

Optimization of Geant4 for the Belle II software library

Swagato Banerjee^{1,*}, Dennis H. Wright^{2,**}, Makoto Asai², and Doris Y. Kim^{3,***}

¹University of Louisville, Louisville KY 40292, USA

²SLAC, Menlo Park, CA 94025, USA

³Soongsil University, Seoul 06978, South Korea

Abstract. During the next decade, the SuperKEKB/Belle II experiment expects to collect almost two orders of magnitude more collision data than the previous generation B-Factory experiments. Study of this data requires monumental computing resources to process and to generate the required simulation events necessary for physics analysis. At the core of the Belle II simulation library is the Geant4 toolkit. To use the available computing resources more efficiently, the physics list for Geant4 has been optimized for the Belle II environment, and various other strategies were applied to improve the performance of the Geant4 toolkit in the Belle II software library. Following the inclusion of this newly optimized physics list in an updated version of Geant4 toolkit, we obtain much better CPU usage during event simulation and reduce the computing resource usage by $\sim 44\%$.

1 Introduction

The computing software library, *basf2* [1], developed for and used at the Belle II experiment [2, 3], is a framework consisting of various modules specified to conduct given tasks. It is written mostly in C++, and executed using Python scripts. The core simulation module of *basf2* is essentially an interface to the Geant4 toolkit [4–6]. There are also other *basf2* modules handling Geant4 related tasks such as Geant4e that allows the propagation of a track with its error parameters and modelling the response of other exotic particles.

The SuperKEKB collider and the Belle II experiment are planning to collect 50 attobarn-inverse (ab^{-1}) of collision data during this decade, which is 50 and 100 times more data than the previous generation B-Factory experiments, Belle and *BABAR*, respectively. The size of Belle II data set is expected to exceed 100 Petabytes (PB). The collision event data set, as well as simulation and cosmic ray data sets, should be processed by *basf2* using the available computing resources provided by the Belle II institutes. The size of these data sets present a strong motivation to optimize the CPU usage and storage resources. The Geant4 toolkit, with its long history, is reliable and accurate in describing detector responses initialized by physics events. However, this general-purpose simulation software is, by default, not optimized for any individual experiment environment. It is thus necessary to review how the Geant4 toolkit is interfaced and implemented in the *basf2* framework to maximize its performance.

*e-mail: swagato.banerjee@louisville.edu

**e-mail: dennis.herbert.wright@cern.ch

***e-mail: dorisykim@ssu.ac.kr

In the Belle II software library, the Geant4 toolkit is imported as an external library. The SuperKEKB/Belle II experiment started data taking with its full detector configuration in March 2019, which is referred to the Phase 3 operational period. From the beginning of physics runs in Phase 3, the version of Geant4 toolkit used in the external library was v10.1.2 with the default physics list parameter set as *FTFP_BERT*.

2 The Belle II physics list

A physics list is a collection of physics processes, models and cross sections that are used to describe the responses of a detector. The popular physics list, *FTFP_BERT*, is optimized for the Large Hadron Collider (LHC) experiments, and thus is not optimal for the Belle II experiment, whose accelerator runs at a much lower energy than the LHC.

The *Belle II* physics list, which is derived from *G4VModularPhysicsList*, is a Geant4 modular physics list designed to fit the requirements of simulation at the Belle II experiment. It is an alternative to the Geant4-provided reference physics lists, and provides several options to further customize the physics that it invokes. The *Belle II* physics list was developed independently of the Geant4 physics lists, which allows for a simpler coding setting and is easier to read, understand and modify. The C++ templating technique is avoided as much as possible and a more modular structure is followed.

In the Geant4 toolkit, the construction of physical processes, models and cross sections is grouped into modules called physics constructors. In the *Belle II* physics list, there are eleven such constructors as shown in the following table:

Source	Constructors
From Standard Geant4 distributions, electromagnetic physics	G4EmStandardPhysics_option1 G4DecayPhysics G4OpticalPhysics
Hadronic physics and other physics	ProtonPhysics NeutronPhysics PionPhysics KaonPhysics HyperonPhysics AntiBaryonPhysics IonPhysics GammaLeptoNuclearPhysics

Table 1. The Belle II physics list constructors

The Geant4 uses these standardized constructor names to modularize detector response algorithms. If necessary, the actual implementation in a physics list constructor was modified, then stored in the Belle II simulation package. All the eleven constructors in *Belle II* physics list as listed in the above table are different from those in the default Geant4 physics lists.

2.1 Electromagnetic physics

This physics list is implemented with the so-called “option1” electromagnetic physics. The physics models used here are older and less precise, but faster, than those used by most physics lists. While almost all of the Geant4 reference physics lists, such as *FTFP_BERT*, use the *G4EmStandardPhysics* constructor, the *G4EmStandardPhysics_option1* was chosen

for the *Belle II* list because of its speed. Using the `G4EmStandardPhysics_option 1` at the LHC, the CMS experiment observed a 30% increase in processing speed with only a small loss in precision.

Command line options control which parts of the physics lists are included at construction time. These options are specified in the Belle2 configuration scripts. An option also exists to add optical photons to the physics list. By default, this is not added in order to save CPU time.

2.2 Decay physics

The decay physics constructor implements the weak and electromagnetic decays of all long-lived hadrons and leptons. Hadronic decays, such as the decay of the ρ or δ resonances, etc., are not included here but are a part of the various hadronic models. The decays of heavy flavor particles, such as D and B mesons are also not implemented, assuming that these decays have been already handled by other external event generator Monte Carlo libraries.

2.3 Hadronic physics

The choice of hadronic physics in the *Belle II* physics list was tuned to match the energy range of the Belle II experiment. In the lower energy range, intra-nuclear cascade models such as *Bertini* are used, while the *FTFP QCD* string models are used for the higher range. To ensure a smooth transition between low and high energy models, the energy range of each model is allowed to overlap. Within the overlap region, the used model is chosen randomly, with the probability of choosing the high energy model increasing linearly from 0 at the lower end of the region to 100% at the upper end. The region of overlap differs for each particle type, and is also different from the physics lists used for the LHC experiments.

The *Belle II* physics list also offers an option to use the Geant4 high precision neutron package below 20 MeV. This option is useful when the background from beamline are studied, but it would not be used in the standard simulation due to the large amount of CPU time required.

2.4 Production thresholds for secondary particles

The *Belle II* physics list allows the Geant4 electromagnetic secondary production thresholds to be set either globally or separately for each detector region (aka cuts per region). Currently, the thresholds are set at the default 0.07cm for all the detector regions. Using a steering script, a different threshold may be set for each sub-detector of the Belle II experiment.

3 Validation of the Belle II physics list

The Belle II detector consists of several sub-detectors arranged in a cylindrical structure around the beam pipe [2, 3]. The innermost sub-detector is the vertex detector (VXD), which includes two layers of silicon pixels and four outer layers of silicon strips. The VXD is surrounded by the main tracking device (CDC), which is a large helium-based small-cell drift chamber. The particle identification detectors, the time-of-propagation (TOP) and the aerogel-based ring-imaging Cherenkov counter (ARICH) come next, surrounded by the electromagnetic calorimeter (ECL) consisting of a barrel and two endcaps made of CsI(Tl) crystals. A superconducting solenoid, situated outside of the calorimeter, provides a 1.5 T magnetic field. The outermost sub-detector made up of resistive plate chambers and scintillators is the K_L^0 and muon (KLM) detector.

These sub-detectors are used to reconstruct final state particles consisting of, but not limited to electrons, photons, pions, kaons, K_L^0 's and muons. Detector validation studies are presented in terms of the properties of particle studied in: (i) samples containing electrons, photons and pions; (ii) samples containing pions and kaons; and (iii) samples containing K_L^0 's and muons. Comparisons are made with 3 versions of the Geant4 toolkit: v10.1.2, v10.5.1 and v10.6.1. Although several hundreds of distributions with full detector simulation have been produced to validate the physics lists in each of these 3 versions, only a handful of them are presented in this note, for the sake of brevity.

3.1 Electrons, pions and photons

A sample of 100K Monte Carlo generated events of the process $e^-e^+ \rightarrow \tau^-\tau^+$ was used to validate the detector response of electrons, pions and photons. Half of the sample consists of $\tau^- \rightarrow e^-\bar{\nu}_e\nu_\tau$, $\tau^+ \rightarrow \pi^+\pi^0\nu_\tau$, $\pi^0 \rightarrow \gamma\gamma$ decays, and the other half consists of the charge-conjugate process. A hemisphere-based topological event selection consisting of one charged track in each hemisphere is used to reconstruct these events. One hemisphere is required to contain a charged track which has a ratio of energy deposited in the ECL (E) to the momentum measured in the CDC (p) of greater than 0.8. This is identified as the electron candidate. The other hemisphere is required to contain a track with (E/p) less than 0.8, which is identified to be the charged pion candidate, associated with two photons reconstructed in the same hemisphere with invariant mass ($m_{\gamma\gamma}$), consistent with that of a neutral pion, e.g. $m_{\gamma\gamma} \in [115, 155]$ MeV.

Two characteristic features of the energy deposited in the ECL are presented here for the reconstructed electron, pion and photon candidates: These features are used to differentiate charged pions from electrons/photons. One feature is the lateral moment of the shower shape of the corresponding candidate, and the other feature is the ratio of the energy deposited in the most energetic cluster divided by the total energy deposited in the surrounding 3x3 matrix (E1E9). Simulated distributions of lateral moment and E1E9 for electrons, photons and pions are shown in Figures 1, 2 and 3, respectively, for *FTFP_BERT* and *Belle II* physics lists. The distributions are found to be consistent with one another within the statistical precision studied, which are much smaller than the respective systematic uncertainties on these quantities. Detailed analysis of systematic uncertainties in particle identification are deferred to separate forthcoming publications from the *Belle II* experiment.

3.2 Pions and kaons

The identification of long-lived particles such as charged pions and kaons is a necessary step in reconstructing their parent particles in the Belle II environment. First, a charged hadron candidate is selected by the requirement of a track in the CDC and/or VXD. Next, a likelihood for each particle hypothesis is constructed using the information from all the particle identification detectors. The constructed likelihoods are compared among one another to select the best hypothesis for the track. To validate the *Belle II* physics list, we checked the individual contribution of the TOP and the ARICH to these particle identification likelihoods.

The TOP detector design is based on the detection of internally reflected Cherenkov light (DIRC) detector [2, 3]. The timing information and spatial distribution of Cherenkov photons produced by a charged particle is used to identify its type. The response of the TOP is studied using a simulated sample of 4000 events corresponding to the process $B^0 \rightarrow K^+\pi^-$ decays from decays of the $\Upsilon(4S)$ resonance produced in e^-e^+ annihilations. The differences in likelihood for a track to be identified as a pion or kaon according to the particle identification

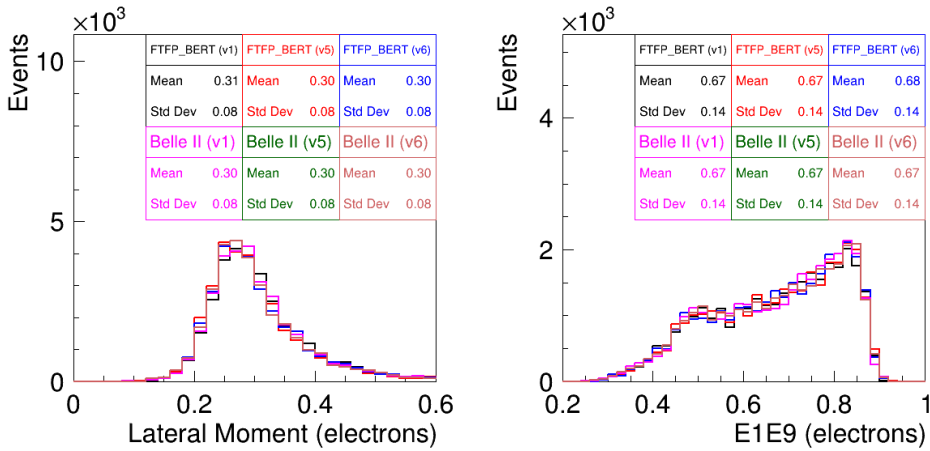


Figure 1. Simulated distributions of lateral moment and E1E9 for electrons with 3 versions of the Geant4 toolkit: v10.1.2, v10.5.1 and v10.6.1, labelled as v1, v5 and v6, respectively, for *FTFP_BERT* and *Belle II* physics lists.

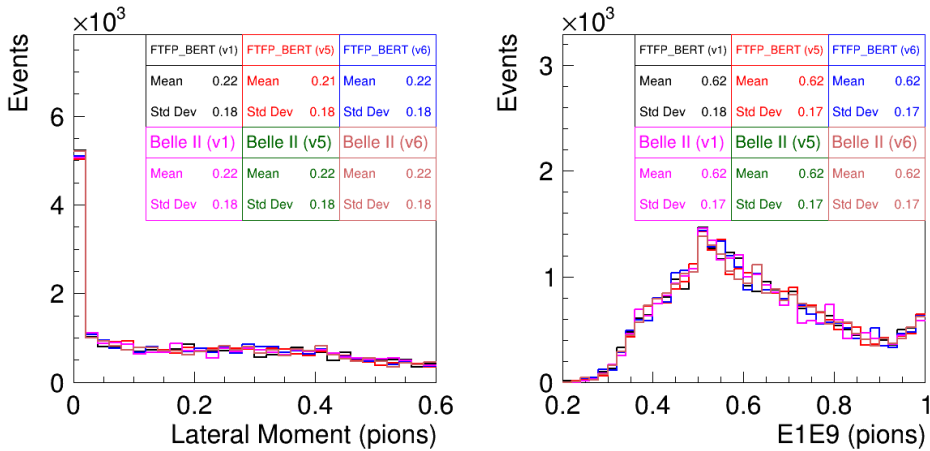


Figure 2. Same as Figure 1, but for pions.

algorithm based solely on the information from TOP detector are studied for tracks with momentum greater and less than 2 GeV in Figure 4. The distributions are found to be consistent between the two physics lists in the three versions of Geant4 under consideration.

Similarly, directions of the Cherenkov light photons in the ARICH are used to identify the particle type of a charged track [2, 3]. In Figure 5, the response in the ARICH detector is studied in terms of the difference in particle detection likelihood using a sample of 25000 events each, for pions and kaons produced with a particle gun Monte Carlo generator, respectively.

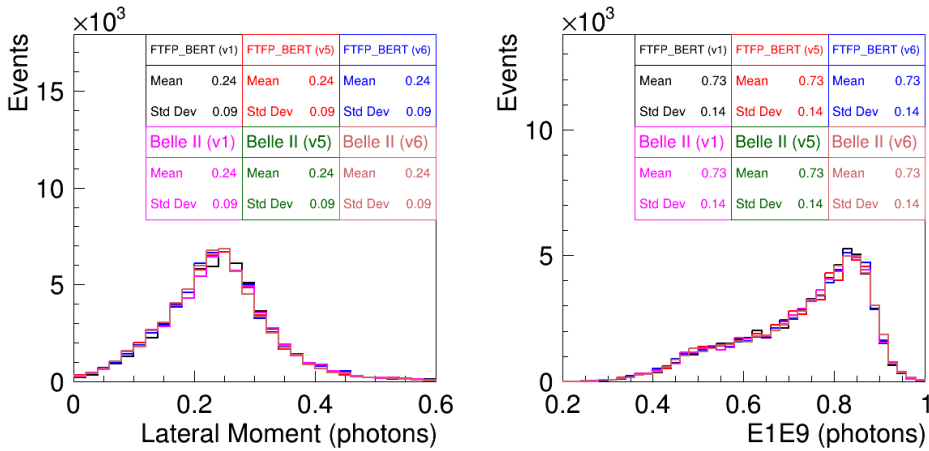


Figure 3. Same as Figure 1, but for photons.

3.3 K_L^0 s and muons

The KLM detector, the outermost component of the Belle II, is mainly used to differentiate K_L^0 s and muons. The KLM consists of glass-electrode resistive plate chambers (RPC) interleaved with the iron plates. The iron plates are part of the 1.5 T solenoid yoke [2, 3]. To validate the response in the KLM for K_L^0 s, a sample of 5000 events containing $B^0 \rightarrow J/\psi K_L^0$ decays was used, while a sample of 10000 events was produced with a particle gun generator for muons. The number of KLM clusters per generated particle and the energy deposited in the KLM detector are found to be consistent among the tested versions of the Geant4 toolkit in Figure 6.

4 Timing comparisons between the physics lists

More recent versions of the Geant4 toolkit provide faster geometry libraries than version v10.1.2. Therefore, it was decided to use Geant4 version v10.6.1 as well as the *Belle II* physics list to optimize the *basf2* performance.

For each combination of the Geant4 toolkit version and the physics list option, event samples corresponding to the $e^-e^+ \rightarrow b\bar{b}$ process were produced and the CPU consumption time used to produce each event was recorded. The analysis on this event timing information is shown in Figure 7. The bottom-right plot in Figure 7 is the summary of the comparison over the combinations. The systematic uncertainties in the summary plot are obtained by running 50 sets of jobs of 1000 events for each combination, and fitting the average event time obtained from each of the 50 sets to a Gaussian distribution.

While moving from the *FTFP_BERT* to the *Belle II* physics list provides a reduction of 21 to 27% in event timing, newer versions of the Geant4 toolkit also reduce the time required to simulate events. Combining both changes, an improvement of $\sim 44\%$ is expected for simulations carried out within *basf2* environment of the Belle II experiment.

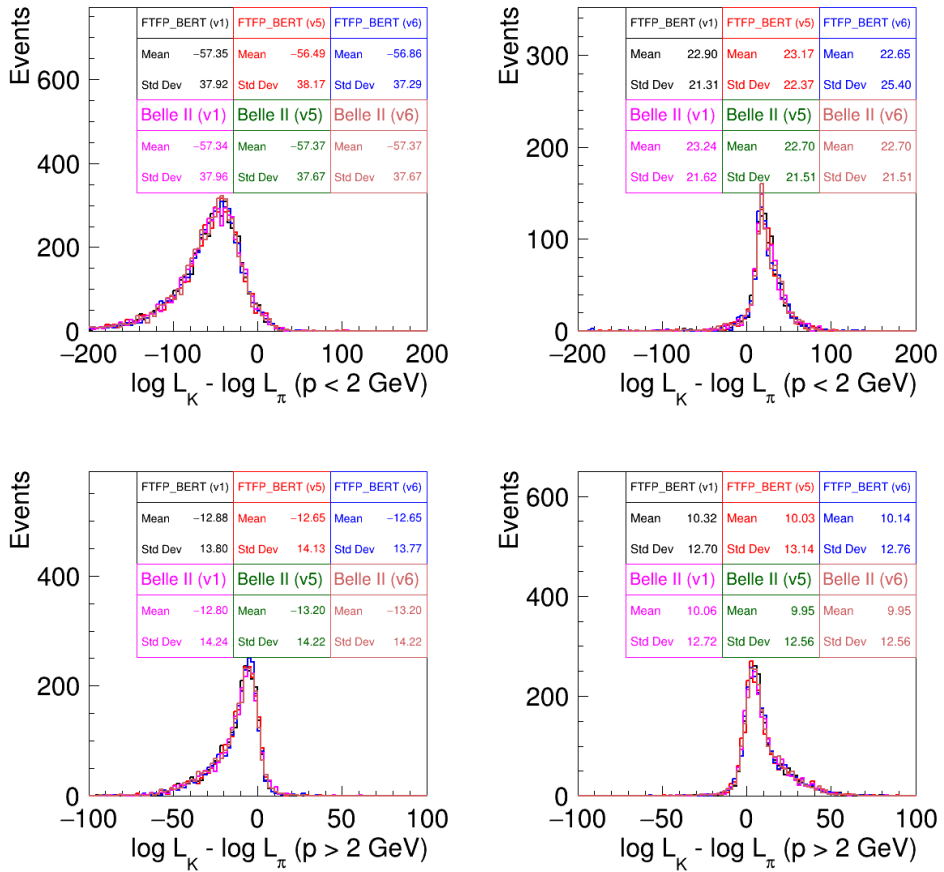


Figure 4. Simulated distributions of differences in particle detection likelihood obtained with the TOP detector for pions (left) and kaons (right) with 3 versions of the Geant4 toolkit: v10.1.2, v10.5.1 and v10.6.1, labelled as v1, v5 and v6, respectively, for *FTFP_BERT* and *Belle II* physics lists, corresponding to tracks with momentum less than (top) and greater than (bottom) 2 GeV.

5 Summary

We find that the implementation of the *Belle II* physics list is an effective improvement of the present defaults in the Geant4 toolkit in the *basf2* computing framework. It reduces the load on available computing resources without degrading the physics performance of the Belle II experiment.

We thank the KEK computer group. We acknowledge support from National Research Foundation (NRF) of Korea Grants No. 2016R1D1A1B02012900 and No. 2019K1A3A7A09034982; Foreign Large-size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD.

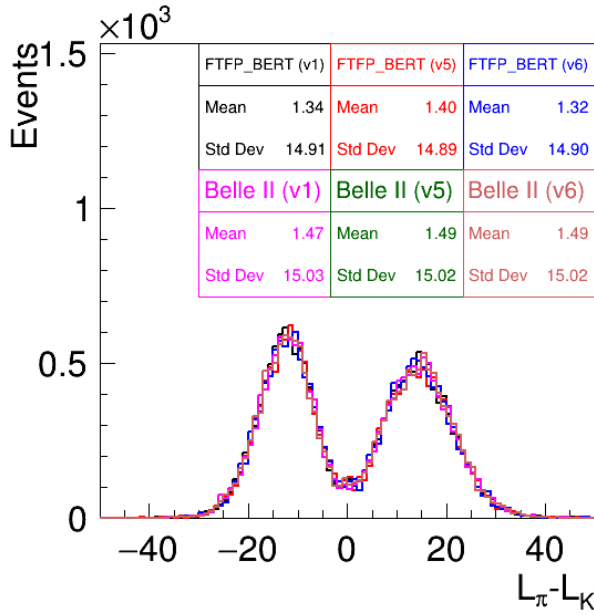


Figure 5. Simulated distributions of differences in particle detection likelihood obtained with the ARICH detector with 3 versions of the Geant4 toolkit: v10.1.2, v10.5.1 and v10.6.1, labelled as v1, v5 and v6, respectively, for *FTFP_BERT* and *Belle II* physics lists.

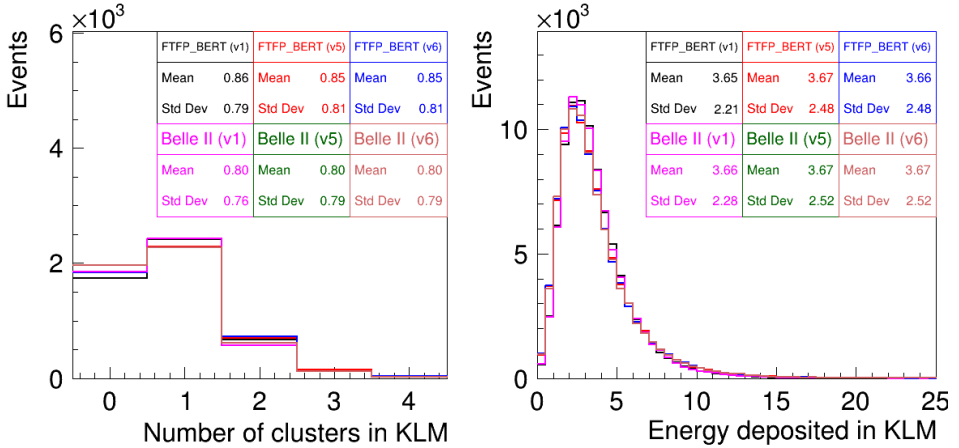


Figure 6. Simulated distributions of number of KLM clusters per generated particle and the energy deposited in the KLM detector for K_L^0 's (left) and muons (right) with 3 versions of the Geant4 toolkit: v10.1.2, v10.5.1 and v10.6.1, labelled as v1, v5 and v6, respectively, for *FTFP_BERT* and *Belle II* physics lists.

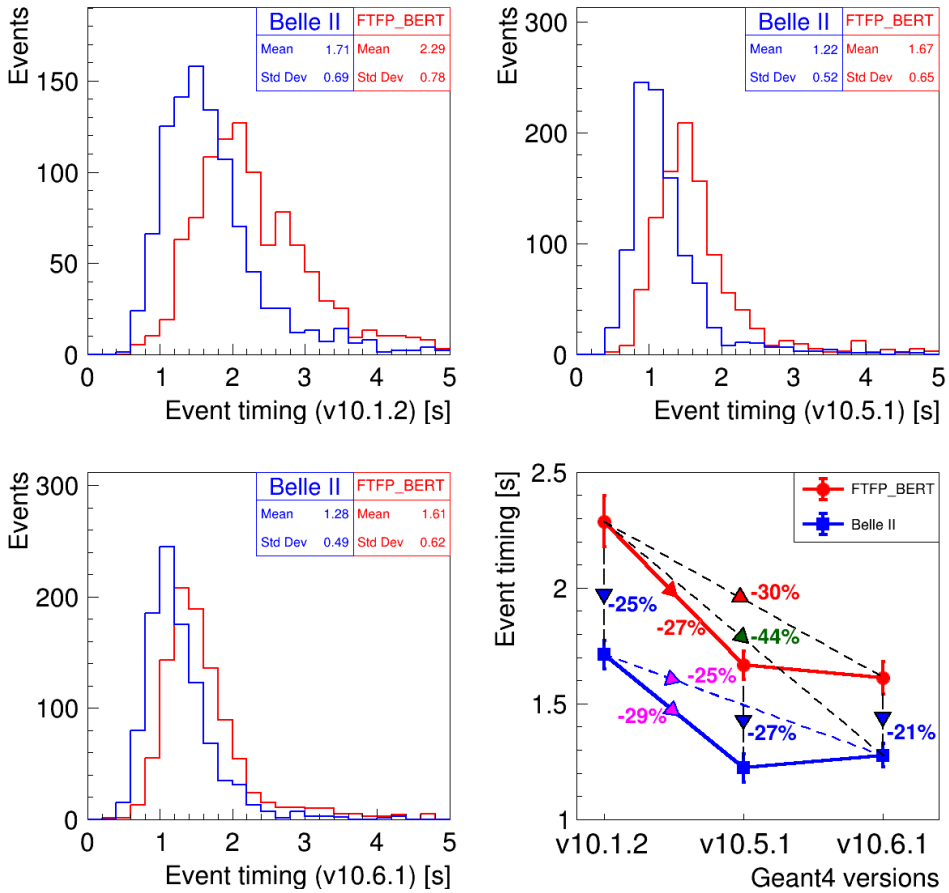


Figure 7. Average time to simulate a $b\bar{b}$ event for Geant4 toolkit versions v10.1.2, v10.5.1 and v10.6.1 is shown at top-left, top-right and bottom-left, respectively, for *FTFP_BERT* and *Belle II* physics lists. The bottom-right plot shows a summary of reductions in event timing in seconds.

References

- [1] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, N. Braun, *Comput. Softw. Big Sci.* **3** 1 (2019).
- [2] T. Abe *et al.*, [Belle-II], *Belle II Technical Design Report*, (KEK-REPORT-2010-1, 2010), [arXiv:1011.0352 [physics.ins-det]].
- [3] E. Kou *et al.*, *Prog. Theor. Exp. Phys.* **2019** 123C01 (2019).
- [4] J. Allison *et al.*, *Nucl. Instrum. Methods. Phys. Res. A* **835** 186 (2016).
- [5] J. Allison *et al.*, *IEEE. Trans. Nucl. Sci.* **53** 270 (2006).
- [6] S. Agostinelli *et al.*, *Nucl. Instrum. Methods. Phys. Res. A* **506** 250 (2003).