Impact of detector simulation in particle physics collider experiments – highlights

V. Daniel Elvira[1],*

[1]Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL, 60510-5011, USA

Abstract. Detector simulation has become fundamental to the success of modern high-energy physics (HEP) experiments. For example, the Geant4-based simulation applications developed by the ATLAS and CMS experiments played a major role for them to produce physics measurements of unprecedented quality and precision with faster turnaround, from data taking to journal submission, than any previous hadron collider experiment. The material presented here contains highlights of a recent review on the impact of detector simulation in particle physics collider experiments published in Ref. [1]. It includes examples of applications to detector design and optimization, software development and testing of computing infrastructure, and modeling of physics objects and their kinematics. The cost and economic impact of simulation in the CMS experiment is also presented. A discussion on future detector simulation needs, challenges and potential solutions to address them is included at the end.

1 Introduction

This paper presents highlights of a recent review article [1] which focuses on the impact of detector simulation in particle physics collider experiments through concrete examples. (The review paper includes a large number of figures and a comprehensive list of references which had to be removed or cut short in the highlights in order to comply with length limitations.) The thesis of this work is that detector simulation is of critical importance to the success of the international high-energy physics (HEP) experimental program and a determinant factor, albeit not the only one, for faster delivery of outstanding physics results. A general introduction is followed by a few words about different types of software applications for detector modeling, including those based on the most widely used toolkit, Geant4 [2]. Examples are presented to show the many uses of simulation in different areas and stages of collider experiments, and on how accurately particle and event properties, and their kinematics, are modeled. For this, it borrows heavily but not exclusively from the CMS (LHC) and D0 (Tevatron) experiments, drawing from the author's personal experience.The subjects of the relationship between simulation and publication turnaround, as well as the economic impact and cost of simulation, are also tackled. The paper ends with an outlook on the future evolution of detector simulation in HEP.

---

*e-mail: daniel@fnal.gov

2 Some history and motivation for detector simulation

Through the years, simulation has become essential to design, build, and commission HEP detectors, as well as to analyze and interpret their data. In the old times, simulation mostly consisted of back-of-the-envelope estimates based on simple analytic calculations. The era of detailed detector simulation started in the late seventies and early eighties with the development of tools such as the "Electron Gamma Shower" code (EGS) [3], followed by the "GEometry ANd Tracking" package (GEANT). Eventually, GEANT3 [4] evolved into a toolkit capable of describing complex geometry and the propagation of particles through material and fields, until it became the most popular and widely used detector simulation software in HEP. There are other tools, such as FLUKA [5] and MARS [6], which will not be discussed here. Geant4-based simulation is the norm these days in collider experiments, as computing performance improved to a level that allows detailed modeling of detector geometry and particle interactions within the available resources.

Detector simulation saves experimenters time and money, improving the quality and accuracy of physics measurements. Simulation enables scientists to design the optimal detector for the best physics at a given cost. Most importantly, it teaches physicists what mark a new particle would leave in the detector if it existed. For example, the Standard Model (SM) predicted a Higgs Boson which would decay into two photons with given properties for the event and the individual particles. In the discovery analysis, the ATLAS and CMS experiments observed two photon events with the predicted detector marks which they were able to identify efficiently with the aid of detector simulation.

A typical HEP detector consists of a tracker to measure charged particle tracks, calorimeters to measure particle energy, and muon systems to measure the muon trajectories. A high quality simulation application models particle tracks and calorimeter showers accurately. For example, deficiencies in the physics models that control particle interactions in the tracker and shower development in the calorimeter systems affect the trajectory of leptons, photons and hadrons, their momentum and energy response and resolution, and the amount and distribution of un-clustered energy. Mis-modeling of the shower leakage beyond the calorimeter boundaries affects energy measurements of hadrons and the reconstruction of muon tracks. An inexact description of the shape or material composition of detector elements also has an effect on particle trajectories and shower development. All of the above translates into inaccuracy in the simulation of particle reconstruction, identification, isolation, and separation efficiencies, which impacts high-level physics predictions including event kinematics and particle characterization in precision measurements, as well as signal and background estimates in searches.

A few facts and numbers help understand the size and importance of the simulation task in a typical collider experiment. For example, the ATLAS and CMS Collaborations have produced tens of billions of Geant4-based events since 2010 at a speed and with physics accuracy never seen before. The Tevatron experiments in the early nineties produced only hundreds of thousands of GEANT3-based events, of relatively poor quality in comparison, because the time necessary to simulate an event in detail at the Tevatron was approximately one hour while it is only about a minute at the LHC. Consequently, Geant4-based simulation has contributed significantly to shorten the time between data taking and submission of physics results for publication in scientific journals. Other dominant factors are the improvements to detector and computing technology, a wealth of experience from pre-LHC experiments, better calibration and analysis techniques, communication tools allowing physicists to collaborate effectively while based at different locations across the planet. There is, however, a price to pay for the benefits of simulation. During the 2010-2016 data-taking period, simulation took more than half of all computing resources available to the LHC program and, in

the next decade, the experiments will need two orders of magnitude more simulated events of increased physics accuracy to make full use of the capabilities of the upgraded accelerator and detectors, while budgets allocated to computing resources are expected to remain flat.

3 Software and toolkits

The simulation software chain or "simulation application" in a HEP experiment consists of four steps: physics generation, detector simulation, calibration and reconstruction. This paper concentrates on detector simulation, which includes both the passage of particles through material and fields, and the modeling of the electronics and contributions of extra collisions from the same or a different bunch crossing (pileup). Throughout the rest of the paper, simulated events, samples, and simulation predictions may be referred to as Monte Carlo (MC) events, samples, and predictions.

3.1 Types of simulation

There are several types of simulation that serve different purposes depending on the accuracy and speed required by a given task. *Toy simulation* is based on simple analytical equations and is used for simple detector or physics studies where only qualitative analysis matters. It is designed to be extremely fast, with speeds of the order of a very small fraction of a second per event. *Parametrized simulation* runs at a speed of the order of one second per event by cutting corners on the geometry and field descriptions, utilizing parametrizations to describe energy response and shower shapes, and smearing high-level quantities to model resolution effects. Current experiments still use parametrized simulation in computing intensive campaigns such as signal production in beyond the standard model (BSM) searches, or production of MC samples for multiple scenarios in detector design studies. Examples of parametrized simulation are the QFL package developed by the CDF experiment in the nineties, and the ATLAS (Atlfast-II) and CMS fast simulation frameworks. All were tuned with the aid of test beam data, single tracks from collider data, and full simulation. *Full simulation* in modern experiments is based on the GEANT toolkit to model geometry, fields, and shower shapes in detail. It is used in studies where geometry and physics accuracy are important, for example in the case of corrections and background estimations for physics measurements. The speed of full simulation is of the order of seconds to minutes per event, and the output data format is identical to that of real data, as it is for parametrized simulation in most of the cases. Fast simulation typically refers to versions of parametrized or toy simulation tools, but the concept is misleading because fast is a relative term. Nowadays, experiments are moving towards simulation frameworks that offer the flexibility to incorporate fast simulation techniques to a base Geant4 application. The result is a hybrid which may include tables, shower libraries, parametrizations, and image recognition in the context of machine learning, one of the most popular techniques under consideration these days.

3.2 Physics validation of full simulation

The impressive success of Geant4-based full simulation applications at the LHC is the result of many years of hard work and partnership between the experiments and the Geant4 team. An element at the core of a Geant4 application is the "physics list", a subset of physics models assembled in a coherent way to describe the interaction of particles of different types with the detector material in an energy range between 250 eV and 100 TeV. The output of the software application is a collection of particle trajectories and "simulated hits" with a position, time,

and energy deposited in the associated detector volume. The physics validation of Geant4 is a collaborative task between Geant4 developers and the experiments. Individual Geant4 models at the single-interaction level are validated against thin-target experiments consisting of beams of particles of different types (e, π, p, etc.) directed onto thin targets made of materials typically used in HEP detectors (Be, C, Si, Cu, Fe, Pb, etc.). HEP experimental data is used to validate higher-level quantities such as energy response distributions and shower shapes. In other words, thin-target data is used for a first principles Geant4 model tuning, while HEP data is used for verification or small adjustments to the models.

It used to take about one hour for the D0 experiment to generate a full simulation event with a GEANT3-based application in the nineties. In addition, the D0 detector lacked a magnetic field surrounding the tracker, test beam programs were limited in scope, and communication technology for international collaboration was deficient. For example, the D0 measurement of the e/π response ratio versus beam energy had large systematic uncertainties and was limited to the 10-150 GeV energy range [7]. This, combined with the fact that the statistical uncertainties in the associated simulation were large due to a computationally costly full simulation application, made it very difficult for the experiment to come up with an accurate, validated, and fast GEANT3-based application for use in data analysis. The compromise solution was to introduce approximations in exchange for time performance: a simplified geometry and average materials, and a shower truncation at 95% of the total energy deposited. Instead, CDF used the QFL parametrized simulation tuned to minimum bias and test beam data, and it was not until 2003 that the experiment was able to develop a GEANT3-based application combined with GFLASH [8] shower parametrizations. In contrast, CMS measured in test beams the single particle hadronic response in the 2-300 GeV range with a ~ 2% absolute uncertainty, and experimenters were able to either select or optimize a Geant4 physics list that provided excellent agreement with experimental data within the small systematic and negligible statistical uncertainties [9]. Similar results were obtained by the ATLAS experiment [10]. Test beams also allowed ATLAS and CMS to verify that the high and low tails of hadronic response functions are modeled with high accuracy, a matter of great importance in the simulation of hadron energy resolution and missing transverse energy. Another quantity for which MC-to-data agreement is within 5% in ATLAS and CMS is the ratio between the energy measured for single charged tracks in the calorimeter and their momentum measured in the tracker [9] [11]. More importantly, the MC-to-data ratio is constant as a function of the track momentum for $p_T > 1$ GeV, an important feature because a normalization factor may be easily applied at the calibration stage.

4 Applications to HEP collider experiments

4.1 Data analysis and interpretation

Data-driven methods are techniques applied to real collider data to measure physics backgrounds, calibration factors, resolutions, efficiencies, etc. They use detector properties, conservation laws, and math tools and they are applied at the detector-level both to real data and simulated data. Simulation is a fundamental element in the development of data-driven methods because corrections in data analysis come mostly from MC truth and are adjusted with small scale factors calculated as the detector-level ratio of data-driven measurements performed on collider data and MC. The trick is that systematic uncertainties mostly cancel in this ratio because the same method is applied to the numerator and the denominator. One example is the di-object p_T balance method used to measure the jet energy response, or jet energy scale, which applied to CMS detector-level data and MC gives a scale factor of 0.98 [12]. The accuracy of a data-driven method improves as the scale factor tends to one within a small

uncertainty, something achievable only with the aid of good quality simulation. A second example is the definition of control regions and the design of methods for background estimation in physics measurements, such as for example the QCD background in the many jets plus $E_{\mathrm{T}}^{\mathrm{miss}}$ SUSY search, estimated with the factorization data-driven method [13]. A third example is the data-driven tag-and-probe method for the measurement of particle reconstruction and identification efficiencies and fake rates. The method uses the a priori knowledge of an identified reconstructed object, called the tagged object, and then measures the fraction of times an algorithm identifies and reconstructs a probe object.

Closure tests are developed to verify that data-driven measurements are correct within the quoted uncertainties, by comparing detector-level MC with the truth information. In other words, for a measured quantity X, a test $T = (X_{\mathrm{MC}}^{\mathrm{reco-level}} - X_{\mathrm{MC}}^{\mathrm{truth}})/X_{\mathrm{MC}}^{\mathrm{truth}}$ must be consistent with zero within the uncertainty of the data-driven method for the method to close. Lack of closure would indicate the need to go back to the drawing board and understand the biases in the procedure, a process that requires that the physics quantities involved in the measurement are simulated accurately. The limitations of the D0 full simulation in the early nineties resulted in the lack of large and accurate MC samples to demonstrate the data-driven methods by closure. This caused a delay in a number of physics measurements. For example, most QCD results involving jets were delayed from the start of the first run in 1992 to 2000, until the jet energy scale uncertainty was reduced to less than 3% and the measurements became competitive.

Signal samples for both SM precision measurements and BSM searches are necessary to design physics analyses for maximum sensitivity and accuracy, as well as for the interpretation of results in the the context of proposed theoretical models.

4.2 Detector design and optimization

HEP detectors typically consist of tracker, calorimeter, muon, and magnet systems. In the design stage, experiments study the physics performance and cost of different technologies, play with the shape, size and materials of sub-detector systems, consider different levels of segmentation and granularity, as well as magnetic field configurations and strengths. For example, ATLAS decided to have a relatively weak solenoidal field surrounding the tracker and quadrupoles in the muon system, plus excellent calorimetry and tracking. CMS uses a stronger magnetic field surrounding both the tracker and the calorimeters, therefore making the design more compact. Experiments conduct MC campaigns which consist of millions of events generated for different detector scenarios in order to make the case for a given design, optimize parameters for best physics, or study the impact of different de-scoping options. These studies are an absolute requirement for every HEP experiment seeking approval from funding agencies. Examples are the studies by ATLAS and CMS to make the case for extending and improving their trackers with the objective to increase tracking efficiency in the forward rapidity regions and prevent efficiency degradation in a HL-LHC environment with 140-200 pileup events. The impact of accelerator conditions and detector design options is also investigated in data analysis performed on simulated datasets, such as for example the mass and cross section reach for particle production in the context of BSM searches [14] [15].

4.3 Software and computing design and testing

The LHC experiments use the Worldwide LHC Computing Grid [16] which is divided into four tiers providing different services. Simulation is utilized to develop the workflow and data flow for this complex system. In CMS, the combined procedure was tested in Computing, Software and Analysis (CSA) challenges at 25%, 50% and 75% capacity [17]. Hundreds of millions of MC events were produced, reconstructed and processed into realistic formats

for analysis. The tests also included data transfers and monitoring of event file size, as well as memory and CPU consumption. The use of simulation to stress test the software and computing infrastructure in ATLAS and CMS is a major reason why there were no significant issues related to computing operations at the start of the first run.

5 Modeling of particle and event properties and kinematics

The level of agreement between the MC predictions of high-level physics observables and the corresponding data measurements is the ultimate accuracy test of the simulation software. Examples are event properties and kinematic distributions for final states including photons, electrons, muons, taus, jets from light and heavy quarks, and $E_{\mathrm{T}}^{\mathrm{miss}}$. These physics objects constitute the basic ingredients of every HEP physics result.

The modeling of the b-jet reconstruction and identification is a critical simulation benchmark because b-jets are present in many SM precision measurements and BSM searches. B-jet identification, also called b-tagging, depends on the impact parameter of charged tracks and reconstructed decay vertices in the jet, as well as on the presence of leptons. The impact parameter (IP) is the point of closest approach between a track and the primary vertex and it is positive for b-quarks and slightly positive for light quarks, close to zero, although resolution effects yield positive and negative values in a real detector. Data driven methods are used to measure the b-tagging efficiencies and fake rates, then accurate simulation of IP variables necessitates good modeling of detector materials, energy loss, ionization, multiple scattering, noise, and pileup, mainly in the tracker systems. The impact parameter distributions for light and heavy jets in a di-jet sample are modeled accurately by the ATLAS and CMS experiments, except in the tails where resolution effects are difficult to simulate. Data-to-MC agreement for b-tagging efficiency is achieved by both experiments within less than 5% [18] [19]. Mis-tag rates are more difficult to simulate because the tracks used in the measurement populate the tails of the light quark IP distributions. Therefore, the MC-to-data disagreement could grow to a large factor depending on the operating point of the selected algorithm. Gauge bosons are also at the core of SM and most BMS searches. MC p_{T} distributions agree within 10% with CMS and ATLAS data, except in the p_{T} range above 200 GeV for Z+jets where the CMS simulation overestimates the data by a difference that grows linearly with p_{T}, most probably coming from generator mis-modeling of event topology and kinematics [20] [21]. The modeling of $E_{\mathrm{T}}^{\mathrm{miss}}$ is a very challenging task because this quantity depends on all physics objects in the event, as well as on un-clustered energy. Simulating $E_{\mathrm{T}}^{\mathrm{miss}}$ is of paramount importance because a large amount is present in some BSM final states. Events with low to medium values of $E_{\mathrm{T}}^{\mathrm{miss}}$ coming from neutrinos are better modeled than the high $E_{\mathrm{T}}^{\mathrm{miss}}$ tails in multi-jet samples with origin in resolution or detector malfunction. This is the reason why MC is not used to estimate QCD backgrounds in searches with high $E_{\mathrm{T}}^{\mathrm{miss}}$ in the signal region. $E_{\mathrm{T}}^{\mathrm{miss}}$ is modeled within 10-20% by ATLAS and CMS in W/Z+jets samples although, in both experiments, the systematic uncertainty grows above 50% in the range where hadronic shower mis-measurement dominates [21] [22]. It is significantly more challenging to simulate $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution because it is affected by fluctuations in the energy measurement of all particles involved in the calculation, including un-clustered energy. That makes the ATLAS and CMS MC-to-data 5-10% agreement in Z+jets and γ+jets a very impressive achievement [21] [22].

6 Publication turnaround

The Tevatron program coincided with the dawn of the era of detector simulation toolkits right at the transition time between sporadic and systematic use of the GEANT3-based full

simulation application. Simulation is not the only factor but it is a major one explaining why it took longer at the Tevatron than at the LHC to understand the detector and the data well enough to submit publications. CDF had a run 0 starting in 1988 and published 20 papers within the first three years [23] [1], while D0's first run started in 1992 and the experiment published seven papers within three years [24] [1]. In relative terms, CDF published faster than D0 in their first run, one of the reasons being the presence of a solenoidal field which was used for calibration and for the tuning of a parametrized simulation. The LHC experiments published much faster. For example, the CMS Collaboration has submitted 796 papers as of September 11th 2018, 19 papers in 2010 during the first run, 64 by June 2011, six months after the end of the run 1, and 90 by the end of 2011, a year after the end of run 1 [25] [1]. These papers are not the low hanging fruit but report on complex measurements which need an excellent understanding of the detector and data. Among the other important factors that contributed to this much shorter publication turnaround are the thousands versus the hundreds of members in the LHC and Tevatron experiments respectively, and the advancements in detector and computing technology. Simulation had a direct impact through the effect on software readiness and the development of calibration, correction, and analysis methods.

7 Economic impact and cost

If the simulation chain is defined to include generation, interaction with matter, readout modeling, reconstruction, and analysis of MC data, the whole process consumed 85% of the CPU resources in the 2010-2016 data taking period, while 40% of the total was spent in the Geant4 module. The rest went to reconstruction and analysis of real collider data. Analyzing two time periods, one the 2012 calendar year, and the other one between May of 2015 and May of 2016, the conclusion is that 45 to 70 thousand CPU cores were used at full capacity by CMS at a purchasing cost of 5-8 million US dollars. The cost of this computing power when purchase from industry, including life-cycle, operation and maintenance is in the 6-10 million US dollars range. (This information was obtained from the CMS Dashboard and from private communication with Oliver Gutsche.) It is interesting to point out that an improvement of 35% in Geant4 time performance would allow CMS to make savings of up to 3 million US dollars per year. Given that the computing needs of the HL-LHC program are 10-100 times higher than those during 2010-2016, depending on the simulation and reconstruction solutions to be implemented, the challenge is daunting. The cost of the design, development, validation, and support of simulation toolkits, such as Geant4, as well as the development of the experiment applications, must be added to the previous estimates. In 22 years of existence, the in Geant4 has been approximately 500 person-years, equivalent to more than 100 million US dollars. How much more would the design and commissioning of HEP detectors and the delivery of the associated physics programs have cost without the Geant4 toolkit? Would it have been at all possible to deliver results of the same quality? How much physics would have been lost? Two corollaries that follow this analysis are that the cost of physics software is a significant fraction of that of detectors, and that the cost of simulation and reconstruction should be considered in the cost-benefit equation during the detector design stage.

8 The future

The Geant4 Collaboration has invested significant effort to improve the computing performance of the toolkit. The introduction of multithreading capabilities in 2013, at the event

level, improved significantly memory usage, although not CPU consumption. Through algorithm re-engineering and optimization, time performance was improved during the 2010-2015 period by about 35% while simultaneously increasing the accuracy of physics models. This performance gain is significant but not enough to satisfy the needs of the HL-LHC program. Therefore, the international HEP simulation community is investing in R&D to increase event throughput using a combination of novel programming techniques, fast simulation methods including machine learning, and efficient use of accelerators in high performance computing systems. A list of the challenges and proposed lines of research are summarized in the "A Roadmap for HEP Software and Computing R&D for the 2020's" [26] and the detector simulation community white paper [27] delivered in the context of the planning process conducted by the HEP Software Foundation (HSF) [28] during 2017-2018.

References

[1] V. Daniel Elvira, Phys. Rep. **695**, 1-54 (2017)

[2] S. Agostineli, et al., Nucl. Instrum. Methods A **506** 250–303 (2003)

[3] R. Ford, W. Nelson, Stanford Linear Accelerator Center Report, SLAC-210, 1978

[4] R. Brun, F. Bruyant, M. Maire, A. McPherson, P. Zanarini, CERN-DD-EE-84-1, 1987

[5] A. Ferrari, P. Sala, A. Fasso, J. Ranft, CERN-2005-10 (2005)

[6] The MARS Code System, http://mars.fnal.gov

[7] The D0 Collaboration, Nucl. Instrum. Methods A **324** 53-76 (1993)

[8] G. Grindhammer, et al., Nucl. Instrum. Methods A **290** 469 (1990)

[9] S. Banerjee for the CMS Collaboration, J. Phys. : Conf. Ser. **898** 042005 (2017)

[10] ATLAS Public Results, 2017,
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileTestBeamResults

[11] The ATLAS Collaboration, Eur. Phys. J. C **77** 26 (2017)

[12] The CMS Collaboration, JINST **6** 11002 (2011)

[13] T. Schum, DESY-THESIS-2012-029, 2012

[14] The ATLAS Collaboration, CERN-LHCC-2015-020, 2015

[15] The CMS Collaboration, CERN-LHCC-2015-019, 2015

[16] WLCG: Worldwide LHC Computing Grid, http://wlcg- public.web.cern.ch/tier-centres

[17] D. Futyan, et al., (CMS Collaboration), J. Phys. Conf. Ser. **219**, 032008 (2010)

[18] The ATLAS Collaboration, JINST **11** 04008 (2016)

[19] The CMS Collaboration, JINST **8** 04013 (2013)

[20] The ATLAS Collaboration, Phys. Lett. **759** 601 (2016)

[21] The CMS Collaboration, JINST **10** 2006 (2015)

[22] The ATLAS Collaboration, Eur. Phys. J. C **77** 241 (2017)

[23] CDF summary Table, 2016, https://www- cdf.fnal.gov/physics/physics.html

[24] D0 paper submissions, 2016, https://www-d0.fnal.gov/Run2Physics/WWW/results.htm

[25] CMS Publications versus Time, 2016,
http://cms-results.web.cern.ch/cms-results/public-results/publications-vs-time

[26] The HEP Software Foundation, Comput. Softw. Big Sci. (2019) 3: 7.
https://doi.org/10.1007/s41781-018-0018-8

[27] The HEP Software Foundation, arXiv:1803.04165 [physics.comp-ph]

[28] The HEP Software Foundation, https://hepsoftwarefoundation.org