematic view of the experimental setup is shown in Fig. 5.

The energy response of the calorimeter is calibrated in units of MIP. The calibration and weighting of the response performed independently in CE-E and in CE-H sections and the total energy is expressed via formula

$$E(\mathrm{GeV}) = \alpha E^{CE-E} + \beta(E^{CE-H} + \delta E^{AHCAL}), \tag{1}$$

where α = 10.5 MeV/MIP obtained from the 50 GeV e^+ beam, β = 80 MeV/MIP obtained from the 50 GeV π^- beam, and $\delta = 0.4$.

The test beam analysis is integrated into the Geant-Val standalone software [15]. This standalone implementation enables the Geant4 team to conduct early comparisons between Geant4 simulations and experimental data while facilitating R&D efforts for software development using GPU acceleration.

The data for π^- beam in interval of the beam energy 20-300 GeV allowing to provide a clean analysis of the energy deposition in each sensitive layer of the calorimeter. The energy response and resolution as a function of beam energy of each detector has been studied in [14]. The total normalized resolution as a function of the beam energy is shown in Fig. 6. In this comparison the simulation is performed for the default Geant4 physics list FTFP_BERT for versions 10.4.3 (used in [14]) and 11.2.2 (the current CMS default). It is confirmed that in the new version of Geant4 the resolution of hadronic showers in sampling calorimeters is underestimated due to the modification in the FTFP string model.

In contrary to hadronic showers in sampling calorimeters, new Geant4 versions provide better agreement between data and simulation (Fig. 7). In this plot the mean energy depositions in units of MIP per layer of CE-E are shown for Geant4 11.2.2. Two physics lists are compared FTFP_BERT and FTFP_BERT_EMZ, the level of agreement between these two is about 2% in the peak and better than 5% at the tail of the shower. In the original publication [14] it was shown that only physics list FTFP_BERT_EMZ provides an agreement to the data. This physics configuration requires 50% more CPU. With the current default CMS physics configuration FTFP_BERT_EMM the accuracy of electromagnetic shower simulation is acceptable. Note, that FTFP_BERT_EMM and FTFP_BERT are the same for the HGCAL calorimeter.

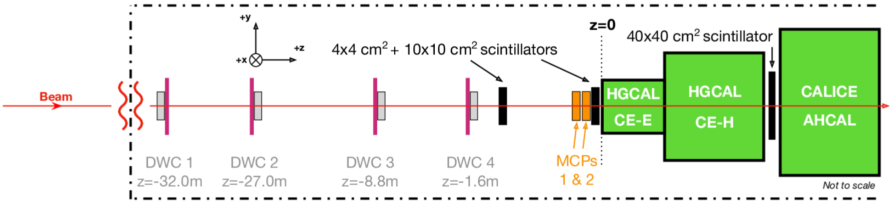

Figure 5. Schematic view of the experimental setup of the test beam [14].

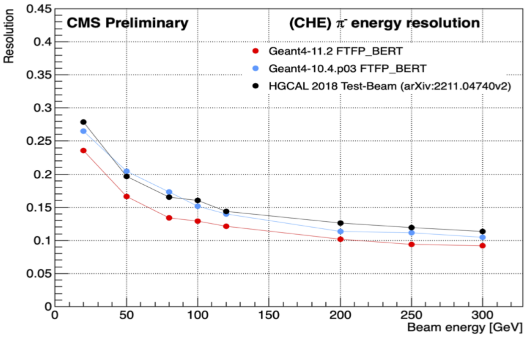

Figure 6. Pion energy resolution as the function of beam energy for Geant4 10.4 (blue) and 11.2 (red) using Geant4 default FTFP_BERT Physics List, data – black points, statistical errors are on level of marker size and are not shown [13].

4 Software performance

CPU monitoring is conducted in Full Simulation, utilizing various Phase-2 geometry design updates in Geant4-based Full Simulation. The performance trends are analyzed over a period of nearly five years, spanning cmssw versions 11_0_X to 14_1_X. Fig. 8 shows the historical trends of Full Simulation CPU time performance. During the period, the CPU time has improved for the ttbar process by 35%. Main improvements are connected with the Geant4 migration from 10.4 to 10.7 (cmssw 11_3_X), to 11.1.1 (cmssw 13_1_X) and to 11.1.2 (cmssw 13_3_X), updates of the HGCAL and Muon geometry (cmssw 12_3_X), the change of the computing platform operating system from CentOS 7 (SLC7) to AlmaLinux 8 (EL8) (cmssw 12_4_X) and the usage of LTO (Link time optimization) build method (cmssw 13_0_X). Some slowdowns relate to addition of more detailed geometries.

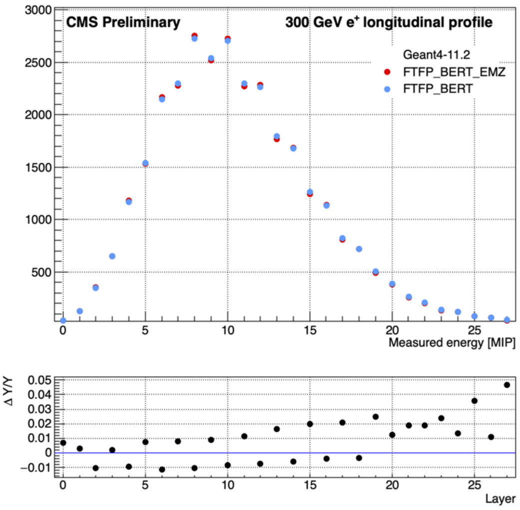

Figure 7. (Top) positron energy deposition in units of MIP as a function of number of layer in EM section of HGCal. The default Geant4 Physics List FTFP_BERT (blue) is compared with FTFP_BERT_EMZ (red), statistical errors are on level of marker size and are not shown. Bottom: relative difference of energy deposition for these 2 variants of simulation is normalized on FTFP_BERT values. [13].

5 Summary

The Full Simulation software of the CMS experiment provides as ongoing Run 3 simulation production and preparation of the Phase-2 high luminosity experiments. The CPU performance between Run 2 and Run 3 is significantly improved (35% for ttbar events), the accuracy of prediction s is under control. This was achieved also due to ability of migration of cmssw to newest version Geant4, the current production version is 11.2.2.

The development of the Phase-2 simulation is performed in parallel in the master branch of cmssw using geometry configuration of Phase-2 detectors and the same physics configuration as for Run 3. It is shown that with the recent Geant4 there is no need to use the configuration of the EM physics EMZ, which requires significantly higher CPU resources. For the first time, we presented results of testing of the new sub-library G4HepEm within cmssw for simulation of gamma, electrons, and positrons, which is designed for more effective simulation of Phase-2 experiments and which is compatible not only for usage with CPU but also for GPU accelerators.

6 Acknowledgement

The presentation and proceeding receives funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B39G670016].

References

[1] CMS Collaboration, JINST **3** S08004 (2008)

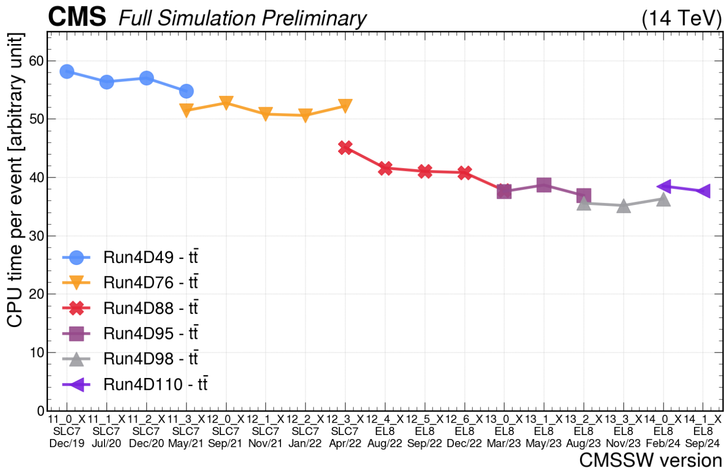

Figure 8. Historical trends of Full Simulation CPU time performance of 14 TeV ttbar process for different Phase-2 geometry design updates (Run4Dxx) [16]. The average CPU run time per event in relative units of the event simulation is shown for 500 events on single threaded jobs.

[2] CMS Collaboration, JINST **19** P05064 (2024)
[3] S. Agostinelli *et al.*, Nucl. Instrum. Meth. A **506** 250-303 (2003)
[4] J. Allison *et al.*, IEEE Trans. Nucl. Sci. **53** 270-278 (2006)
[5] J. Allison *et al.*, Nucl. Instrum. Meth. A **835** 186-225 (2016)
[6] S. Banerjee, J. Phys.: Conf. Ser. **396**, 022003 (2012)
[7] D. J. Lange *et al.*, J. Phys.: Conf. Ser. **608**, 012056 (2015)
[8] M. Hildreth *et al.*, J. Phys.: Conf. Ser. **664**, 072022 (2015)
[9] V. Ivanchenko *et al.*, EPJ Web of Conf. **251**, 03016 (2021)
[10] CMS HCAL/ECAL Collaborations, Eur. Phys. J. C **60**: 359–373 (2009)
[11] N. Srimanobhas *et al.*, EPJ Web of Conf. **295**, 03017 (2024)
[12] G4HepEm code repository https://github.com/mnovak42/g4hepem
[13] CMS Collaboration, CMS DP-2024/087 (2024)
[14] B. Acar *et al.*, JINST **18** P08014 (2023)
[15] https://geant-val-web.cern.ch/
[16] https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults