# Report on G4-Med, a Geant4 benchmarking system for medical physics applications developed by the Geant4 Medical Simulation Benchmarking Group

Arce, P.; Bolst, D.; Bordage, M. C.; Brown, J. M.C.; Cirrone, P.; Cortés-Giraldo, M. A.; Cutajar, D.; Desorgher, L.; Wright, D. H.; More Authors

**DOI**
10.1002/mp.14226

**Publication date**
2021

**Document Version**
Final published version

**Published in**
Medical Physics

# Magphan Phantoms for MR for radiation therapy and quantitative imaging applications.


Smári


Sub-voxel geometric distortion measurements and critical image quality metrics engineered for ease of use in the clinical workflow.

The Phantom Laboratory manufactures high-precision phantoms coupled with Smári image analysis service and innovative custom solutions for the medical imaging and radiation therapy fields.

**Click to see our latest phantoms and schedule a demo of our Smári image analysis service.**

# Report on G4-Med, a Geant4 benchmarking system for medical physics applications developed by the Geant4 Medical Simulation Benchmarking Group

P. Arce
*CIEMAT, Madrid, Spain*

D. Bolst
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

M.-C. Bordage
*CRCT (INSERM and Paul Sabatier University), Toulouse, France*

J. M. C. Brown
*Department of Radiation Science and Technology, Delft University of Technology, Delft, The Netherlands*

P. Cirrone
*INFN LNS, Catania, Italy*

M. A. Cortés-Giraldo
*Universidad de Sevilla, Sevilla, Spain*

D. Cutajar
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

G. Cuttone
*INFN LNS, Catania, Italy*

L. Desorgher
*Institute of Radiation Physics (IRA), Lausanne University Hospital, Lausanne, Switzerland*

P. Dondero
*SWHARD srl, Genova, Italy*

A. Dotti
*SLAC National Accelerator Laboratory, Stanford, CA, USA*

B. Faddegon
*University of California, San Francisco, CA, USA*

C. Fedon
*Radboud University Medical Center, Nijmegen, The Netherlands*

S. Guatelli[a)]
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

S. Incerti
*Université de Bordeaux, CNRS/IN2P3, UMR5797, Centre d'Études Nucléaires de Bordeaux Gradignan, Gradignan, France*

V. Ivanchenko
*Tomsk State University, Tomsk, Russian Federation*
*CERN, Geneva, Switzerland*

D. Konstantinov
*NRC "Kurchatov Institute" - IHEP, Protvino, Russian Federation*

I. Kyriakou
*Medical Physics Laboratory, University of Ioannina, Ioannina, Greece*

G. Latyshev
*NRC "Kurchatov Institute" - IHEP, Protvino, Russian Federation*

A. Le
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

C. Mancini-Terracciano
*Physics Dep., University of Rome "Sapienza" and INFN, Sec. of Rome, Rome, Italy*

M. Maire
*LAPP, IN2P3, Annecy, France*

  

A. Mantero
*SWHARD srl, Genova, Italy*

M. Novak
*CERN, Geneva, Switzerland*

C. Omachi
*Nagoya Proton Therapy Center, Nagoya, Japan*

L. Pandola
*INFN LNS, Catania, Italy*

A. Perales
*Medical Physics Department of Clínica Universidad de Navarra, Pamplona, Spain*

Y. Perrot
*IRSN, Fontenay-aux-Roses, France*

G. Petringa
*INFN LNS, Catania, Italy*

J. M. Quesada
*Universidad de Sevilla, Sevilla, Spain*

J. Ramos-Méndez
*University of California, San Francisco, CA, USA*

F. Romano
*INFN Catania Section, Catania, Italy*
*Medical Physics Department, National Physical Laboratory, Teddington, UK*

A. B. Rosenfeld
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

L. G. Sarmiento
*Lund University, Lund, Sweden*

D. Sakata
*Centre For Medical Radiation Physics, University of Wollongong, Wollongong, Australia*

T. Sasaki
*KEK, Tsukuba, Japan*

I. Sechopoulos
*Radboud University Medical Center, Nijmegen, The Netherlands*
*Dutch Expert Center for Screening (LRCB), Nijmegen, The Netherlands*

E. C. Simpson
*Department of Nuclear Physics, Research School of Physics, Australian National University, Canberra, Australia*

T. Toshito
*Nagoya Proton Therapy Center, Nagoya, Japan*

D. H. Wright
*SLAC National Accelerator Laboratory, Stanford, CA, USA*

(Received 10 July 2019; revised 26 April 2020; accepted for publication 30 April 2020;
published xx xxxx xxxx)

**Background:** Geant4 is a Monte Carlo code extensively used in medical physics for a wide range of applications, such as dosimetry, micro- and nanodosimetry, imaging, radiation protection, and nuclear medicine. Geant4 is continuously evolving, so it is crucial to have a system that benchmarks this Monte Carlo code for medical physics against reference data and to perform regression testing.
**Aims:** To respond to these needs, we developed *G4-Med*, a benchmarking and regression testing system of Geant4 for medical physics.
**Materials and Methods:** *G4-Med* currently includes 18 tests. They range from the benchmarking of fundamental physics quantities to the testing of Monte Carlo simulation setups typical of medical physics applications. Both electromagnetic and hadronic physics processes and models within the

prebuilt Geant4 physics lists are tested. The tests included in *G4-Med* are executed on the CERN computing infrastructure via the use of the geant-val web application, developed at CERN for Geant4 testing. The physical observables can be compared to reference data for benchmarking and to results of previous Geant4 versions for regression testing purposes.

**Results:** This paper describes the tests included in *G4-Med* and shows the results derived from the benchmarking of Geant4 10.5 against reference data.

**Discussion:** Our results indicate that the Geant4 electromagnetic physics constructor *G4EmStandardPhysics_option4* gives a good agreement with the reference data for all the tests. The *QGSP_BIC_HP* physics list provided an overall adequate description of the physics involved in hadron therapy, including proton and carbon ion therapy. New tests should be included in the next stage of the project to extend the benchmarking to other physical quantities and application scenarios of interest for medical physics.

**Conclusion:** The results presented and discussed in this paper will aid users in tailoring physics lists to their particular application.  [https://doi.org/10.1002/mp.14226]

Key words: benchmarking, Geant4, medical physics, Monte Carlo

## 1. INTRODUCTION

Geant4[1–3] is a Monte Carlo (MC) toolkit describing particle transport and interactions in matter. Originally developed for high energy physics, it was later extended to the low energy domain (down to the eV scale) for medical physics and space science applications. Today Geant4 is widely used in medical physics in critical applications such as verification of radiotherapy treatment planning systems, and the design of equipment for radiotherapy and nuclear medicine. It is also used in medical imaging for dosimetry, to improve detectors and reconstruction algorithms, and for radiation protection assessments.[4,5] In the medical physics domain, Geant4 can be used in stand-alone applications or via software tools like GAMOS,[6] GATE,[7] PTSim,[8,9] and TOPAS.[10] Given the extensive use of Geant4 in medical physics, a systematic benchmarking of the accuracy of the Geant4 physics models in this domain is paramount. This is crucial in a field where MC simulations are often regarded as a gold standard.

In this work, we describe *G4-Med*, a group of currently 18 tests, executed on the CERN computing infrastructure to benchmark and to perform regression tests of new development tags and public releases of Geant4.

The tests are contributed by the members of the Geant4 Medical Simulation Benchmarking Group[11] to benchmark Geant4 for medical physics applications. They range from the benchmarking of fundamental physics quantities to more complex simulations typical of medical physics applications. We show the results of the tests obtained for Geant4 version 10.5 and report on their benchmarking against reference data. For the sake of brevity, we show only a subset of the results. The full results can be downloaded using the *geant-val* web interface.[12] *G4-Med* has been set up with the aim to assess the appropriateness of any Geant4 release in the medical domain and to monitor the impact of software development, including physics refinements, on physical quantities and applications of interest for medical physics. The results of the benchmarking are also intended to provide recommendations on the most appropriate physics configuration provided by Geant4 with the prebuilt physics constructors and physics lists[3] to adopt in specific user applications. The project documentation can be found at the *G4-Med* web page.[11]

Section 2 describes the general method adopted in this study. This is then followed by sections that describe tests devoted to benchmarking electromagnetic physics (10 tests, Section 3), hadronic physics (3 tests, Section 4), and testing both these physics sets together (5 tests, Section 5). For clarity, Table 1 lists the tests of *G4-Med*, their name in the *geant-val* interface[12] and corresponding subsections. The subsections briefly describe the tests and their results. This table also reports the source of the tests. We conclude our report with a summary of our findings in Section 6.

This report has gone through the peer review and approval processes of the Geant4 Editorial Board, which reviews all scientific papers emerging from the Geant4 Collaboration (Geant4 Collaboration website[13]).

## 2. METHODOLOGY

All tests included in *G4-Med* have been integrated into the *geant-val* environment[14] developed at CERN to perform Geant4 benchmarking and regression testing. The *geant-val* system provides a convenient web-based validation tool for Geant4 developers. It allows for storage of the Geant4 tests results in analysis objects, such as histograms and scatter plots, together with meta information, such as: Geant4 version, name of the test, energy and momentum of the incident particle, name of the target/detector, and physics list used in the simulation. The main web interface[12] allows for visual and statistical comparison among results produced using different versions of Geant4, or a comparison with reference experimental results. In addition, users can download plots or data in various formats. The tests are executed automatically for all the global development tags and public releases of

TABLE I. Tests of *G4-Med* with their name in the *geant-val* interface and the subsections of this report with their description and results.

| Test | Name in *geant-val* interface[12] | Source | Subsection |
|---|---|---|---|
| Photon attenuation | *PhotonAttenuation* | Amako et al.[15] | 3.A |
| Electron electronic stopping power | *ElectDEDX* | * | 3.B |
| Electron backscattering | *ElectBackScat* | Dondero et al.[17] | 3.C |
| Electron forward scatter from foils at 13 and 20 MeV | *ElecForwScat* | Faddegon et al.[18] | 3.D |
| Bremsstrahlung from thick targets | *Bremsstrahlung* | Faddegon et al.[19] | 3.E |
| Fano cavity | *Fano cavity* | ** | 3.F |
| Low-energy electron dose point kernels | *LowEElectDPK* | ** | 3.G |
| Microdosimetry | *microyz* | **Kyriakou et al.[20] | 3.H |
| Brachytherapy | *Brachy-Ir* | *** | 3.I |
| Monoenergetic x-ray internal breast dosimetry | *Mammo* | Fedon et al.[21,22] | 3.J |
| Nucleus–nucleus hadronic inelastic scattering cross sections | *NucNucInelXS* | NA | 4.A |
| 62 MeV/μ $^{12}$C fragmentation | *LowEC12Frag* | Mancini et al[23] | 4.B |
| Charge-changing cross section for 300 MeV/μ carbon ions | *C12FragCC* | Toshito et al[24] | 4.C |
| 67.5 MeV proton Bragg curves in water | *LowEProton BraggBeak* | Faddegon et al[25] | 5.A |
| Light ion Bragg peak curves | *LightIonBraggPeak* | NA | 5.B |
| Neutron yield of protons with energy 113 and 256 MeV and 290 MeV/u carbon ions | *ProtonC12 NeutronYield* | NA | 5.C |
| Fragmentation of a 400 MeV/μ $^{12}$C ion beam in water | *FragTest* | Bolst et al[26] | 5.D |
| Test on cell survival modeling for proton therapy | *Hadrontherapy* | ***, Petringa et al[27] | 5.E |

*: tests already executed by Geant4 developers for regression testing purposes; **: tests released as Geant4 *extended* examples; ***: tests released as Geant4 *advanced* examples.

Geant4 at the CERN computing facility. All results derived from the benchmarking and regression testing, including plots and statistical analysis, are generated automatically and are accessible through the web interface.[12]

In general, the tests use the same set of electromagnetic (EM) and hadronic physics constructors, with exceptions for tests aimed at validating the Geant4 EM multiple and single scattering models and specific hadronic cross sections and models. Simulation parameters, such as the secondary particles production threshold (cut) or the maximum step size, were appropriately set for each simulation scenario and are included in the description of each test.

In all the tests, the % difference between the results of the Geant4 simulations and reference data was calculated, then compared to the 1 $\sigma$ uncertainty affecting the reference data (called $\sigma_{ref}$ in the following sections). If the difference was smaller than $\sigma_{ref}$ the agreement was considered satisfactory.

## 3. ELECTROMAGNETIC PHYSICS BENCHMARKING TESTS

This section describes the 10 tests in *G4-Med* where only the Geant4 EM physics component is activated and all the other processes are disabled. The tested Geant4 EM physics constructors are briefly described here as released within Geant4 10.5.

In all the EM constructors under investigation, the set of EM processes are the same. Photon interactions include the photoelectric effect, Compton and Rayleigh scattering, and pair/triplet production (gamma conversion). Electron and positron interactions are ionization, bremsstrahlung, and elastic scattering. Positrons can also annihilate at rest and in flight. Protons and heavier ions have ionization, elastic scattering and nuclear stopping.

*G4EmStandardPhysics*, *G4EmLivermorePhysics*, *G4EmPenelopePhysics*, *G4EmStandardPhysics_option3* and *G4EmStandardPhysics_option4* EM constructors, called here *Opt0*, *Livermore*, *Penelope*, *Opt3*, and *Opt4*, respectively, were considered in this work and correspond to different combinations of models deriving from either Geant4 *Standard*, *Livermore* or *Penelope* packages. The EM physics constructors and models, reported briefly in Table 2, are described in detail in the Geant4 Physics Reference Manual.

The *Standard* sublibrary describes electromagnetic interactions in the range between 1 keV and 10 PeV, and is focused on high energy physics applications such as the simulation of LHC experiments.[3,28] The *Livermore* models provide a more accurate description of EM physics processes in the low energy domain. They are based on the Livermore Evaluated Data libraries and are documented in Ivanchenko et al,[29] Allison et al,[30] and Chauvie et al.[31] *Penelope* features the specific low-energy models for electrons, positrons, and photons, originally developed for the PENELOPE Monte Carlo code[32] and then implemented in Geant4. They can describe particle interactions down to 100 eV.

The *Opt4* constructor contains a combination of models for each EM physics process deemed to offer the best performance in term of precision at the cost of CPU efficiency.[33] The benchmark results of the current paper will also be used to discuss the expected better physical accuracy of *Opt4* against the other EM constructors under investigation. *Opt4* uses the *G4GoudsmitSoundersonMscModel* to describe multiple

TABLE II. Geant4 physics models to describe EM physics processes in Geant4 EM constructors under investigation.

| Geant4 | *Opt0* | *Opt3* | *Opt4* | *Livermore* | *Penelope* |
|---|---|---|---|---|---|
| Rayleigh scattering and photoelectric effect | Livermore | | | | PENELOPE |
| Compton scattering | Standard | *G4KleinNishina Model* | *G4LowEPComptonModel* for E < 20 MeV* Brown et al[36] | Livermore for E <1 GeV* | PENELOPE for E <1 GeV* |
| Gamma conversion | Standard | Standard | PENELOPE for E <20 MeV Standard for E >20 MeV | *G4BetheHeitler5DModel* for E <1 GeV, Standard for E >1 GeV Bernard et al 2013[37] | PENELOPE for E<1 GeV, Standard for E >1 GeV |
| $e^-$ and $e^+$ ionisation | Standard | Standard | Livermore for $e^-$ for E <100 keV, PENELOPE for $e^+$ for E <100 keV, Standard for E >100 keV | Livermore for E <100 keV Standard for E >100 keV | PENELOPE |
| $e^-$ and $e^+$ bremsstrahlung | Standard | *G4SeltzerBergerModel* for E <1 GeV, *G4eBremsstrahlungRelModel* for E >1 GeV | | | PENELOPE |
| $e^+$ annihilation | Standard | | | | PENELOPE |
| $e^-$ and $e^+$ multiple scattering | Urban model for E < 100 MeV (Ivanchenko et al[39]), Wentzel model for E >100 MeV | Urban model | Goudsmit–Saunderson model (Incerti et al 2018a[38]) for E <100 MeV Wentzel model for E >100 MeV | | |
| Coulomb scattering | on | off | On | | |
| Bremsstrahlung angular distribution | *ModifiedTsai* | *2BS* | | | PENELOPE |

*: *G4KleinNishinaModel* for higher energies. Geant4 10.5 is considered. For details on the models the reader should refer to the Geant4 Physics Reference Manual.[28]

TABLE III. Geant4 EM parameters of the EM constructors under investigation.

| Geant4 EM parameter | *Opt0* | *Opt3* | *Opt4 livermore penelope* | *SS* | *GS* | *WVI* |
|---|---|---|---|---|---|---|
| *Minimum energy* (eV) | 100. | 10. | 10. | 100. | 100. | 10. |
| *Lowest electron energy* (keV) | 1. | 0.1 | 0.1 | 0.01 | 1. | 0.01 |
| *Number of bins per decade* | 7 | 20 | 20 | 7 | 7 | 20 |
| *Angular generator* | False | True | True | False | False | True |
| *Mott corrections* | False | True | True | False | False | True |
| *dRoverRange* for $e^-$ and $e^+$ | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| *finalRange* for $e^-$ and $e^+$ (mm) | 1. | 0.1 | 0.1 | 1. | 1. | 0.1 |
| *dRoverRange* for muons and hadrons | 0.1 | 0.05 | 0.05 | 0.1 | 0.1 | 0.05 |
| *Lateral displacement* for muons and hadrons | False | True | True | False | False | True |
| *Skin* | 1 | 1 | 3 | 1 | 1 | 0 |
| *Range factor* | 0.04 | 0.04 | 0.08 | 0.04 | 0.06 | 0.04 |
| *Theta* (rad) | $\pi$ | $\pi$ | $\pi$ | 0. | $\pi$ | 0.15 |
| Luorescence | Off | On | On | On | Off | On |
| Auger electrons and PIXE modeling | Off | Off | Off | On | Off | Off |

Geant4 10.5 is considered.

scattering for e± below 100 MeV and the ICRU73 stopping power data for ions heavier than helium (ICRU report No.73[34]).

In all the EM constructors, the multiple scattering of protons, muons, and other hadrons is modeled with the Wentzel model and the single scattering. In the case of ions, the Urban model is used for all energies.

*Opt3*, *Livermore*, and *Penelope* constructors are considered as relatively accurate but less CPU demanding for medical applications. We therefore expect the *Opt4* constructor to perform best overall in our benchmarking system, but we nevertheless study *Opt3* and *Livermore* constructors, given their appeal in terms of computational efficiency.

For electron scattering and Fano cavity tests, additional EM physics constructors are used, the *G4EmStandardPhysicsSS*, *G4EmStandardPhysicsGS*, *G4EmStandardPhysicsWVI*, called here *SS*, *GS*, and *WVI*, respectively. They have the same physics parameters of *Opt0*, apart from different modeling of either the Coulomb scattering or the multiple scattering. In *SS*, the single scattering model is used for the Coulomb scattering. *GS* and *WVI* adopt the Goudsmit–Saunderson (GS) and the Wentzel models, respectively, for multiple scattering. These physics constructors are mainly for Geant4 internal tests allowing study of various EM models alone, and are not meant for production physics configurations.

For completeness, Table 3 shows the values of EM parameters relevant for medical physics applications, listed for each of the EM constructors under investigation. The *minimum energy* is the minimum kinetic energy used to build the EM physics tables. The *lowest electron energy* parameter defines tracking cut for electrons and positrons. If after a step in a material, the particle energy is below this cut, this energy is added to the energy deposition and this particle is stopped. This parameter may be changed by the user and can be set to zero. The *number of bins per decade* is the number of bins per decade to be used when building the physics tables. The *angular generator* parameter activates the angular generator interface of the ionization process. The *Mott corrections* activate the Mott corrections in the electron multiple scattering. *dRoverRange* and *finalRange* are parameters of the method *SetStepFunction* used in the modeling of the multiple scattering. The lateral displacement due to multiple scattering is enabled by default for electrons and positrons, while it is set up differently in the EM constructors for muons and hadrons (see the *Lateral displacement* parameter in Table 3). The *Skin* and *Range factor* parameters are used to limit the step in multiple scattering of electrons and positrons. The *Theta* parameter is the angular limit between single and multiple scattering. The last two rows of Table 3 report if the atomic de-excitation, including fluorescence x ray and Auger electrons emission, is active in the EM constructors.

All model parameters are unchanged with respect to their defaults, that is, no fine tuning or optimization is performed within the constructor.

## 3.A. Photon attenuation test

### 3.A.1. Simulation setup

The *Photon Attenuation* test, described and published in Amako et al,[15] has the name of *PhotonAttenuation* in the *geant-val* web interface.[12] This Geant4 application calculates the attenuation coefficients of photons with energy between 1 keV and 1 GeV striking a water target with normal incidence. The number of incident photons $N$ emerging without interacting in the target is counted and then the attenuation coefficient is calculated as $\frac{\mu}{\rho} = -\frac{1}{\rho d} ln\left(\frac{N}{N_0}\right)$, where  is the target density, $d$ the target thickness and $N_0$ the number of incident photons.

The test calculates the photon attenuation coefficient in water of individual photon processes: Rayleigh scattering, photoelectric effect, Compton scattering and gamma conversion, as well as the total one. The photon attenuation coefficients are calculated in water, modeled as Geant4 NIST material G4_WATER (documented in the Geant4 Application Developer Manual[39]).

We compare the results of the Geant4 simulations to reference data of the United States National Institute of Standards and Technologies (NIST), included in the NIST XCOM database available online..[16,40] The NIST XCOM database[16] was chosen as it is often used as a reference in medical radiation

physics. It provides photon attenuation coefficients between 1 keV and 100 GeV for all the elements of the periodic table, calculated with a theoretical approach based on Hubbell et al.[41] The quoted uncertainty for the reference data is about 1% for high energies and away from the atomic edges, while it can be substantially larger, up to 10–20%, for energies close to the atomic absorption edges.[42] From here, it was decided to adopt an uncertainty ($\sigma_{ref}$) of up to 10% and 1% for energies below and above 100 keV, respectively.

The number of histories (photon events) in the simulations is adjusted depending on the energy at which the attenuation coefficients are calculated in order to obtain simulation statistical uncertainties below 1%.

### 3.A.2. Results and discussion

Figure 1 shows the results concerning the total attenuation coefficient of photons in water, while Fig. 2 shows the attenuation coefficients for each individual photon process. The agreement between the Geant4 EM constructors and the reference data is summarized in Table 4.

For the total attenuation coefficient, all the Geant4 EM constructors agree with the reference data within the uncertainty of the NIST XCOM in the entire energy range under study (up to 1 GeV). The 2% local difference at 20 keV (see Fig. 2) is due to differences in the photoelectric and Compton cross sections with respect to the NIST XCOM data. This is

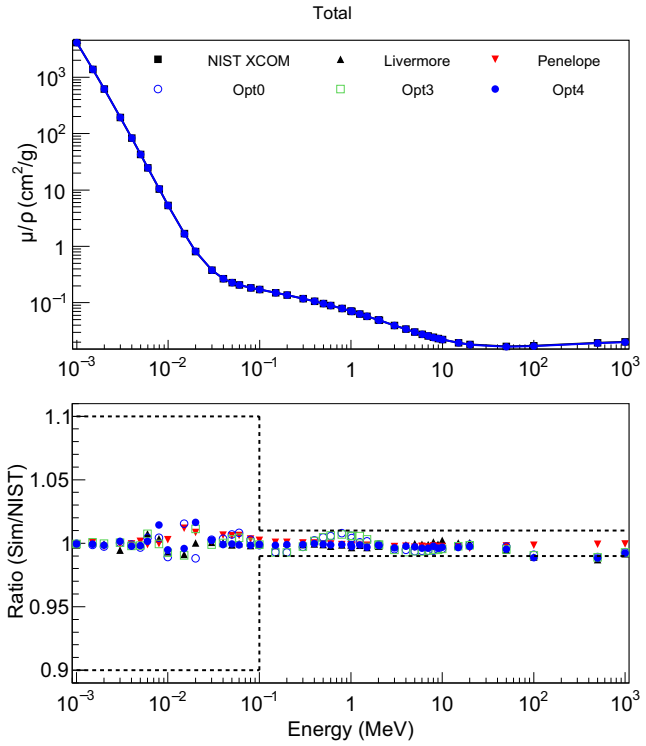


FIG. 1. Top plot: Total attenuation coefficient in water with respect to the photon energy. Bottom plot: Ratio of the results obtained with Geant4 and the NIST XCOM reference data. The black dashed lines in the bottom plot represent the uncertainties of the NIST XCOM data ($\sigma_{ref}$). The statistical uncertainty affecting the Geant4 simulation results is below 1%.

the point where we have the transition from photoelectric effect to Compton scattering as dominant process.

The results show a maximum difference of approximately 10% in the case of Rayleigh scattering for energies below few keV (see Table 4 for more details). Then such differences decrease to <1% for higher energies. This is due to different modeling of this process in the Livermore Evaluated Data libraries and NIST, highlighted in Amako et al[15] and Cirrone et al.[42] Among the EM constructors considered, *Opt0* is the one showing the biggest differences with respect to the reference data.

In the case of photoelectric effect attenuation coefficient, all the EM constructors agree with the reference data within 1 $\sigma_{ref}$ (1 $\sigma$ uncertainty of the reference data), with the exception at around 1–2 MeV, where the photoelectric attenuation coefficient changes significantly in slope and at high energies (above 500 MeV) where the photoelectric effect is not important anymore. There is a local difference of about 2% at 20 keV between the Geant4 EM constructors and the reference data, which contributes to the difference at this energy in the case of the total attenuation coefficient.

In the case of Compton scattering, an agreement within 1 $\sigma_{ref}$ was observed for the entire energy range under investigation for *Opt4*, with maximum differences of <1.5%. *Opt0*, *Opt3*, and *Penelope* show differences above 5% at low energies (still within 1 $\sigma_{ref}$, see details in Table 4). *Livermore* and *Penelope* show local differences of a few percent at around 500 MeV–1 GeV; however, here the Compton cross section is one order of magnitude lower than the gamma conversion cross section.

The gamma conversion attenuation coefficient vanishes below the pair production threshold of 1.022 MeV, as expected. In this case, we observed differences equal or above 5% close to the threshold energy for all the EM constructors. *Opt0*, *Opt3*, and *Livermore* have differences up to 3% in the energy range between 3 and 10 MeV and up to 1% for higher energies.

In general, *Opt3*, *Opt4*, *Livermore*, and *Penelope* have a good agreement with the reference data. The results are material dependent, but in water, all models are consistent with each other and with NIST XCOM within a few percent. A slightly overall better agreement with the reference data was found when considering *Opt4*. Nevertheless, it is important to note that the Geant4 EM physics constructors were compared here to theoretical data with modeling limitations. Therefore, this test should be extended in the future to include comparisons to available experimental measurements and other data libraries.

The results of this test are in agreement with those reported in Cirrone et al[42] and Amako et al.[15] The readers should consult these publications for results of this test when considering other target materials.

## 3.B. Electron electronic stopping power test

In this test, the electron electronic stopping power is calculated in a set of target materials and compared to the ESTAR

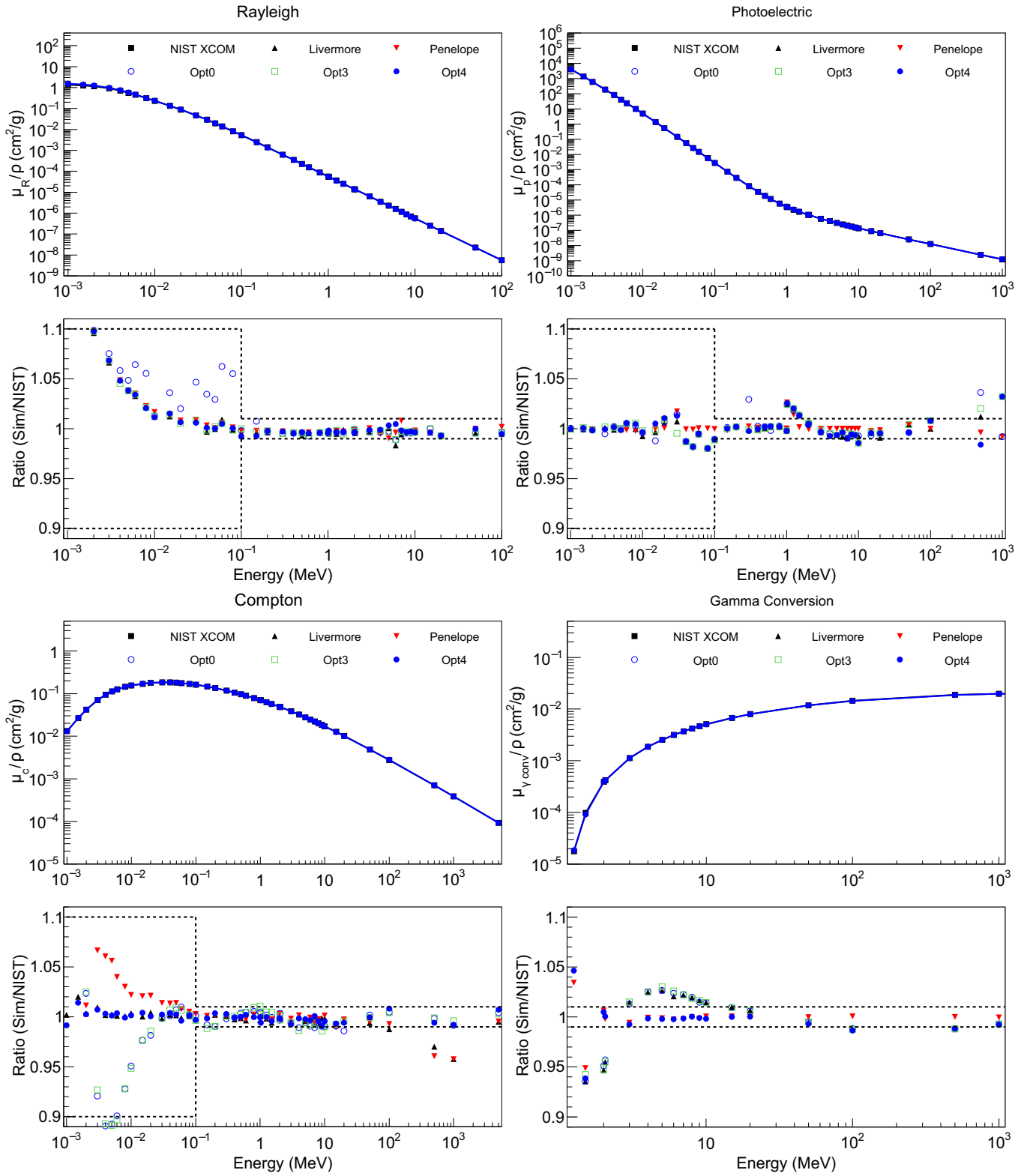


FIG. 2. Comparison of the attenuation coefficients for each individual photon process: Rayleigh scattering (top left), photoelectric effect (top right), Compton scattering (bottom left), gamma conversion (bottom right). The plots show the attenuation coefficients as calculated using Geant4 against the NIST XCOM data; the ratio of Geant4 results and the reference data is shown as well. The black dashed lines in the lower plot represent the uncertainties of the NIST XCOM data ($\sigma_{\text{ref}}$). The statistical uncertainty affecting the Geant4 simulation results is below 1%.

TABLE IV. Difference between Geant4 simulation results and NIST XCOM.

| Test | Difference with respect to the reference data | | | | |
|---|---|---|---|---|---|
| | *Opt0* | *Opt3* | *Opt4* | *Livermore* | *Penelope* |
| Total | $\leq 1\ \sigma_{ref}$ | | | | |
| Ray | $<1\sigma_{ref}$, in particular: <10% for E <10 keV, ≤ 5% for 10 keV < E <100 keV, < 1% for E >100 keV | <1 $\sigma_{ref}$, in particular: <10% for E ≤ 3 keV, ≤ 5% for 3 keV < E < 10 keV, < 1% for E > 10 keV | | | |
| Photo | | < 1 $\sigma_{ref}$, with exceptions at 1–2 MeV and E ≥ 500 MeV | | | |
| Comp | $\leq 1\sigma_{ref}$, in particular: >5% for E < 10 keV, ≤5% for 10 keV < E <20 keV, ≤ 1% for E > 20 keV | | $<1\sigma_{ref}$ | $\leq 1\sigma_{ref}$ | $\leq 1\sigma_{ref}$, in particular: > 5% for E < 6 keV,≤ 5% for 6 keV < E < 60 keV ≤ 1% for E > 60 keV |
| Conv | ≤5% for 2 MeV, ≤ 3% for 3 MeV < E < 10 MeV,≤ 1% for E > 10 MeV | | ≤ 1% for E ≥ 2 MeV (< 1 $\sigma_{ref}$) | ≤ 5% for 2 MeV, ≤ 3% for 3 MeV < E < 10 Me V ≤ 1% for E > 10 MeV | ≤ 1% for E ≥ 2 MeV (≤ 1 $\sigma_{ref}$) |

$\sigma_{ref}$ is the 1 $\sigma$ uncertainty of the reference data. Ray: Rayleigh scattering; Photo: photoelectric effect; Comp: Compton scattering; Conv: pair and triplet production.

collision stopping power data, that is available online in the NIST Reference Database.[43] Electronic and collision stopping powers are the same physical quantity; however, the ICRU Report No. 85[44] recommends the use of the specific term *electronic*. From this observation, the term *electronic* was adopted here.

This technical test was developed for the regression testing of Geant4 for both low energy and high energy physics applications to demonstrate the level of agreement between available models of electron stopping powers.

Here details of the simulation, results, and discussion are provided for electrons with an incident energy between 10 keV and 10 GeV. In the *geant-val* web interface,[12] the test has the name *ElectDEDX*.

### 3.B.1. Simulation setup

Monoenergetic electrons originate in the center of a target. The simulation tests only the ionization processes of the *Opt0*, *Penelope*, *Livermore*, *Opt3*, and *Opt4* EM constructors, in the case of no energy loss fluctuations and no generation of secondary particles.

Assuming small energy losses, the stopping power is calculated as the energy deposited in the first step of the track, divided by the true step length, which takes into account multiple scattering correction.

The electronic stopping power is calculated for targets made of Al, Ar, Cu, Au, and Pb, in order to represent a range of atomic numbers. Water is considered as a target as well because it is of interest for medical physics. The target materials are modeled from the Geant4 NIST material database, described in the Geant4 Application Developer Manual.[39]

The ESTAR electronic stopping power values are calculated based on Bethe theory, with a density-effect correction evaluated according to Sternheimer, as described in the documentation of the NIST Standard Reference Database.[43] The uncertainties of the ESTAR electronic stopping powers for electrons are estimated to be 1–2% above 100 keV, 2–3% in low atomic number materials, and 5–10% in high atomic number materials between 10 and 100 keV. At energies below 10 keV the stopping powers from ESTAR are expected to be too large due to the omission of shell corrections and are recognized not to be accurate.[45] Therefore, we do not report data below 10 keV.

### 3.B.2. Results and discussion

Figure 3 shows the electronic stopping power and ratio of the simulation results with respect to the ESTAR data libraries, in the case of water, Al and Au. These two elements have been chosen as they represent low and high atomic numbers, respectively. The simulation results do not have any statistical uncertainty because there is no multiple scattering, no energy loss fluctuations and no secondary particles generation.

The agreement between the simulation results and ESTAR data is within the uncertainty of the reference data for all target materials. Between 100 keV and 1 GeV, the difference between Geant4 simulation results and ESTAR data is <1%, below 100 keV the difference is within 3% for all models except for *Penelope* in Au, which is within a 5% agreement (still within the uncertainty of the reference data $\sigma_{ref}$). Above 1 GeV, the difference increases due to different bremsstrahlung models in Geant4 and ESTAR computations. The results also indicate some interpolation problems in Livermore stopping powers below 100 keV, which is however within the model uncertainty.

## 3.C. Electron backscattering test

Electron backscattering is an important process in several physics applications including medical physics. It occurs when electrons, incident on a target, are scattered in the backward direction. The capability of properly reproducing this complex interaction in Geant4 has considerably improved in

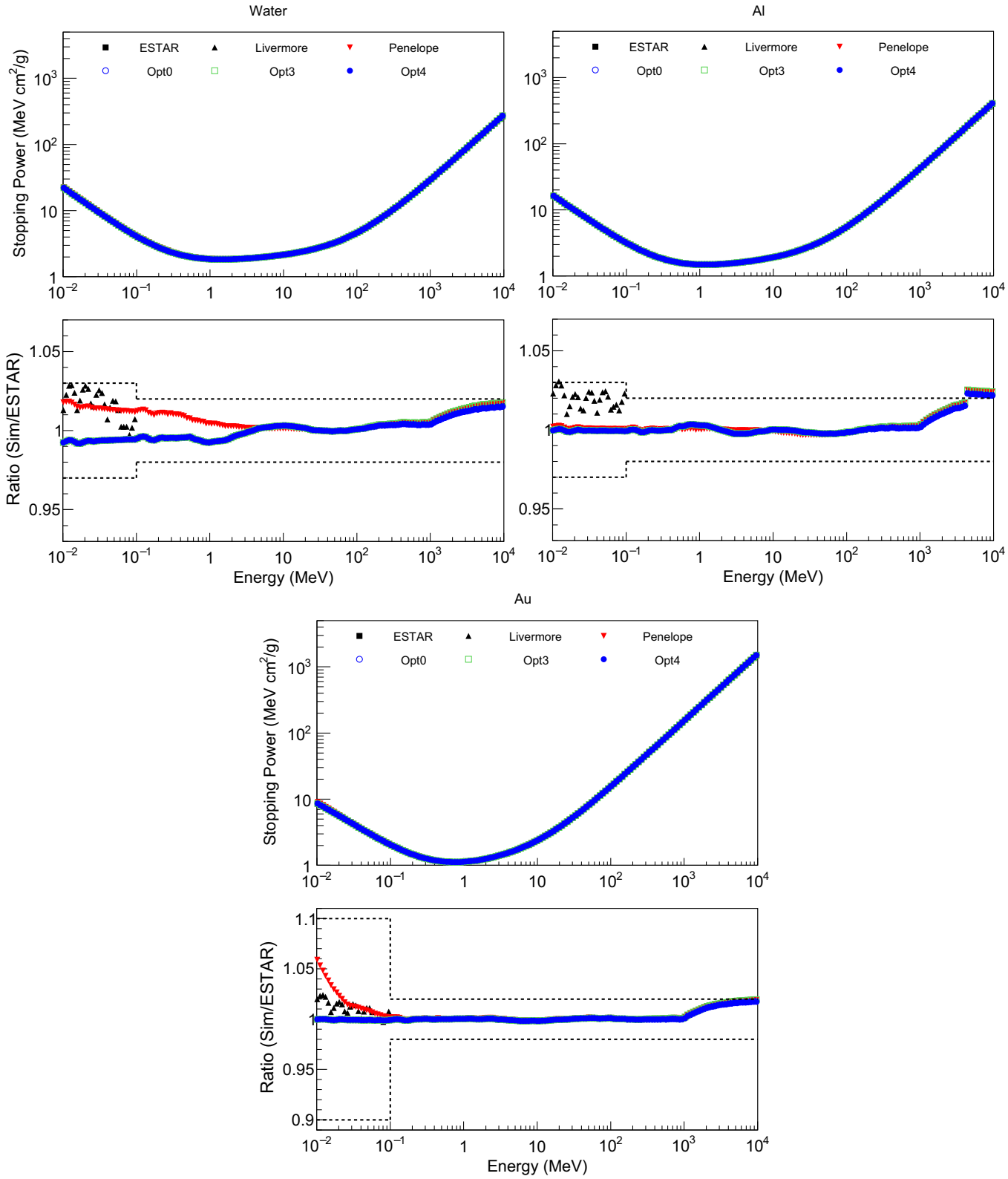


FIG. 3. Electron electronic stopping power calculated by means of different Geant4 EM physics constructors, compared to the ESTAR reference data. The lower plots show the ratio of the Geant4 simulation results and ESTAR data. The dashed lines represent the uncertainty of the reference data $\sigma_{\text{ref}}$. The simulation results do not have any statistical uncertainty because there is no multiple scattering, no energy loss fluctuations and no secondary particles generation.

TABLE V. Electron beam energies, incidence angle and target materials used in the Geant4 simulations.

| Case study | Incidence angle (°) | Electron energies (MeV) | Target materials |
|---|---|---|---|
| Angular dependence | 0, 60 | 1.033, 0.521, 0.314, 0.109, 0.084, 0.058, 0.032 | Al |
| Energy dependence | 0 | From 50 eV to 1.033 MeV | Au, Si |

All the combinations are subject of the test.

the last years, reaching a satisfactory agreement with experimental measurements published in the literature, performed in different periods and using different techniques.[17] This work considers the wide selection of the available experimental data that is documented in Dondero et al,[17] where a detailed description of each experimental technique used in the reported experiments can be found. This selection includes all the datasets known to the authors that respects basic quality criteria, as discussed in the reference article. This test is named *ElectBackScat* in the *geant-val* web interface.[12]

### 3.C.1. Simulation setup

The Geant4 simulation used in this test models a simplified geometry, with an ideal monoenergetic electron beam impinging on a circular metallic target in vacuum. Several energy values, ranging from 50 eV to 1 MeV, and several incidence angles, from 0 to 75°, were simulated for the electron beam. Different target materials have also been tested. The complete list of these parameters is reported in Table 5. All the possible combinations of the reported values have been simulated, in order to reproduce the experimental conditions of the reference data.

The backscattering coefficient, $\eta$, is defined as follows:

$$\eta = \frac{e_{\text{back}}}{e_{\text{tot}}},$$

where $e_{\text{back}}$ is the number of backscattered electrons and $e_{\text{tot}}$ is the total number of incident electrons. For each simulation, the backscattering coefficient was calculated. The backscattered electrons can be either primary or secondary particles. Two case studies are addressed. The first considers the angular dependence of $\eta$, varying the incidence angle and the electron energies. The second case focuses on the energy dependence at low energies, only when the electrons are normally incident on the target.

For the reported simulations, the following Geant4 EM physics constructors have been used: *Opt0*, *Opt3*, *Opt4*, *SS*, *GS*, *Livermore*, and *Penelope*.

All the EM constructors have been used with the default parameters apart from the electron *minimum energy* (see Table 3) that was reduced with respect to the default value (from 100 to 50 eV) to reach the lowest energies available in experimental data. The simulation results have a statistical uncertainty below 1%.

### 3.C.2. Results and discussion

Table 6 reports the references of the experimental data used in this test to benchmark simulation results. Figure 4 shows backscattering coefficients as a function of the electron beam incident energy for an Al target for normal incidence and for 60° incidence angle. The average number of backscattered electrons increases for larger incidence angles. Above 0.2 MeV, there is an agreement within $3\sigma_{\text{ref}}$ between experimental data and simulation results, for all the considered EM physics constructors except *GS*.

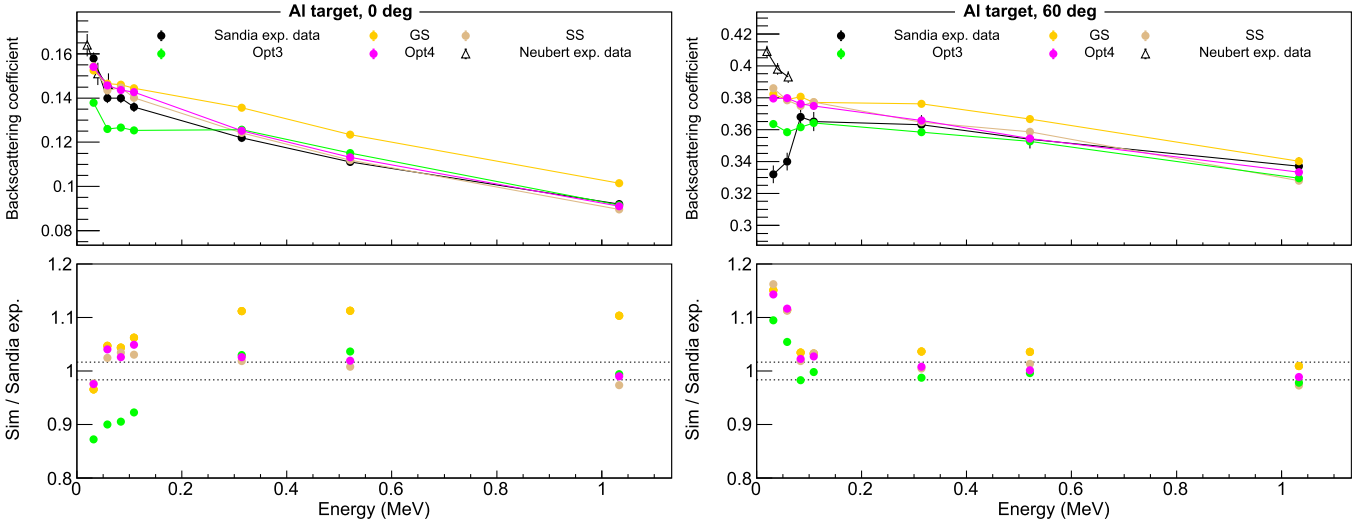


FIG. 4. Backscattering coefficient as a function of the electron beam energies for an aluminum target, in the case when the electron beam is incident normally on the target (left plot) and with an angle of 60° (right plot). Coloured symbols are the simulated coefficient for some of the EM constructors considered. Experimental results from two different research groups are also reported. The lower plots show the ratio between the simulations and Sandia experimental data. The black dashed lines represent 1 $\sigma$ experimental error (1 $\sigma_{\text{ref}}$). The simulation results have a statistical uncertainty below 1%. See Table 6

At lower energies, the considered experimental data show significant differences as reported in Fig. 4, with a dependence on the beam incidence angle. For normal incidence, *Opt4* and *SS* show an agreement within 5% with the Sandia experimental data, while *Opt3* seems to underestimate the low energy backscattered electrons. *GS* is comparable to *Opt4* below 100 keV but then it seems to overestimate $\eta$.

In the case of a 60° incidence, all EM constructors show significant differences (within 15%) below 100 keV. After that energy, there is an agreement between all the EM constructors and Sandia experimental data within 1–2 $\sigma_{\text{ref}}$. It is important to note that the comparison between experiments and simulations can't be quantitatively reliable below 0.2 MeV because of the differences found in the experimental data sets themselves.

The results of energy dependence for silicon and gold targets are reported in Fig. 5. Large discrepancy among the experimental data is shown, due to different experimental conditions and measurement thresholds applied to the scattered electrons. In general, *SS*, *GS*, *Opt4*, and *Livermore* produce results at lower energies with similar trend to the experimental data. In particular for high Z targets, the agreement with some physics constructors results can be rather poor for energies below 10 keV. The calculation of $\eta$ is influenced in the simulations, across the low energy region, by the electron *minimum energy* parameter applied. This parameter is related to the physical cutoff on the electron energy used in the experiments that is observed to have an impact in the lowest energy $\eta$ distribution profile. Usually this cutoff ranges between 50 and 100 eV, due to the use of polarized grids or polarized targets, depending on the particular experimental technique of each experiment. In Dondero et al[17], a more detailed description of the experimental setups used in these electron backscattering measurements is reported. This cutoff starts to influence the electron behavior below 10 keV.

For clarity, Figs. 4 and 5 show the results for a subset of tested Geant4 EM constructors. The results obtained with *Opt0* are very close to the ones generated with *Opt3* as both constructors use the Urban multiple scattering model[38] in the electron energy range under study. *Penelope* and *Livermore* agree very well with *Opt4* as they use the GS model to simulate multiple scattering (see Section 3).

In conclusion, *SS*, *GS*, *Livermore*, *Penelope*, and *Opt4* EM constructors show the best agreement with experimental results. The agreement above 0.2 MeV is within 3 $\sigma_{\text{ref}}$ and between 5 and 0.2 MeV is within 15%. The agreement between simulation and experimental data below about 5 keV is, in general, worse than at higher energies due to differences of up to 40% for certain datasets. As discussed above, in this region the simulation results are particularly sensitive to experiment-related parameters, like the measurement thresholds applied to the detection of scattered electrons. Also for this reason, further investigation is still needed for electron backscattering simulation in Geant4. When this process is important, caution should be used in the meantime, even using the suggested EM physics constructors.

### 3.D. Electron forward scatter from foils at 13 and 20 MeV

The main physical processes in the transport of electrons at clinical energies are bremsstrahlung, collisional energy loss, and scattering. The components of the treatment head of a linear accelerator used for electron therapy are intentionally thin in order to minimize energy loss and the generation of x rays, which contaminate the treatment beam. In this case, accurate simulation of electron scattering is paramount.

An experimental benchmark of the scatter of electrons (13 and 20 MeV) is available that represents the higher range of energies generally available in the clinic (5–25 MeV) for a comprehensive set of scattering materials (Be, C, Al, Ti, Cu, Ta, and Au) for thicknesses that result in a characteristic angle (or root mean square scattering angle) of 2–8° (Ross et al[60]). The test is called *ElecForwScat* in *geant-val*.

The benchmark gives the characteristic angle for each energy, material, and thickness, and the angular distribution

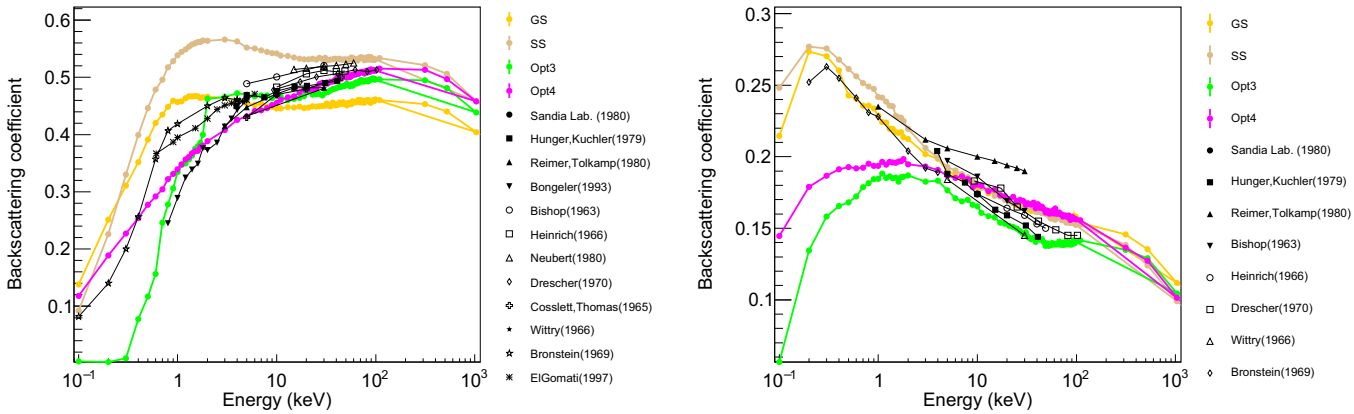


FIG. 5. Comparison among electron backscattering coefficients simulated using various Geant4 EM physics constructors (colored symbols) and a selection of experimental data (black and white symbols), see Table 6 for the complete list of references. Results for gold (right plot) and silicon (left plot) are reported. The simulation results have a statistical uncertainty below 1%.

of fluence out to 0.9–2.5 times the characteristic angle, limited by the lateral extent of the helium bag that was placed between the scattering foil and detector. This helium bag was used to minimize scatter of the electron beam as it passed through this intervening space. The published measurements include a rigorous uncertainty analysis. Previously, the MC systems (in alphabetic order) EGS, Geant, Geant4, and PENELOPE have been benchmarked against these measurements.[18,61] It was found that the characteristic angle alone was insufficient to quantify the discrepancy between the measurement and simulation. Thus, both the characteristic angle and the angular distribution at points near or beyond the characteristic angle are shown. The comparison is limited to a single, representative foil thickness, chosen to illustrate the characteristic angle closest to 5°, and to the more commonly used lower energy of 13 MeV. This is justified since any discrepancy between measurement and calculation is expected to appear at all foil thicknesses and at both tested energies. The comparison for the angular distribution was for representative results from a select set of scattering foils.

#### 3.D.1. Simulation setup

A monoenergetic electron beam of 13 MeV and Gaussian circular spot of 0.1 cm FWHM was normally incident on the exit window, a scattering foil, a monitor chamber, and mylar slabs on either side of a region filled with helium. The foils were 0.926 g/cm$^2$ Be, 0.546 g/cm$^2$ C, 0.14 g/cm$^2$ Al, 0.0910 g/cm$^2$ Ti, 0.0864 g/cm$^2$ Cu, 0.443 g/cm$^2$ Ta, and 0.0312 g/cm$^2$ Au. The scoring plane was perpendicular to the beam axis and located 1.182 m from the exit window. Details of the geometry along with the various scattering foil materials and thicknesses are the same as those used in Faddegon et al.[18] The fluence of electrons was scored in radial spatial bins of 1 mm width. A two parameter Gaussian function was fitted to the angular distribution of fluence to calculate the characteristic angle, limited to the region with the fluence greater than 1/3 of the maximum value, out to a radius of 18 cm on the scoring plane as done previously in Faddegon et al.[18] The physics constructors were *Livermore, Penelope*, *Opt0*, *Opt3* and *Opt4*. A global production cut of 0.01 mm for secondary particles was used.

#### 3.D.2. Results and Discussion

The characteristic angle from the benchmark measurement is compared with that from Geant4 in Fig. 6 for all scattering foil materials. In the top left panel, the simulation with *Opt0* agrees with the measured benchmarks to within 1 standard deviation (which is 1.0%) of the experimental uncertainty for all materials except carbon. Note that this reasonable agreement was not seen in the fluence distributions of all the other panels, where fluence at angles larger than the characteristic angle are shown. All the other constructors significantly underestimated the characteristic angle of most of the foils, by up to 3% for some foils. Results for the characteristic angle calculated with 10.5 version of Geant4 are comparable to those calculated with Geant4 9.2.[18]

The angular distribution beyond the characteristic angle (below for the carbon foil) is shown in Fig. 6 for all the foils. *Opt3* and *Opt4* for Geant4 10.5 show a comparable acceptable match to the measured angular distributions and to those calculated with the other MC codes than past comparisons with Geant4 9.2.[18] In general, *Opt4* systematically underestimates forward-scattered electron fluence by up to 2–5% in the MeV range. The mitigation of these differences remains as an open problem.

### 3.E. Bremsstrahlung from thick targets

The measured yield of bremsstrahlung from electrons normally incident on thick targets at radiotherapy energies as a function of target material, electron energy, and angle provides a key benchmark for the modeling of linear accelerator treatment heads since this requires accurate simulation of both electron scatter and bremsstrahlung within the target.

The benchmark reference dataset are the measured photon fluence per incident electron and differential in energy: (a) along the axis for 10–30 MV x-ray beams from thick targets of aluminum and lead, and (b) from 0 to 90° of 15 MV beams from thick targets of beryllium, aluminum, and lead. Photon fluence per unit energy per incident electron, and total photon fluence, integrated over energy, per incident electron was experimentally determined at 1 m from the target. Bremsstrahlung yield from 0.22 MeV to the incident electron energy was measured on the axis of 10.09, 15.18, 20.28, 25.38, and 30.45 MeV electron beams.[62] In a separate experiment, bremsstrahlung yield down to 0.145 MeV was measured at angles out to 90° for 15.18 MeV electrons.[63] The published measurements include a rigorous uncertainty analysis. Several MC systems have been previously benchmarked against these measurements.[19] This benchmark is available in the *geant-val* web interface[12] with the name *Bremsstrahlung*.

#### 3.E.1. Simulation setup

The sources in the simulation were monoenergetic, normally incident 0.35 cm diameter beams of constant fluence with energies of 10, 15, 20, 25, and 30 MeV. The beam first impinges on a titanium exit window, followed by a silicon transmission current monitor, then a pure target encased in a steel target chamber. Separate simulations were done for the 15.18 MeV beam with and without the target chamber, as the chamber was not included in measurements for angles over 10°. Details of the simulation geometry are described in Faddegon et al.[19] Photon fluence was scored on the surface of several concentric spherical rings in a sphere of 1 m radius centered at the intersection of the beam axis with the upstream face of the target. The rings covered 1° in the polar angle and the full $0° - 360°$ in the zenithal angle. Photons with energies larger than 0.22 MeV for the forward-directed benchmarks and 0.145 MeV for the angular distribution

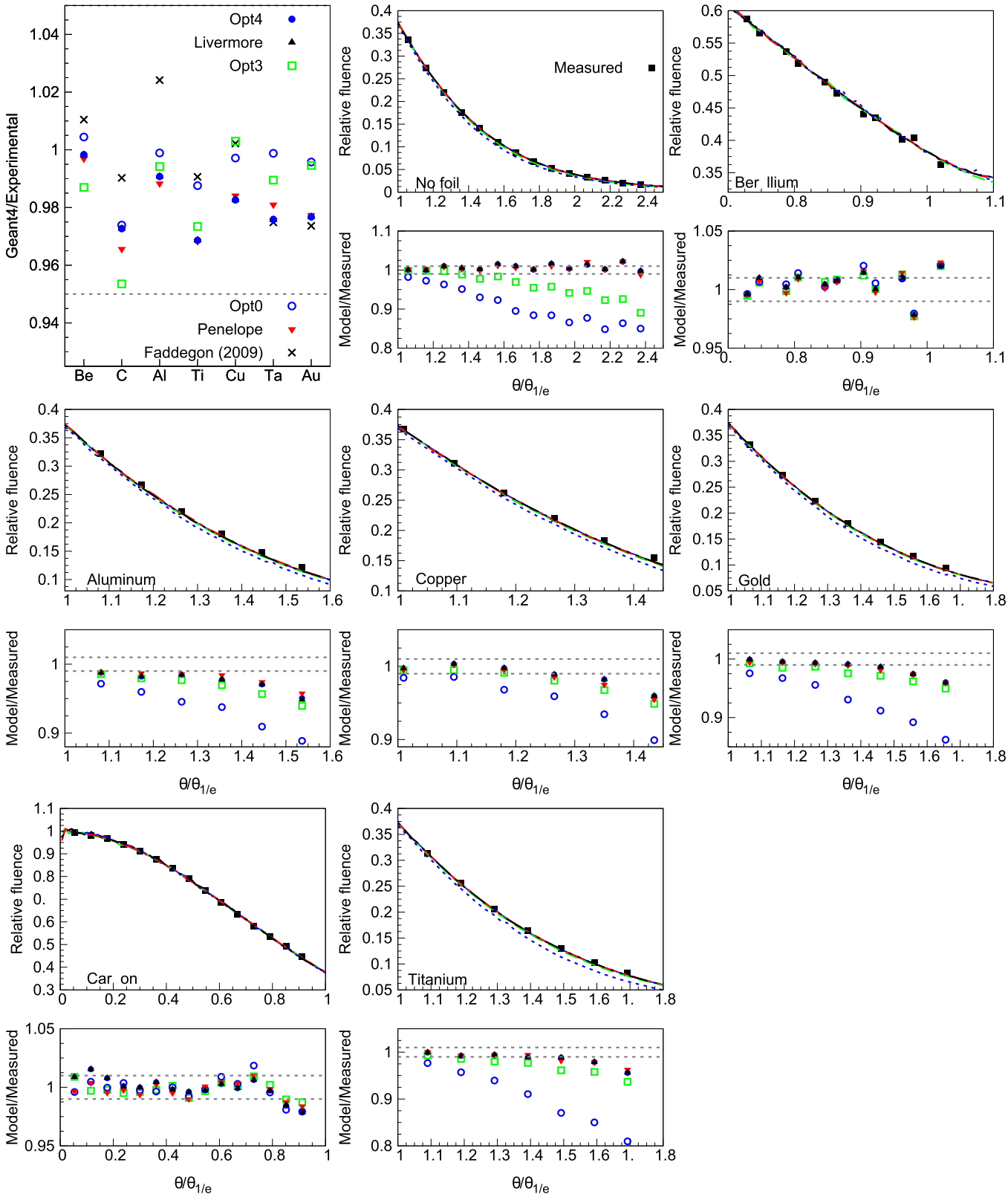


FIG. 6. Ratio of calculated characteristic angle to that measured by Ross et al.[60] for all foil materials (top left), the calculated data from Faddegon et al [18] were obtained with Geant4 9.2. The measured angular distribution normalized to unity on the beam axis for select foils at points near or beyond the characteristic angle $\theta_{1/e}$ are also shown (symbols) compared to calculated results (lines). The gray dashed lines in the plot represent the experimental uncertainties, 1 standard deviation ($\sigma_{\text{ref}}$). The statistical uncertainties of the simulations are within 1%.

benchmarks were scored in 100 log spaced bins and compared with the published experimental benchmarks.[62,63]

The tested physics constructors were *Penelope*, *Livermore*, *Opt0*, *Opt3* and *Opt4*. A global production cut of 0.01 mm for secondary particles was used.

### 3.E.2. Results and discussion

Figure 7 shows selected results of the angular and spectrum distributions for beryllium (top), aluminum (middle), and lead (bottom) for all Geant4 EM physics constructors. These results, obtained with Geant4 10.5, are in better agreement with measurement than Geant4 9.0 patch01, where some yields were well outside 1 standard deviation of experimental uncertainty (see results reported in Faddegon et al[19]). In particular, the results obtained with *Opt3* and *Opt4* agree within 1 standard deviation experimental uncertainties for all energies and all angles below 60° with the exception of the 10 MeV yield for *Opt4* on the beam axis, which is just outside 1 standard deviation. At 90° for the Al and Pb targets and 60° for the Pb target, the simulated results exceed the measurement by 1–2 standard deviations for all options, a larger discrepancy than found previously, but in better agreement with EGSnrc and PENELOPE. The energy spectra are within 1–2 standard deviations as shown in Fig. 7 except at

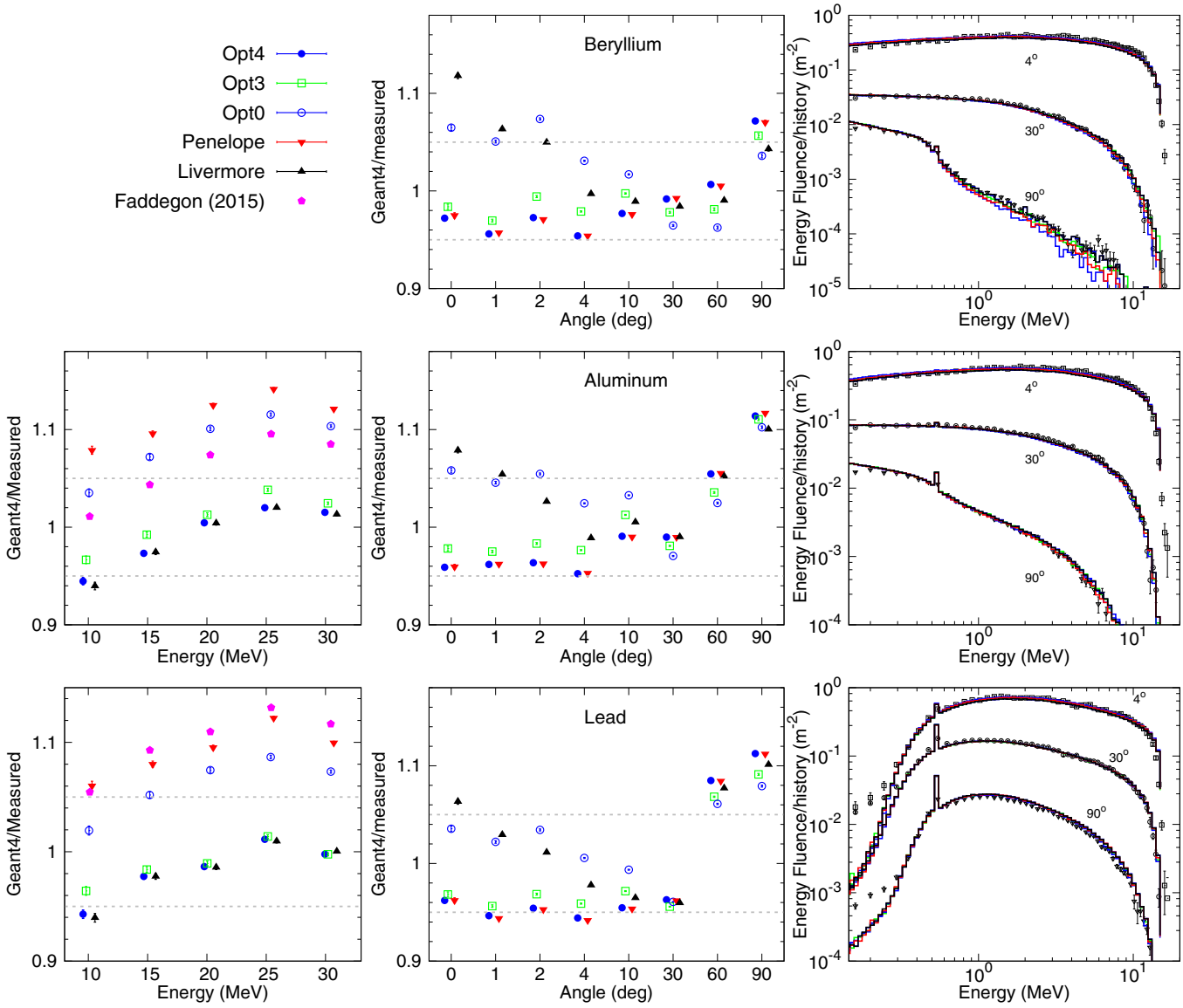


FIG. 7. Bremsstrahlung from thick targets of beryllium (top), aluminum (middle) and lead (bottom). Ratio of simulated to measured 10–30 MeV bremsstrahlung yield on the beam axis (left) for aluminum and lead targets, including published results[19] from an earlier version of Geant4 (left), ratio of simulated to measured 15.18 MeV bremsstrahlung yield from 0-90° (center) and spectral distributions of energy fluence at 15.18 MeV and representative beam angles of 4°, 30° and 90° (right). The measured spectra are shown with markers. Ratios are displayed at slightly shifted energy and angle for clarity. Associated error bars, in most of the cases smaller than the symbols, represent calculated statistical uncertainties, 1 standard deviation. The gray dashed lines in the plot represent the experimental uncertainties, 1 standard deviation ($\sigma_{\text{ref}}$).

low energy fluence where contributions of experimental artifacts to the uncertainty may be underestimated.

## 3.F. Fano Cavity test to verify the multiple scattering and boundary crossing algorithm

This test is released as an *extended* example of Geant4 with the name *FanoCavity*. It is based on Poon et al.[64] and Kawrakow.[65] It was designed to check the accuracy of the condensed history (CH) electron transport, especially the stability of the related stepping algorithms with respect to increasing values of the maximum allowed energy loss along an individual electron step. In case of Geant4, the corresponding continuous step limit parameter is the *dRoverRange* which is the maximum allowed step length in units of fraction of the (charged) particle range. This test is available under the name *FanoCavity* in the *geant-val* web interface.[12]

### 3.F.1. Simulation setup

The different factors involved in the electron transport, that is, the step limitation, effects of energy loss, modeling of multiple Coulomb scattering, are tested using the Fano cavity principle described in Fano.[66]

The model of ionization chamber used is the one described in Poon et al.[64]: a cylinder made of 5 mm water (G4_WATER) walls and a 2 mm cavity filled with steam (G4_WATER with a density of 1.0 mg/cm$^3$). A beam of 1.25 MeV gamma rays parallel to the cylinder axis traverses it. With this setup, under idealized conditions, the ratio of the dose deposited divided by the beam energy fluence must be equal to the mass energy transfer coefficient of the wall material.

The needed equilibrium condition for charged particles is realized using the beam regeneration after each Compton interaction: the scattered photon is reset to its initial state, energy, and direction after the Compton process. Consequently, interactions are uniformly distributed within the wall material.

It is important to mention, that unlike the other tests used in this benchmark, the Fano cavity test requires its special physics modeling conditions. Therefore, the test fully relies on custom, local to the test EM physics constructors with special models for ionization and Compton scattering. Ionization is simulated using a model similar to the standard *G4MollerBhabha* (see the Geant4 Physics Reference Manual for details) with the density dependent correction term of the corresponding stopping power removed. Moreover, in order to have the same stopping power both in the wall and cavity, the bremsstrahlung process is not modeled. The special model for Compton scattering guarantees the conservation of the charged particle fluence by utilising the above mentioned beam regeneration.

To speed up the simulation, it is possible to increase the Compton cross section and the secondary particles that have no chance of reaching the cavity (when the range is smaller than 0.8 times the distance to the cavity) are killed.

To prevent the generation of $\delta$-rays, the global production cut is set to 10 km, in order to be in the continuous slowing down approximation (CSDA). On top of these options, the *finalRange* (see Table 3) of the energy loss is set to 10 μm, which corresponds to a kinetic energy of 20 keV in water.

As it was already mentioned, this test was designed to check the accuracy of the CH electron transport. Therefore, five different local physics lists were used in the benchmark all with exactly the same special description of the physics interactions except the Coulomb scattering. The Coulomb scattering was modeled exactly as in the *Opt0*, *GS*, *Opt3*, *Opt4*, and *WVI* constructors in case of *Opt0**, *GS**, *Opt3**, *Opt4**, and *WVI**, respectively. The * is used in the notations in order to clearly indicate that the physics constructors used in this test are not identical to the corresponding Geant4 EM physics constructors except the modeling of the electron Coulomb scattering and the corresponding stepping algorithm.

### 3.F.2. Results and discussion

Figure 8 shows the dependency of the ratio of the simulated dose and the theoretical one as a function of the *dRoverRange* parameter value for different physics constructors discussed in the previous section. *Opt0* and *GS* constructors utilize the Urban[38] and the GS[37] models, respectively, to simulate multiple Coulomb scattering of electrons (below 100 MeV) with their special settings recommended for high energy physics (HEP) simulation applications. The corresponding settings include looser stepping algorithms since HEP applications, in general, are more tolerant of mistakes in the electron stepping especially when it comes with a significantly increased computing performance. As it is expected, both local physics constructors *Opt*0* and *GS** show a strong dependence on the *dRoverRange*. *Opt0** and *GS** show a

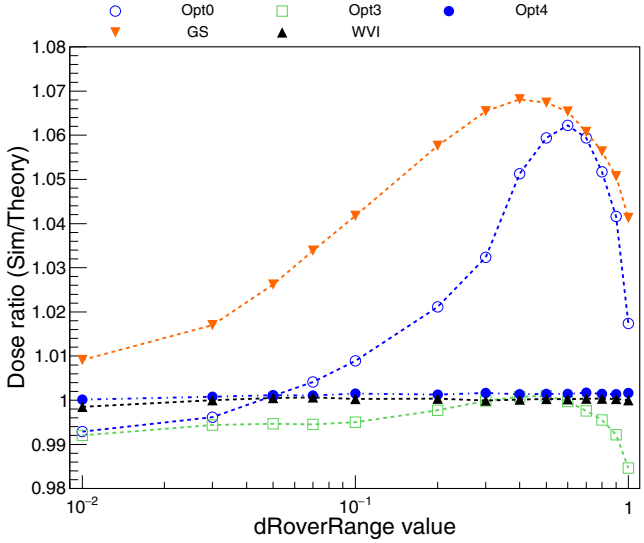


FIG. 8. Dependency of the ratio of the simulated dose and the theoretical one as a function of the *dRoverRange* for different EM physics constructors. The statistical error bars affecting the simulation results are smaller than the symbols.

deviation from the theoretical data values between approximately 1% and 7%, depending on the value of the *dRoverRange*. The same two multiple scattering models are used to describe the Coulomb scattering in *Opt*3* (Urban) and *Opt*4* (Goudsmit–Saunderson) but with their more accurate settings, including the stepping algorithm, that are recommended for precision in critical applications. Accordingly, both multiple scattering models show significantly better stability under these settings. *Opt3** shows an agreement between 1% and 2%, depending on the value of the *dRoverRange*. For the default value of the *dRoverRange* (0.2) (see *dRoverRange* for electrons and positrons in Table 3), the agreement is approximately 0.3%. *Opt*4* provides a remarkably small (<0.2%) deviation from the theoretical dose value practically independently from the *dRoverRange* parameter value. In addition to *Opt*4*, that contains the GS multiple scattering model with its most accurate configuration,[37] the *WVI** constructor, that utilizes the so called mixed-simulation description of Coulomb scattering, also provides a high level of accuracy and stability (<0.2%). These results are in a perfect agreement with the findings described in a recent study,[67] in which a maximum deviation from the theoretical dose values of 0.16% was reported when the GS model with its *Opt*4* settings were used. Therefore, this test shows that Geant4 can transport electrons accurately without the need of applying significant step-size reduction, and irrespective of *dRoverRange* when using the *Opt4* constructor, in the geometrical setup considered.[67] The Fano test subject of this section could be repeated by including photon transport and secondary particle production for completeness. It is recommended to repeat the test in the user's geometry of interest.

## 3.G. Low-energy electron dose point kernels

The simulation of radial energy deposition profiles from isotropic sources of electrons has proven to be a useful method of evaluating the performance of MC codes used in medical physics, in particular for the usage of radionuclides in targeted cancer therapy.[68,69] Although experimental data on electron Dose Point Kernels (DPK) in liquid water — the main component of biological medium — currently do not exist, preventing a full validation, we propose a benchmark comparing Geant4 performance to the EGSnrc Monte Carlo code. This application code can be found in Geant4 as the *TestEm12 extended* electromagnetic example. The test has the name *LowEElectDPK* in the *geant-val* web interface.[12]

### 3.G.1. Simulation setup

In this setup, electrons are emitted isotropically into a $4\pi$ solid angle from a point source placed in a spherical liquid water (defined as G4_WATER) volume. In this work, results are presented for 10, 15, 100, and 1000 keV incident monoenergetic electrons. At each simulation step, the energy deposition is randomly distributed along the step; the radial distribution of energy deposition from the emission point is then recorded in a histogram with a weight equal to the energy deposition value collected in the step. Radius values are scaled to the CSDA range ($r_0$) of the electron at the selected incident energy ($E_0$). The DPK distribution is then normalized to the number of incident electrons, to the selected histogram bin width, and to the incident kinetic energy value. Results are presented for $10^5$ incident electrons. The simulation results have a statistical uncertainty lower than 3%. *Opt0*, *Opt3*, *Opt4*, *Livermore*, and *Penelope* have been adopted to describe the EM physics interactions. In addition, a maximum step size equal to the bin width of the histogram is applied to control the spatial accuracy of the energy deposition.

### 3.G.2. Results and discussion

Figure 9 presents the scaled DPK distributions obtained for each incident electron energy using the Geant4 electromagnetic physics constructors. The curve obtained with *Opt0* is not shown as this constructor produces the same results of *Opt3*. These profiles are compared to EGSnrc predictions (Mainegra-Hing et al,[70] computed as described in the PhD thesis of Perrot[71]). In this comparison with EGSnrc, energy loss fluctuations have not been considered in the simulations. All Geant4 EM physics constructors produce similar profiles as a function of scaled radius and incident energy. *Opt3* shows lower maxima at 10, 15, and 100 keV, and larger profiles below about $(r/r_0) = 0.4$ than the EGSnrc simulation data.

In the case of *Opt4*, *Livermore*, and *Penelope* constructors, one can observe a better agreement (within approximately 3%, corresponding to the statistical uncertaity of the Geant4 simulation results) with EGSnrc simulation data at all incident energies, thanks to the GS multiple scattering model,[37] which has been recently introduced in those three physics constructors (in Geant4 version 10.5, as the best alternative to the *G4UrbanMscModel*). This agreement is obtained under the condition that energy loss fluctuations are ignored in Geant4. Indeed, EGSnrc simulation data do not take into account these fluctuations. Such agreement is not observed when energy loss fluctuations are taken into account in the Geant4 simulations (data not shown). This underlines that, under this condition, the GS model used in the three above constructors is able to perform as EGSnrc and can provide more accurate simulation data. It remains, however, impossible to fully validate such simulations in the absence of experimental data in liquid water in this energy range. At higher incident electron energies, effects of energy loss fluctuations are not important and Geant4 *Opt4* with GS model and EGSnrc results agree well, so it is possible to conclude that both simulation tools (on the one hand Geant4 *Opt4* constructor with GS model, and EGSnrc on the other hand) can be regarded as a gold standard for such simulations.

We show in this study the full benefit of the newly implemented GS model for the simulation of electron multiple scattering in liquid water, when compared to EGSnrc Monte Carlo code.

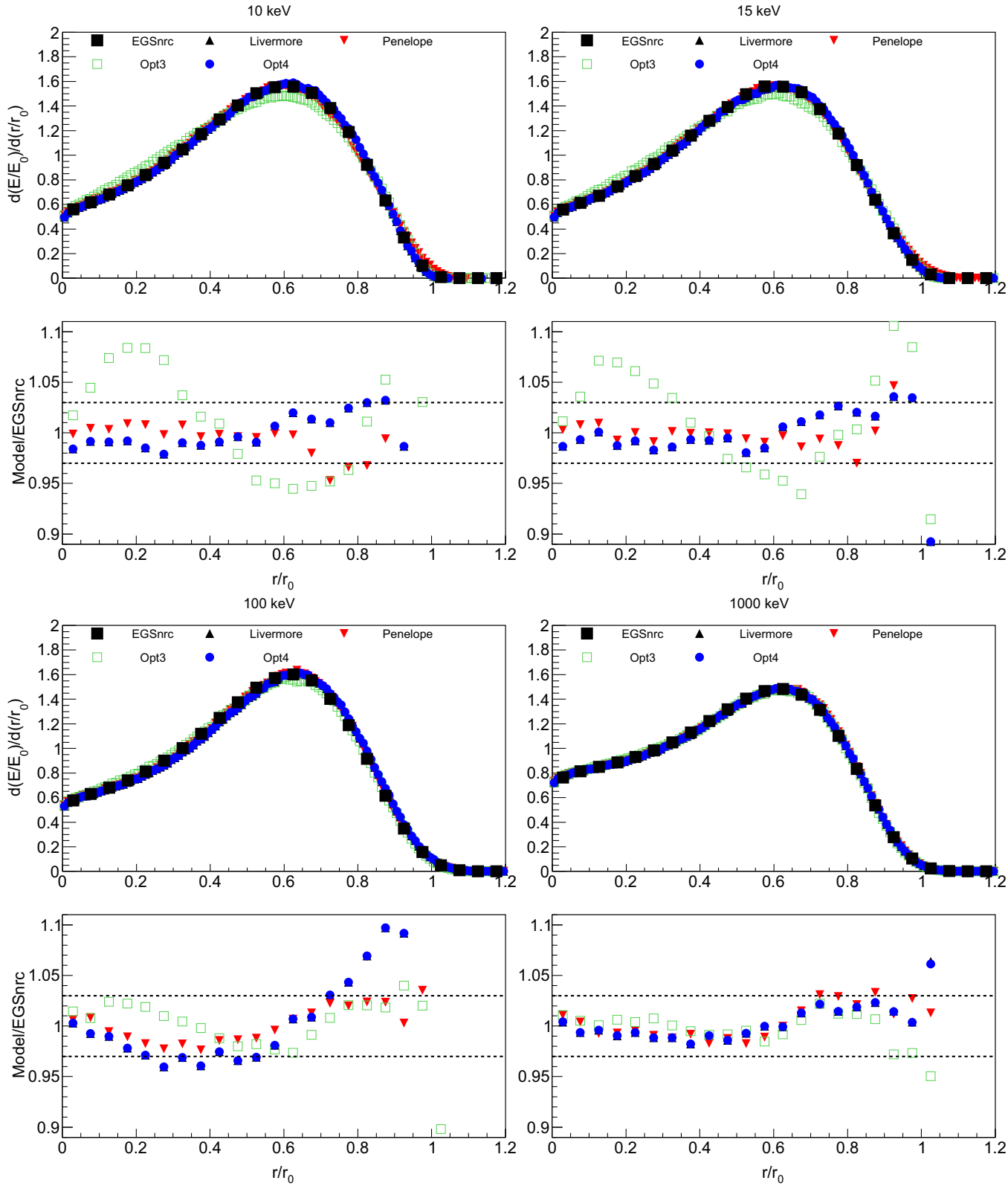


FIG. 9. Scaled DPK distributions for incident electrons of 10, 15, 100, and 1000 keV in liquid water, simulated using the five Geant4 EM physics constructors. Ratios of Geant4 and EGSnrc results are reported as well. Dashed lines indicate an agreement of 3%, which corresponds to the statistical uncertainty affecting the Geant4 Monte Carlo simulation results.

## 3.H. Microdosimetry test

### 3.H.1. Introduction

Microdosimetric spectra of lineal energy (or specific energy) are historically being used for describing radiation quality (see ICRU Reports 40[72] and 36[73]), and many theoretical estimates of the relative biological effectiveness (RBE) of therapeutic beams are based on such calculations.[74,75]

The use of the MC method for the calculation of stochastic energy deposition in an irradiated volume is facilitated by the use of condensed-history models in order to reduce simulation time, especially when we are dealing with radiotherapeutic energies. A dedicated example available in the Geant4 *extended* examples, called *microyz*, offers a way to calculate microdosimetric spectra in liquid water target spheres.

Systematic studies of the microdosimetric performance of the low-energy condensed-history models available in Geant4 (i.e., *Livermore* and *Penelope*) and the track structure models available in Geant4-DNA[76,77-79] have been recently carried out for submicron volumes for the energy range 50 eV—10 keV.[20,80,81] Here we investigate the microdosimetric performance of the *Opt4* constructor at energies of radiotherapeutic interest (10 keV–1 MeV) in terms of calculating the frequency-mean lineal energy in liquid water spheres with a diameter of 1 μm (ICRU sphere). This test has been included in the *G4-Med* benchmarking system for regression testing purposes only. The test has the name *microyz* in the *geant-val* web interface.[12]

### 3.H.2. Simulation setup

The extended example *microyz* was used to calculate the probability density function of lineal energy by scoring the energy deposited by monoenergetic electrons within target spheres of liquid water (1 μm diameter) randomly overlapping their track. The procedure of scoring the energy deposition is provided in detail in Kyriakou et al.[20] For all simulations, the global production cut was set equal to the tracking cut.

The effect of the step-size limit and tracking cut is also investigated. To obtain a statistical uncertainty below 1–2%, the number of electron tracks simulated was $10^6$ for electron energies up to 100 keV and $10^5$above 100 keV.

### 3.H.3. Results and discussion

Figure 10 presents the frequency–mean lineal energy as a function of incident energy for 10, 50, 100, 500, and 1000 keV electrons using two different values for the tracking cut, 100 eV and the commonly used value in condensed-history simulations of 1 keV.

Results are presented for three different step-size limits (called SL in the Figure), the default step-size limit which for this test is equal to 1 mm, a step-size limit equal to the sphere diameter (1 μm), and a small step-size limit equal to 1/10 the sphere diameter (100 nm).

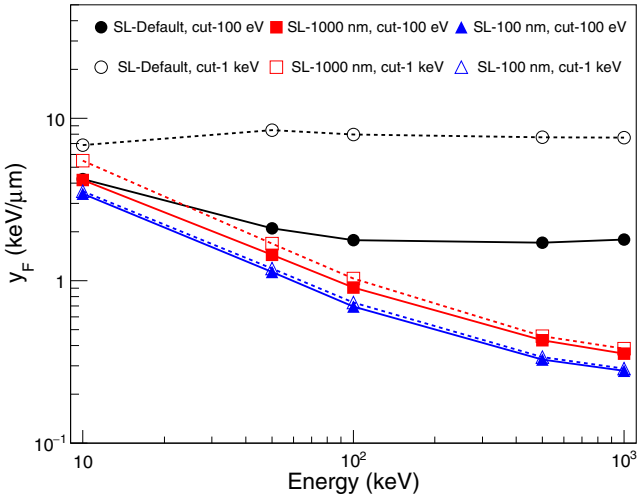


FIG. 10. Frequency–mean lineal energy as a function of the incident electron energy, using two different values for the tracking cut 100 eV and 1 keV. SL is the step-size limit. SL was set equal to 1 mm (default value for this test), 1 μm and 100 nm. Error bars are not included since they are smaller than the symbols. The plot demonstrates that SL should be substantially less than the size of scoring volume.

We observe that for both values of the tracking cut (1 keV and 100 eV), the default step-size limit (equal to 1 mm in the present energy range) results in a significant overestimation that increases with electron energy (at 1 MeV incident electron energy it reaches a factor of 6 for the 100 eV cut, and a factor of 26 for the 1 keV cut). A step-size limit equal to the sphere diameter (1 μm) also results in sizeable overestimation (20–30% for 100 eV cut and 30–50% for 1 keV cut) compared to the small step-size limit. The influence of the tracking cut is reduced when smaller step-size limits are chosen. For example, with the small step-size limit examined (100 nm), decreasing the tracking cut from 1 keV to 100 eV affects the frequency–mean lineal energy by <7%.

The test *microyz* assumes by design discrete simulation, that is, each individual interaction leading to any modification of the particle trajectory (energy, direction) is modeled explicitly as point-like interaction and is recorded during the simulation. This assumption is broken when employing the CH simulation technique which leads to the observed sensitivity to the CH step size. The extent of this artificial step size dependence is determined by the relation of the step size distribution to the size of the target sphere. Reducing this step size dependence requires that several steps are done by electrons inside the scoring sphere which may be achieved by defining a step-size limit less than the scoring sphere dimensions. For the commonly used tracking cuts studied in the present work (100 eV to 1 keV), a safe value for the step-size limit is equal to the 1/10 of sphere diameter.

## 3.I. Brachytherapy test

The *Brachytherapy* Geant4 advanced example is used to calculate the dose rate distribution of a high dose rate

brachytherapy $^{192}$Ivr source in water. We compared the results of the simulation to the reference data published in Granero et al,[82] which were obtained with Geant4 version 7.1. Therefore, the test has a regression testing focus and is available in the *geant-val* interface[12] with the name *Brachy-Ir*.

### 3.I.1. Simulation setup

The $^{192}$Ir Flexisource, described in Granero et al,[82] has been modeled in the center of a water box (modeled as G4_WATER) with 30 cm size. The photon radiation field, detailed in Granero et al,[82] is emitted from the iridium radioactive core. *Opt0*, *Opt3*, *Opt4*, *Livermore*, and *Penelope* have been tested. A global production cut was fixed equal to 0.05 mm. A Geant4 *scoring mesh* was defined to calculate the energy deposition in the plane containing the source. The plane is subdivided in voxels with size equal to 0.25 mm along x, y and z.$10^8$histories were simulated to obtain a statistical uncertainty of 1.5% in the results. The same statistical uncertainty affects the reference data.[82]

### 3.I.2. Results and discussion

Figure 11 shows the radial dose rate distribution about an $^{192}$Ir brachytherapy source, as a measure of the energy deposition per unit of mass, along the transverse axis of the source, 90° from the source axis.

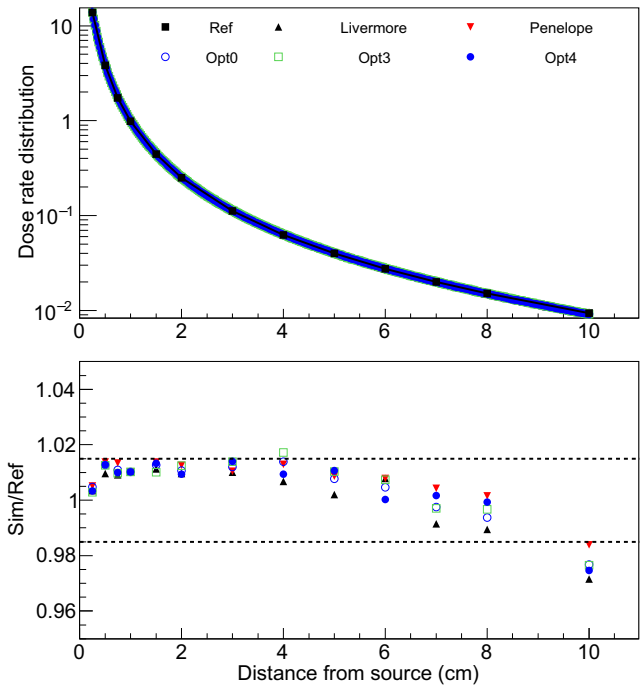


FIG. 11. Radial dose rate distribution with respect to *r*, distance from the center of the $^{192}$Ir brachytherapy source. The radial dose distribution is normalized at $r = 1$ cm. Black squares in the top figure represent the reference data (from Granero et al82), the statistical uncertainties of the data are within the symbols. The bottom plot shows the ratio of the Geant4 simulation results and the reference data. The dashed lines indicate the statistical uncertainty affecting both the simulation results obtained with Geant4 10.5 and the reference data published in Granero et al,82 obtained with Geant4 7.1.

As it can be observed in Fig. 11, the simulation results obtained with any Geant4 EM constructor agree with the reference within the uncertainty of the simulation data $\sigma_{ref}$ (1.5%), for almost all points, up to a distance of 8 cm from the radioactive core. At 10 cm, depth the agreement between Geant4 results and the reference data is within 3% (corresponding to 2 $\sigma_{ref}$) for all the EM constructors. No recommendation to use any of the EM constructors can be done at this stage with this test, because the reference data are derived from Geant4 as well (Geant4 7.1).

## 3.J. Monoenergetic x-ray internal breast dosimetry test

An accurate and controlled evaluation of the radiation dose delivered during x-ray-based breast imaging is part of quality control procedures[83] and necessary for risk estimation (Dance et al[84]). This simulation aims to compare the radiation dose predicted by means of Geant4 and experimental measurements performed at the SYRMEP beamline of the ELETTRA synchrotron light source (Trieste, Italy). Dose measurements were performed using thermoluminescent dosimeters, TLD-100H (ThermoFisher Scientific, Waltham, MA, USA), in absolute terms and down to the local deposition level, in a mammographic acquisition setup installed at the SYRMEP beamline.[85] The test has the name *Mammo* in the *geant-val* interface.[12] Details about the test and the experimental procedure can be found in Fedon et al.[21,22]

### 3.J.1. Simulation setup

The geometry implemented in the simulation is shown in Fig. [12]. A homogeneous semi-cylindrical breast phantom (with the dimensions of 18 cm × 10 cm) consisting of four 1-cm thick slabs was positioned within the plates of the compression system. The phantom (CIRS Inc.,Norfolk, VA, USA) reproduces the x-ray attenuation property of a breast with a 50% adipose and 50% fibroglandular tissue composition. Monoenergetic, parallel 20 keV photons are emitted from a rectangular, planar x-ray source (20 cm × 12 cm) toward the breast phantom. This source geometry models the monoenergetic nature of the SYRMEP beamline of the ELETTRA synchrotron when the double-crystal Si monochromator is set to 20 keV. The scored physical quantity in the simulation was the dose deposited in thirty sensitive volumes (i.e., each TLD) positioned on the *xy* phantom plane at four different phantom depths (i.e., the scoring planes shown in Fig. [12]). The TLDs were modeled in terms of physical dimensions (i.e., 3.2 mm × 3.2 mm × 0.38 mm) and relative chemical composition ( 99.5% LiF, 0.2% Mg, 0.004% Cu, and 0.296% P) according to the manufacturer. For each simulation, $2 \times 10^9$ photons were simulated to obtain a statistical uncertainty lower than 1% in all the scored quantities, estimated using the method of Sempau et al.[86] The experimental TLD procedure and the uncertainty analysis is

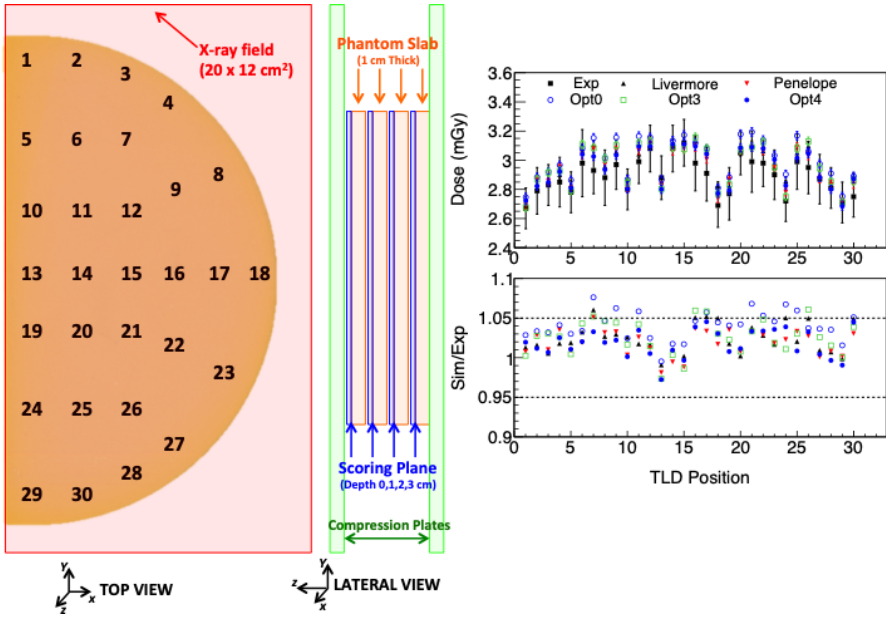


FIG. 12. Left and center: Dosimeter placement and irradiation geometry modeled in the test. Drawings are not to scale. Right: Dose comparison between experimental data (black symbols) and the five Geant4 EM physics constructors for the depth of 3 cm. All the other depths are available in the *geant-val* interface. The black error bars represent the combined standard experimental uncertainty (with a coverage factor, k = 1), while the color ones are referred to the simulations data. The bottom plot shows the ratio between Geant4 and experimental data. The black dashed lines represent the 5% accuracy, a typical threshold accuracy considered in breast dosimetry.

described in detail in Fedon et al.[21] To normalize the photon fluence in the simulation to that used in the experiments, a scale factor was used, defined as the ratio between the experimentally kinetic energy released per unit mass in air (kerma) and the simulated air kerma, analytically evaluated in the MCsimulation. The constructors *Opt0*, *Opt3*, *Opt4*, *Livermore*, and *Penelope* were tested. A global production cut of 0.7 mm was adopted.

### 3.J.2. Results and discussion

Dose comparison results for the depth of 3 cm are presented on the right side of Fig. 12. Data for the other depths can be found on the *geant-val* web interface.[12] In general, all physics constructors produce results that agree with each other within the combined experimental uncertainty. The mean combined uncertainty for the depth of 3 cm is 5.7% (uncertainty range from 5% to 7.7%).[21] It should be pointed out that a systematic bias appears to be present, since the MC data results consistently overestimated the experimental data. This overestimation is, however, within the combined experimental uncertainty. Among the *Livermore*, *Penelope*, *Opt3*, and *Opt4* constructors, no clear trend is observed that can suggest the use of one over the others, since the results are within an accuracy of 5% with only few exceptions for *Opt3* (ratio plot in Fig. 12). However, we noticed that the performance of *Opt0* worsens with increasing depth, confirming previously observed trend.[87] A maximum relative difference of 7.7% is observed at 3 cm depth when using this constructor (position 7 in Fig. 12), whereas with the other EM constructors there is a better agreement with a largest difference of 6%. An accuracy of 5% might be considered large for a MC benchmark. However, in the case of breast dosimetry, this is a remarkable results considering the difficulties and inherent uncertainty in these types of measurements, where the recommended uncertainty range ±12.5%.[21] Thus, these results, within the experimental uncertainty, can be considered a valid benchmark for this specific application.

## 4. HADRONIC PHYSICS BENCHMARKING TESTS

Hadronic interactions are paramount in hadron therapy. In proton therapy, incident protons generate a secondary neutron field which scatters in the patient producing recoil protons, which will then deposit energy in the patient outside the target tumor and potientally in organs at risk.[88] In addition, it recently became evident that the few fragments generated in proton–tissue interactions cannot be neglected[89] because they can affect the average linear energy transfer (LET) values, thus changing the biological outcome. Such fragments are mainly produced at the entrance in the patient where the energy of the protons is higher .[89]

In carbon ion therapy, a beam with an initial energy of 400 MeV/u, will have approximately 70% of the initial carbon ion beam undergo fragmentation before reaching the tumor target. Such fragments contribute to the dose in-field and are responsible for the dose delivered out-of-field, that is, laterally to the beam and beyond the Bragg peak.[90] The total contribution to the dose is approximately 30% for a carbon ion spread out Bragg peak with a maximum energy of 290 MeV/u. It is therefore crucial to benchmark the hadronic physics component of Geant4, which is often used as MC

TABLE VI. References for the experimental datasets used in the electron backscattering test with the energy range and the target material.

| Dataset name | Reference | Material | Energy range |
|---|---|---|---|
| Sandia Lab. (1980) | 46,47 | Al, Au, Si | 0.032 MeV–1.033 MeV |
| Reimer, Tollkamp (1980) | 48 | Au, Si | 1 keV - 30 keV |
| Hunger, Kuchler (1979) | 49 | Au, Si | 4–40 keV |
| Cosslett, Thomas (1965) | 50 | Si | 5–45 keV |
| Bongeler (1993) | 52 | Si | 1–10 keV |
| Bishop (1963) | 53 | Au, Si | 5–30 keV |
| Heinrich (1966) | 54 | Au, Si | 30 keV |
| Neubert (1980) | 55 | Al, Si | 15-60 keV |
| Drescher (1970) | 56 | Au, Si | 10–100 keV |
| Wittry (1966) | 57 | Au, Si | 5, 30 keV |
| Bronstein (1969) | 58 | Au, Si | 0.1–4 keV |
| El Gomati (1997) | 59 | Si | 0.6–6 keV |

The first column reports the dataset names used in Figs. 4 and 5.

code for hadron therapy. An accurate description of the same physical interactions is important for radiation protection studies as well.

This section describes tests in *G4-Med* where only hadronic physics processes and models are tested and the electromagnetic physics is not considered in the simulation (Section 5 is dedicated instead to tests which activate both the electromagnetic and the hadronic physics). The tests described in both sections (here and Section 5) were performed in the energy range of interest for hadron therapy.

The test of section 4.A benchmarks the total hadronic inelastic cross section, which is described with the Glauber–-Gribov model in Geant4 (see the Geant4 Physics Reference Manual) against experimental measurements.

The tests subject of Sections 4.B. and 4.C benchmark different hadronic ion inelastic scattering constructors, *G4IonBinaryCascade*, *G4IonQMDPhysics* and *G4IonINCLXXPhysics*, to describe the final state of carbon ion hadronic inelastic interactions. Table 7 summarizes the main features of the three constructors modeling ion fragmentation under study. It is important to note that the constructors adopt the same total inelastic cross section based on the Glauber–Gribov model, while providing different descriptions of the final state of the interactions (yield, energy, and angular distributions of the secondary particles). The three constructors handle the interactions of deuteron, triton, alpha particle, $^{3}$He, and heavier nuclei.

*G4IonBinaryCascade* (*BIC*) activates the *LightIon BinaryCascade*, which describes the interaction between the projectile and each participating single nucleon of the target nucleus described as a Gaussian wave packet .[91] *G4IonQMD Physics* activates the Quantum Molecular Dynamics (*QMD*) model. In this case, all nucleons of the target and projectile have their own wave function.[92]

*G4IonINCLXXPhysics* uses the approach of the Liège intranuclear-cascade model called here *INCL*.[93,94] The target nucleons are treated as a free Fermi gas in a static potential well, whereas the projectile is modeled without Fermi motion. As result of this asymmetric treatment, the projectiles which can be modeled with *INCL* are limited to mass numbers less than A = 19; otherwise, when the target mass is below A = 19, target and projectile are interchanged internally and when both mass numbers are above A = 19, the fragmentation interaction is modeled with *BIC*.

In the high energy range, all the ion hadronic inelastic scattering constructors adopt the Fritiof parton string model (*FTF*).[95] In the energy overlap region (see Table 7), an interpolation between the two models is done.

## 4.A. Test of nucleus–nucleus hadronic inelastic scattering cross sections

This test calculates the total cross section of hadron–nucleus and nucleus–nucleus collisions. The cross sections are then compared to reference experimental measurements publicly available in the Experimental Nuclear Reaction Data (EXFOR) database.[96] EXFOR provides libraries containing an extensive compilation of experimental nuclear reaction data.

The total inelastic scattering hadronic cross sections are calculated for incident protons and carbon nuclei. The test was named *NucNucInelXS* in the *geant-val* web interface.[12]

### 4.A.1. Simulation setup

The benchmark test retrieves the total hadron–nucleus and nucleus–nucleus inelastic cross sections, which are stored in a data table via the use of the class *G4HadronicProcessStore* in the initialization phase of the simulation. The data table is then compared to the reference data. Only the Geant4 prebuilt physics list *QGSP_BIC* was benchmarked in this test as the total inelastic hadronic cross section in all the hadronic physics constructors, used in any prebuilt Geant4 physics list, are

TABLE VII. Physics models activated in the ion hadronic inelastic scattering constructors under investigation.

| Constructor | Physics models | | |
|---|---|---|---|
| *G4IonBinaryCascade* | *G4BinaryLightIonReaction** for $E < 4$ GeV/u | | *FTF* for $E > 2$ GeV/u |
| *G4IonQMDPhysics* | *G4BinaryLightIonReaction** for $E < 110$ MeV/u | *QMD* for 100 MeV/u $< E < 10$ GeV/u | *FTF* for $E > 1$ GeV/u |
| *G4IonINCLXXPhysics* | *INCL* for $E < 3$ GeV/u | | *FTF* for $E > 2.9$ GeV/u |

*: the *G4BinaryLightIonReaction* activates the Geant4 Precompound Model (*G4PreCompoundModel*, Geant4 Physics Reference Manual).

based on the Glauber representation with the Gribov screening correction on inelastic screening (GG model).[97–99]

The total inelastic scattering cross sections of $p + {}^{12}_{6}C$, $p + {}^{16}_{8}O$, $p + {}^{27}_{13}Al$, $p + {}^{40}_{20}Ca$, ${}^{12}_{6}C + {}^{12}_{6}C$, and ${}^{12}_{6}C + {}^{27}_{13}Al$ are compared with the experimental data available in EXFOR. Such reactions were selected because of relevance for hadron therapy and because EXFOR provided an adequate number of reference experimental data.

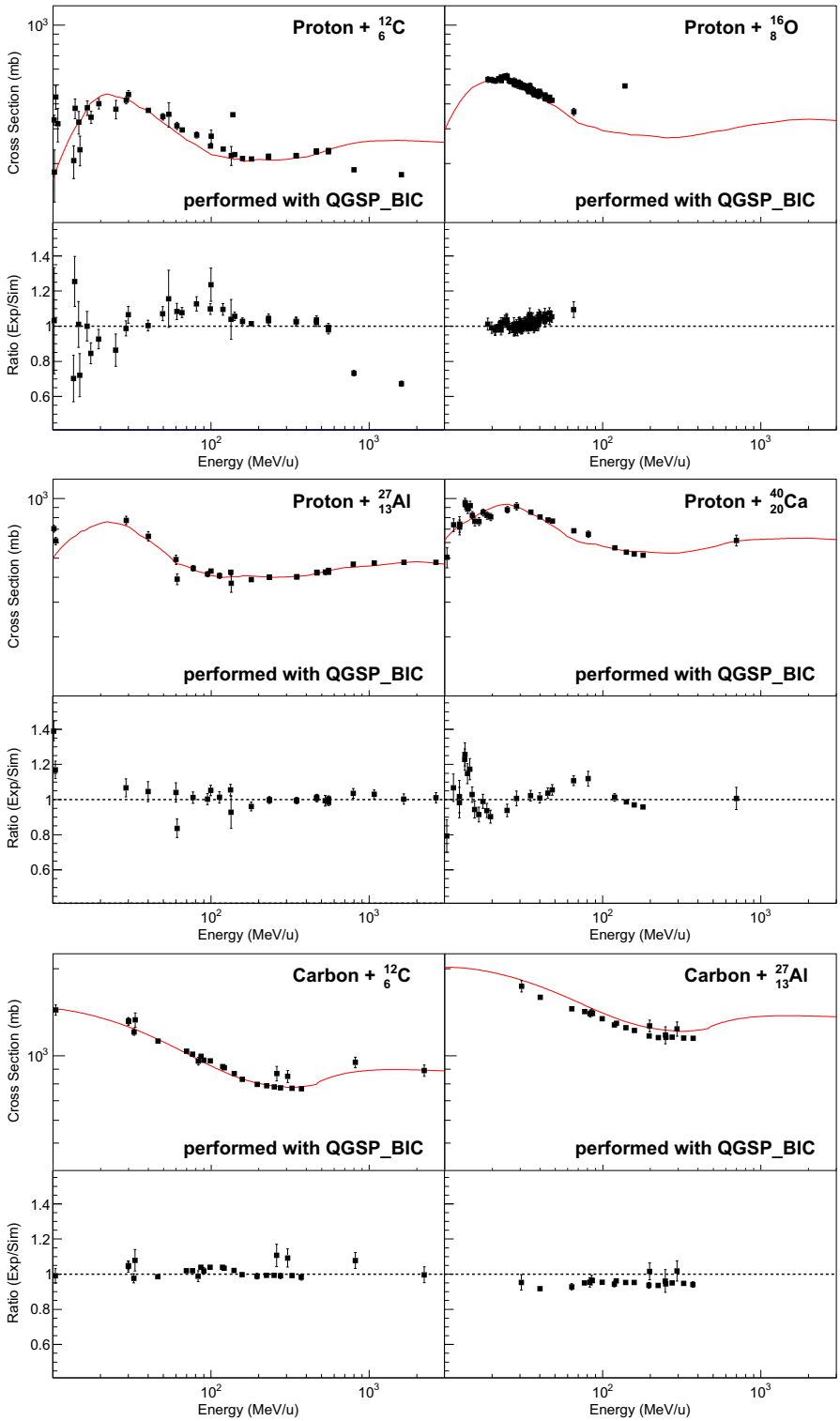


FIG. 13. Total hadronic inelastic scattering cross sections as a function of the kinetic energy of the projectile, calculated by means of the *QGSP_BIC* physics list. To note, differently from the other figures depicting the results of the tests of *G4-Med*, in the bottom plot ratios of reference data and Geant4 simulation results (and not the other way round) have been plotted for clarity reasons. Red curve: Geant4 cross section; data points: EXFOR reference experimental data.[96]

#### *4.A.2. Results and discussion*

Figure 13 shows the total inelastic cross sections calculated by means of the *QGSP_BIC*, as a function of the kinetic energy of the projectile, compared to the experimental data of the EXFOR database. Overall, for both incident protons and carbon ions, an agreement within 10% was observed for most of the incident projectile energies. Significant differences (20%–40%) were found for proton energies below 20 MeV in the case of a carbon target. These differences are caused by the large variations in the experimental inelastic cross sections for p+$^{12}$C, likely due to resonances that enhance the cross section, particularly near 10.5 MeV proton energy (e.g., Dyer et al,[100] Davids et al,[101]). These variations will tend to arise for low energy proton induced reactions, particularly for light targets where the density of states is low. In the case of $^{12}$C+$^{27}$Al, the data were consistently overestimated of about 10%. This reaction should be further investigated.

### 4.B. 62 MeV/u $^{12}$C fragmentation test

This test concerns the nuclear interaction models available in Geant4 for low-energy carbon ions. The name of the test is *LowEC12Frag* in the *geant-val* web interface.[12] The models were benchmarked against double-differential cross sections of the secondary fragments produced in the $^{12}$C fragmentation at 62 MeV/u on a thin carbon target. This dataset has been acquired from De Napoli et al.[102]

#### *4.B.1. Simulation setup*

As described before, the two models available in Geant4 at low energy are the *Binary Light Ions Cascade* (*BIC*) [91] and the *Liege Intranuclear Cascade* (*INCL*).[93,94] There is a third model available in Geant4 to simulate the first part of a nuclear interaction, namely, the *Quantum Molecular Dynamics* (*QMD*).[92] However, *QMD* has, in its default physics constructor class (*G4IonQMDPhysics*), a lower energy threshold at 100 MeV/u. Below this threshold it uses *BIC*. Therefore,

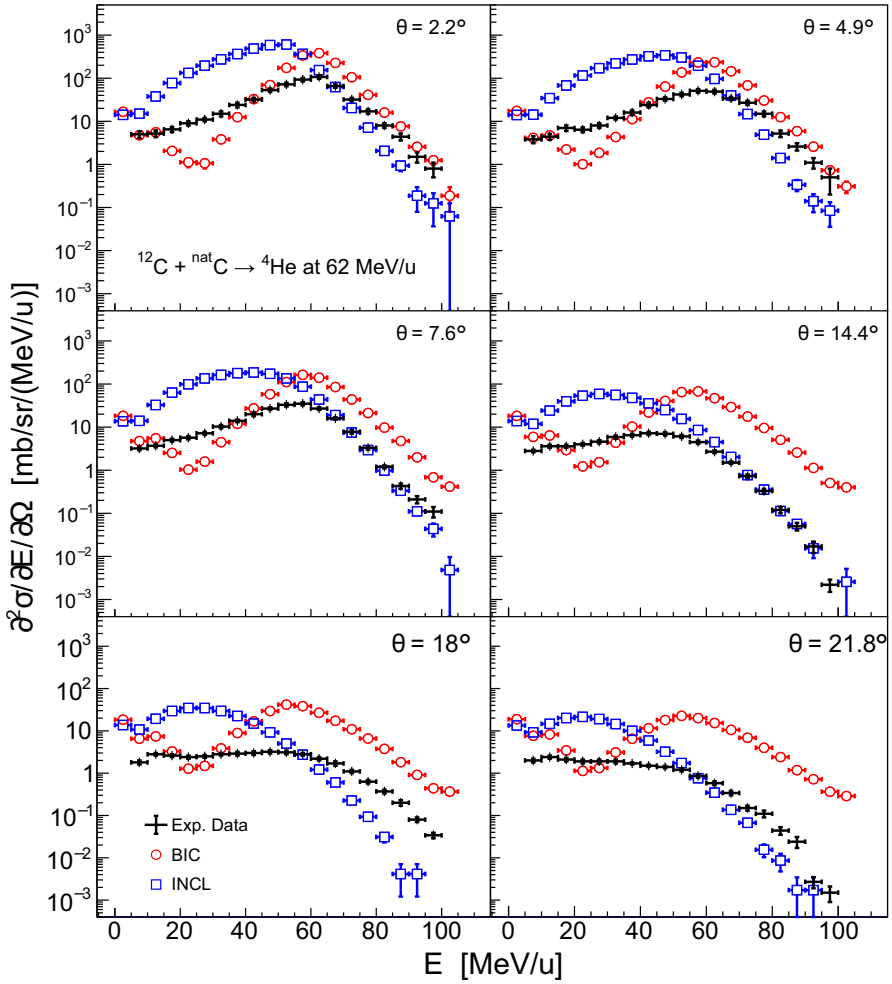


FIG. 14. Double-differential cross sections measured at different angles of $\alpha$-particle production in the $^{12}$C interaction with a thin C target at 62 MeV/u as a function of the $\alpha$-particle kinetic energy. Experimental data from De Napoli et al[102] (in black) are compared with the two models available in Geant4 for ion interactions at this energy, namely *INCL* (in blue) and *BIC* (in red).

in this test, we decided not to benchmark the *QMD* model. This selfsame benchmark with *QMD* included, performed with Geant4 10.4, can be found in Mancini et al.[23]

In order to reduce the computation time, all interactions but the hadronic inelastic ones have been switched off. All secondaries are discarded after being produced and the target is much longer than the hadronic inelastic interaction length. In this way, all primaries undergo an inelastic interaction, which is also the only process simulated. Secondaries are saved immediately after being produced and the event interrupted. Data are selected to match the geometrical acceptance and the energy resolution of the experiment. The obtained spectra are scaled by the total inelastic cross section.

#### 4.B.2. Results and discussion

Figure 14 shows the comparison of the double-differential cross sections of $\alpha$-particles production at different emission angles. *BIC* shows a doubly peaked structure. The one at higher energy is due to the fragment produced in the statistical de-excitation of the projectile remnant, while the lower energy one comes from the de-excitation of the target. *BIC* underestimates the formation of fragments in the mid-rapidity region, here at roughly 30 MeV/u of kinetic energy of the emitted fragment, by up to one order of magnitude at small angles. Its minimum is approximately two decades smaller than the two neighbors regions at all angles, while the projectile-like and target-like fragments formation is overestimated for all angles. A possible explanation could be that *BIC* approach is based on a time-invariant optical potential and this leads to an underestimation of the formation of a neck fragmentation events[103,104] in the overlapping region, roughly at mid-rapidity in semicentral reactions. On the contrary, *INCL* shows a single distribution peaked at lower energy, with respect to the experimental peak. The *INCL* predictions overestimate by around a factor ten the production of $\alpha$-particles below 62 MeV/u but are compatible with the experimental data for particles produced with higher kinetic energy. This could be due to its complete-fusion model.

Here we show only the plot with the comparison of the $\alpha$-particle production. All other plots, with the different fragments, can be found at the *geant-val* web interface.[12]

In conclusion, the results show that both of the tested models, *BIC* and *INCL*, exhibit limitations in reproducing doubly differential cross sections of fragments produced in the interaction of $^{12}$C with a thin carbon target. Efforts are underway to interface new models to Geant4 dedicated to nuclear interactions below 100 MeV/u.[105,106]

### 4.C. 300 MeV/u $^{12}$C ion charge-changing cross section test

The goal of this test is to validate the partial and total charge-changing cross sections of $^{12}$C ions with energy 300 MeV/u as simulated by Geant4 against experimental data published in the literature by Toshito et al,[24] obtained with an emulsion plate in the NIRS P152 experiment. The test is named *C12FragCC* in the *geant-val* web interface.[12]

The partial charge-changing cross section to B, Be, and Li fragments is about $(428 \pm 21)$ mb, while the total is $(1183 \pm 52)$ mb (including the production of B, Be, H, He, Li, N, O, $\alpha$, deuteron, proton, neutron, and triton), accounting for approximately 36% with respect to the other ion species in the experiment. Therefore, a correct modeling of the carbon charge-changing process is needed for an accurate calculation of the dose in carbon ion therapy, especially out-of-field where organs at risk may be located, the LET and the radiobiological effectiveness (RBE).

#### 4.C.1. Simulation setup

In this simulation, all electromagnetic processes and decay physics are switched off and only hadron and ion transport processes are used to retrieve the carbon ion charge-changing cross section. The target geometry is a cubic water phantom with a 10 m length. The water is defined as G4_WATER in the Geant4 simulation. Carbon ions are irradiated from the surface of the phantom.

Only projectile-like fragments were considered in this work. Target-like fragments are also important to the dose; however, it is difficult to measure their cross section in an experiment. When an interaction occurs in the simulation, the distance from the initial point to the interaction point is summed up as the track length only for projectile-like fragments which can be identified as they travel to the forward region (momentum along the initiated direction $>$ 600 MeV/c/u).

The partial cross sections of B, Be, and Li isotopes were calculated by means of the simulation. This choice was dictated by the availability of experimental data in Toshito et al.[24]

#### 4.C.2. Results and discussion

The results of the test are shown in Fig. 15 and summarized in Table 8. The partial cross sections calculated by means of the *G4IonQMDPhysics* of B, Be, and Li isotopes are 144.1, 77.90, and 97.66 mb, respectively. An agreement within the uncertainty of the reference data was observed for the total $^{12}$C charge-changing cross section and for the Li isotope partial cross section. Differences of approximately 18% and 34% were found for the production of Be and B isotopes, respectively. The isotope production cross sections have a similar trend when substituting *QMD* with either *BIC* or *INCL*. The total fragmentation cross section is almost the same for all the ion physics constructors.

## 5. ELECTROMAGNETIC AND HADRONIC PHYSICS BENCHMARKING TESTS

The tests included in this section activate both electromagnetic and hadronic physics processes, using prebuilt Geant4 physics lists. *QGSP_BIC_HP*, *QGSP_BERT_HP*, and

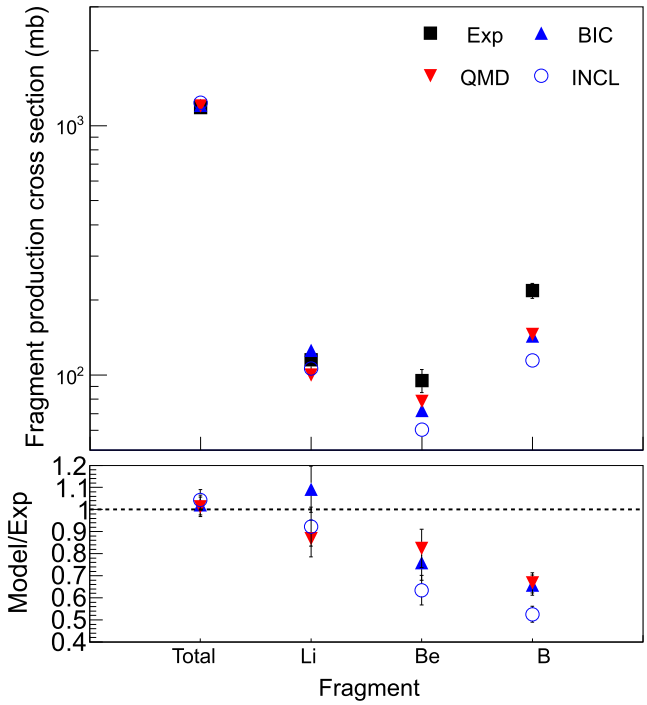


FIG. 15. Comparison of Geant4 against experimental data for $^{12}$C total and partial charge-changing cross sections. The total charge-changing cross-section includes the production of B, Be, H, He, Li, N, O, $\alpha$, deuteron, proton, neutron, and triton. The energy of the incident $^{12}$C ion beam is 300 MeV/u.

TABLE VIII. Differences between Geant4 simulation results and the reference experimental data published in Toshito et al.[24]

| Charge-changing cross section | Geant4 ion fragmentation model | | |
|---|---|---|---|
| | *QMD* | *BIC* | *INCL* |
| Total | <1 $\sigma_{ref}$ | <1 $\sigma_{ref}$ | <1 $\sigma_{ref}$ |
| Li partial cross section | <1 $\sigma_{ref}$ | <1 $\sigma_{ref}$ | <1 $\sigma_{ref}$ |
| Be partial cross section | 18% | 26% | 38% |
| B partial cross section | 34% | 36% | 48% |

$\sigma_{ref}$ represents the uncertainty of the experimental measurements.

*QGSP_BIC_EMY* were tested for the application of Geant4 in proton therapy. Details of those physics lists are reported below in the energy range of interest for proton therapy and summarized in Table 9.

In *HadronElasticPhysics*, elastic hadronic scattering of protons and neutrons is modeled by means of the CHIPS model.[107] Below an energy of 20 MeV, the high precision (HP) data libraries are used to model the neutron elastic scattering. For deuteron, triton, and $\alpha$ particles the elastic scattering model is based on the Geant3/GHEISHA approach.[98] In both *QGSP_BIC_HP* and *QGSP_BIC_EMY*, the ion elastic scattering is modeled. The cross section is calculated with the Glauber–Gribov model while the diffuse model describes the final state (documented in Grichine[108]).

In *QGSP_BIC_HP* and *QGSP_BIC_EMY*, the hadronic inelastic scattering of incident neutrons and protons on nucleons

TABLE IX. Table summarising the main features of the prebuilt Geant4 physics lists under investigation.

| Physics list | *QGSP_BIC_HP* | *QGSP_BIC_EMY* | *QGSP_BERT_HP* |
|---|---|---|---|
| EM constructor | *Opt4* | *Opt3* | *Opt0* |
| Hadron elastic scattering | *HadronElasticPhysics* | | |
| Hadron inelastic scattering | *BIC* < 9.9 GeV | | BERT < 9.9 GeV |
| Ion elastic scattering | Active | | Not active |
| Ion inelastic scattering | *G4BinaryLightIonReaction* for E < 4 GeV | | |
| HP data libraries | On | Off | On |
| Radioactive decay | On | On | Off |

In the case of *HP* on, the hadronic inelastic scattering of neutrons is described by means of the HP data libraries below 20 MeV (Allison et al[3]).

of the target nuclei is described by means of the Binary Intranuclear Cascade (*BIC* in Table 9) model, below approximately 10 GeV.[91] The *BIC* is followed by the Geant4 *Precompound* model describing the de-excitation of the remnant nucleus.[3] This is followed in turn by the evaporation of particles until the nucleus is fully de-excited (see the Geant4 Physics Reference Manual). In *QGSP_BERT_HP*, the proton and neutron inelastic scattering is described with the Bertini model,[109] which has its own precompound and evaporation models.

In all the investigated physics lists, the *LightIonBinaryCascade* describes the intranuclear cascade of deuteron, triton, He isotopes, and heavier nuclei (Allison et al[3]). Also in this case, the cascade is followed by the de-excitation of the remnant nucleus via the *Precompound* model and the nuclear evaporation.[3] HP (see Table 9) means that neutron inelastic scattering, capture, and fission are modeled using the HP data libraries below 20 MeV. Particle decay is modeled in the physics lists under study. The radioactive decay is activated by default in the *QGSP_BIC_HP* and *QGSP_BIC_EMY*, while it is not included in the *QGSP_BERT_HP*.

## 5.A. 67.5 MeV proton Bragg curves in water

The accurate calculation of the proton range is essential in MC calculations for proton therapy. While submillimeter accuracy in-depth penetration can be achieved by fine tuning of the proton energy, depth–dose curves measured from a proton beam with an energy accurately known are desirable for benchmarking purposes. In this test, the setup for comparing MC calculated Bragg peaks with experimental data of submillimeter (< 0.2 mm) uncertainty is described. The test is available in the *geant-val* web interface[12] with the name *LowEProtonBraggBeak*.

### 5.A.1. Simulation setup

A 67.5 MeV proton beam with a Gaussian energy spread of 0.4 MeV, and a uniform 5 mm × 10 mm oval spot travels

through a tantalum (Ta) foil of either 101.6 μm or 381 μm thickness. The geometry up to and including the water phantom is described in detail in Faddegon et al.[25] The sensitive region in the water phantom is 5 × 5 × 50 mm3 divided into 1000 slices. The depth–dose curves in water (density = 0.998 g/cm$^3$ and ionization potential I = 78 eV,[110] see Faggedon et al[25]) are calculated and then fitted with an analytical expression for the Bragg curve described in Bortfield.[111] The fitting parameters range, $R_0$, and spread, $s$, are compared against the corresponding experimental parameters from depth–dose measurement from Faddegon et al.[25] The range, $R_0$, is the depth at 80% distal falloff of the Bragg curve, whereas $s$ is the spread of a Gaussian describing the width of the Bragg peak according to Bortfeld et al.[111]

The physics list included alternatively *Livermore*, *Penelope*, *Opt3*, and *Opt4*. This test calculates the proton range which depends on the EM physics processes only.

Therefore, the hadronic physics component was modeled only by means of the *QGSP_BIC_HP* as described in Testa et al,[112] where a comprehensive validation of the Geant4 physics list for proton therapy was performed. A global production cut for secondary electrons of 50 μm was used everywhere with exception of the beam plug region. In there, a production cut of 20 μm was used. Simulations using the Bertini hadronic model gave the same results for the four tested EM physics constructors within statistical uncertainty.

#### 5.A.2. Results and discussion

The left side of Fig. 16 shows the depth–dose distributions in water for simulated (lines) data and measured data (symbols). In the right side of Fig. 16, the differences in Bragg peak spread ($s$) and range ($R_0$) are shown in the top and bottom of the figure, respectively. Differences in $R_0$ were within 2 standard deviations whereas ($\sigma$) for the $s$ were within 1 $\sigma_{\text{ref}}$, for all the EM constructors. The uncertainties shown in the figure include experimental, fitting and statistical uncertainties from the MC simulations. As depicted, the MC calculation underestimates $R_0$ within 0.5 mm, and overestimates $s$ within 0.05 mm.

In the end, the proton range and Bragg curve spread calculated with Geant4 10.05 were in acceptable agreement with the measured depth–dose curves within better than 0.5 mm.

### 5.B. Light ion Bragg Peak curves

The accurate modeling of Bragg curves of proton and light ion beams in water is of key importance to ensure code reliability for applications in hadron therapy. In this test, we compare our Geant4 calculations against mean range values estimated experimentally at GSI[113] for proton and $^{12}$C beams at energies of interest in particle therapy. The test is available in the *geantval* web interface[12] under the name *LightIonBraggPeak*.

#### 5.B.1. Simulation setup

The Bragg curve measurements were carried out by means of a water tank, whose thickness could be varied with micrometric accuracy, and two ionization chambers (IC) placed upstream (IC1) and downstream (IC2) in the tank. These measurements were carried out for protons and $^{12}$C ions, among other beam species, at various nominal energies to provide range-in-water values roughly from 3 to 30 cm. The ratio between the ionization registered by IC2 over that registered by IC1 was used to estimate the mean energy deposition at each depth with respect to the energy deposited at the entrance. According to a previous work,[114] the mean range of the beam was estimated to be at the depth where the distal edge of the Bragg curve was equal to 82% of the peak value, $R_{d82}$. More details of the experimental setup can be found in Schardt et al.[113]

For our calculations with Geant4, the geometry consisted of a simple water tank, modeled using G4_WATER material with density corrected according to room temperature (24° C, 0.997 g/cm$^3$). The energy deposition was scored by means of cylindrical volumes with a radius of 28 mm (equal to the radius of the ICs) and a thickness of 50 μm (similar to the water equivalent thickness of both ICs). Pencil beams were defined in our calculations with a mean energy equal to the reported nominal energy; the energy distribution was Gaussian, whose spread was estimated by fitting the width of the experimental distal fall-off, defined from the 90% and 10% of the peak value. For each type of beam, all

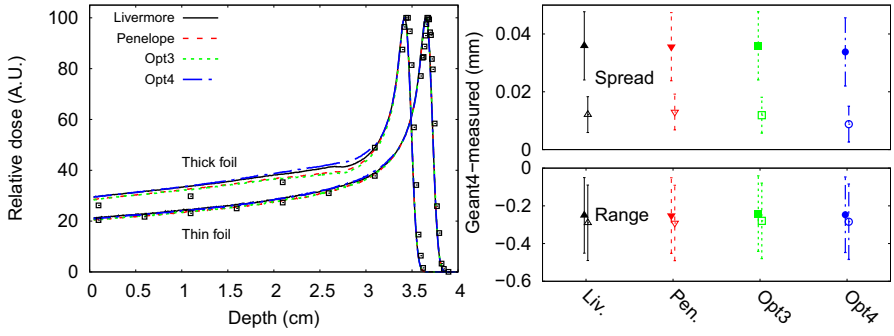


FIG. 16. Left: Depth–dose distributions in water for protons scattered with a Ta thin foil and a Ta thick foil calculated with four EM physics constructors, while keeping the hadronic physics modeled by means of the *QGSP_BIC_HP*. Experimental data[25]is shown with empty squares. Right: Differences between calculated to experimental parameters spread $s$ (top) and range $R_0$ (bottom) for the thin (empty markers) and thick (filled markers) foils. Errors and error bars represent 1$\sigma$ and include experimental and Monte Carlo uncertainties.

EM models matched experimental data with the same energy spread, which FWHM was of the order of, or below, 0.2% of the nominal energy in all cases. We tested the electromagnetic physics constructors: *Opt0*, *Opt3*, *Opt4*, *Livermore*, and *Penelope*; the hadronic physics was activated as defined in *QGSP_BIC_HP* physics list. Consistent with the experimental data, we calculated the depth, at distal edge, where the energy deposition was equal to the 82% of the maximum; this value was compared with those obtained from the experiments.

#### 5.B.2. Results and discussion

Figure 17 shows the absolute differences obtained in the value of $R_{d82}$ between measurements (suffix "exp") and Geant4 calculations ("G4") as a function of depth in water for protons (125–200 MeV/u) and $^{12}$C beams (100–400 MeV/u), respectively. In all cases, dose was calculated with statistical uncertainty below 1% (1$\sigma$) at the scoring volumes around the peak of the Bragg curve to ensure a small propagation to the calculated $R_{d82}$; thus, the uncertainty of $R_{d82}$ was estimated to be equal to the scoring volume thickness (50 μm). As for proton beams, the agreement was better than 0.2 mm ($\lesssim$0.1%) in all cases, which is considered remarkable, given that (a) the precision of the algorithm itself is within 2% and (b) particle therapy calculations require a precision of about 1 mm. As for $^{12}$C beams, on one hand, we could observe that *Opt0* showed range deviations of about 1.2%–2.5%, that is, similar to the precision of the Bethe–Bloch model used. On the other hand, the other EM physics constructors showed deviations about three times smaller, with a maximum absolute difference about 0.8 mm (0.3%) for the largest energy (400 MeV/u), and maximum relative difference of 1.1% for the smaller energy (100 MeV/u). This can be explained by the fact that the ion stopping is calculated in *Opt0* by means of an algorithm based on Bethe-Bloch formula, whereas in the other constructors a more precise method is implemented, based on a parameterization of stopping power values published by the ICRU[34] for some ions and materials, as described in Lechner et al.[115]

In summary, all electromagnetic physics constructors are able to successfully model the stopping of protons in water, whereas for $^{12}$C beams only *Opt3*, *Opt4*, *Livermore*, and *Penelope* are recommended.

### 5.C. Neutron yield of protons with energies 113 and 256 MeV and carbon ions at 290 MeV/u

Absolute neutron yields from protons and carbon ions at therapeutic energies interacting in targets of materials of interest for medical physics are presented. This benchmark shows the current status of two of the main Geant4 physics lists for hadronic physics used by the community in medical applications. For protons, two energies (113 and 256 MeV) have been selected, as they are important in clinical applications. The proton beams are incident on one of three targets made of aluminum, carbon, and iron, with thickness large enough to fully stop the beam. For carbon ions of 290 MeV/u, the beam is incident on a water (G4_WATER material) phantom of 18 cm thickness.

The results of the test are available in the *geant-val* web interface under the name of *ProtonC12NeutronYield*.[12]

#### 5.C.1. Simulation setup

Experimental neutron yields from thick targets of different materials bombarded by protons of energies below 256 MeV were obtained from Meier et al[116,117]; for carbon ions

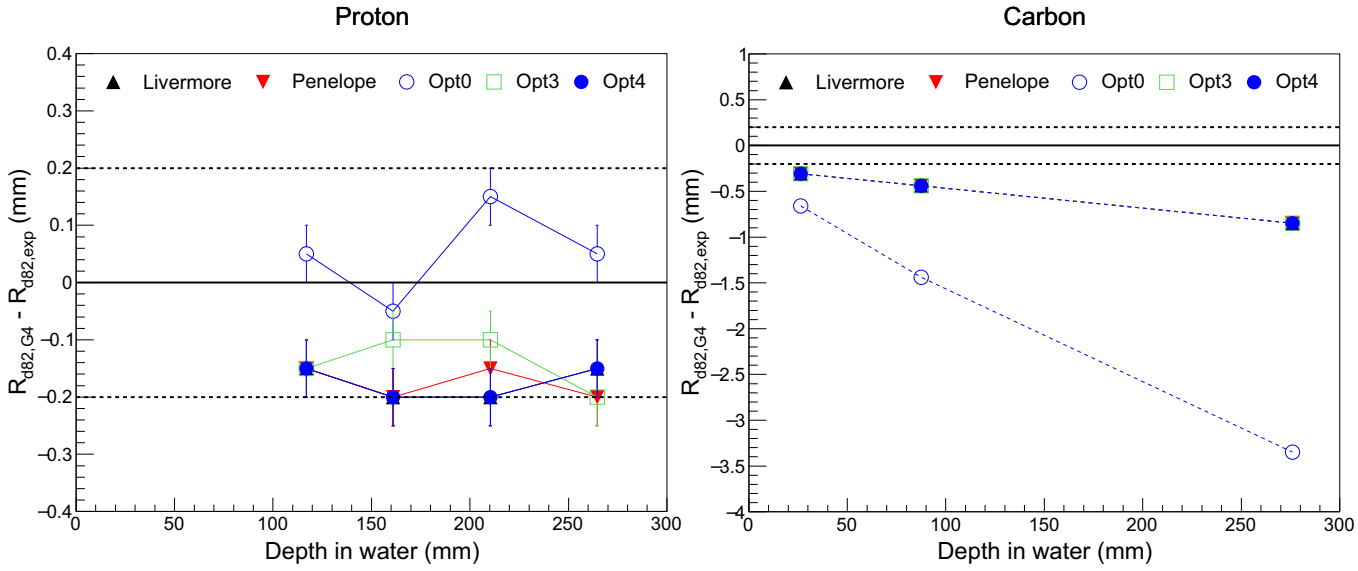


FIG. 17. Absolute differences obtained between the depth at which the energy deposited was 82% of peak, at distal Bragg Peak, calculated with Geant4 ($R_{d82,G4}$) and obtained experimentally ($R_{d82,exp}$), represented as function of the reported range in water for each beam. Left: Results for proton beams at 125, 150, 175, and 200 MeV/u; right: Results for $^{12}$C beams at 100, 200, and 400 MeV/u; here, *Livermore*, *Penelope*, *Opt3*, and *Opt4* symbols exactly overlap. The estimated uncertainty of $R_{d82,G4}$, represented with error bars, was 50 $\mu$m in all cases (see main text for details); for $^{12}$C, symbols are larger than the error bars.

TABLE X. Target dimensions, material and density for each proton energy and carbon ion beam configuration.

| Material | Radius (cm) | | Thickness (cm) | | Density (g/cm$^3$) |
|---|---|---|---|---|---|
| | | Protons | | | |
| | 113 MeV | 256 MeV | 113 MeV | 256 MeV | |
| Aluminum | 3.65 | 8.0 | 4.00 | 20.0 | 2.699 |
| Carbon | 3.65 | 8.0 | 5.83 | 30.0 | 1.867 |
| Iron | 3.65 | 8.0 | 1.57 | 8.0 | 7.867 |
| | | Carbon ions | (290 MeV/μ) | | |
| G4_WATER | – | – | | 18.0 | 1.0 |

interacting in water the data was obtained from Satoh et al.[118] For protons, the MC simulations modeled a cylindrical target irradiated with a uniformly distributed, zero divergence, monoenergetic (113 or 256 MeV) proton beam with a radius equal to the radius of the target. The beam was located upstream of the target. The radius, thickness, and material of each target are shown in Table 10.

Secondary neutrons produced in the inelastic nuclear interactions within the targets and escaped were scored at the surface of spherical rings (radius of 90 m) that covered specific angular bins of 4° of thickness. For carbon ions, a monoenergetic carbon ion beam of 0.25 cm radius, 290 MeV/u energy, and zero divergence was simulated. The target was a water phantom 18 cm thick. Detailed dimensions are shown in Satoh et al.[118] The neutron yield per incident primary particle per steradian per energy (equally spaced logarithmic energy intervals) was scored at specific angular bins. For proton beams, the neutron yield was scored at 7.5°, 30°, 60°, and 150° for 113 MeV, and at 30°, 60°, 120°, and 150° for 256 MeV. For carbon ions, the neutron yield was scored from 15°-90° at 15° angular steps.

The physics lists under study included alternatively *QGSP_BIC_HP* and *QGSP_BERT_HP_EMZ*. Both adopt the EM constructor *Opt4* (indicated with the suffix *EMZ* in the prebuilt Geant4 physics lists).

A global production cut of 0.05 mm was used.

### 5.C.2. Results and discussion

In Figure 18, top left, representative results are shown for the absolute neutron yield for protons (incident on aluminum) and carbon ions (on water) for several angles. Experimental data are shown with empty circles. In the figure, the ratio between the calculated to experimental integrated yields is shown for several angles and materials. For the carbon ion case, the yields were integrated from 2 MeV as reported in Satoh et al.[118] As shown, for protons there exists an overall better agreement (within 2 standard deviations) between *QGSP_BIC_HP* and the experimental yield for all the angles and materials compared to *QGSP_BERT_HP_EMZ*. However, the largest differences between the two physics lists

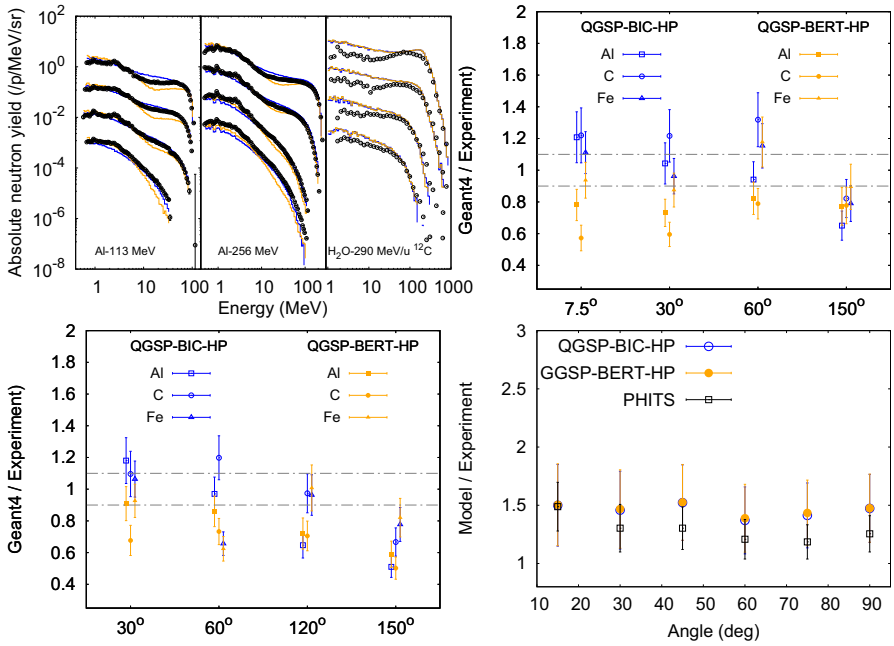


FIG. 18. Absolute neutron yield (top left) for protons of 113 and 256 MeV impinging an aluminum target and 290 MeV/u carbon ions in a water phantom. Geant4 calculation using *QGSP_BIC_HP* is shown with blue solid line, whereas results with *QGSP_BERT_HP* is shown with orange solid lines. Experimental data are shown with empty markers. Ratio between calculated to experimental integrated neutron yield for 113 MeV protons (top right), 256 MeV protons (bottom left), and 290 MeV/u carbon ions (bottom right). The gray dashed line represents the experimental uncertainty, one standard deviation ($\sigma_{ref}$). Error bars represent one standard deviation.

were found at larger angles for the aluminum target. For water material, there was no significant difference between the two physics list for all the angles. In that scenario, the yields were overestimated by up to 50%. As a comparative with other MC engines, results from Satoh et al.[118] using PHITS are shown and they agreed with Geant4 calculation within 1 standard deviation.

In summary, a marginally closer agreement with measured data was achieved with *QGSP_BIC_HP* physics list for high *Z* materials using the proton beams. For carbon ions, no statistically significant difference was found between the two hadronic physics lists used, because they both use the same physics models to describe the hadronic interactions of heavier ions. For the case involving carbon ions, the neutron yields were underestimated compared with experimental data by 50% on average, and by 11% compared with published data calculated with the PHITS code.

## 5.D. Fragmentation of a 400 MeV/u $^{12}$C ion beam in water

This test, published in Bolst et al,[26] benchmarks Geant4's modeling of the fragmentation process against experimental measurements of a $^{12}$C ion beam incident upon a water target performed at GSI.[119] The measured data includes fragment yields from hydrogen to boron for different thicknesses of water irradiated by a mono-energetic 400 MeV/u $^{12}$C ion pencil beam, as well as the fragment's angular and energy distributions. The full results of the test can be downloaded from the *geant-val* web interface.[12] The name of the test is *FragTest*.

### 5.D.1. Simulation setup

The simulation consists of a mono-energetic $^{12}$C ion pencil beam with an initial energy of 400 MeV/u and energy $\sigma$ of 0.15%, with a FWHM of 5 mm. The beam is incident upon a variable thickness of water with an area of 50 × 50 cm$^2$ and thicknesses of: 59, 159, 258, 279, 288, 312, and 347 mm.

Fragments were collected in a 2.94 m radius hemisphere centered on the water slab. For fragment yields, fragments were scored in a 10° cone from the center of the hemisphere. Fragment's angular distributions were scored in arcs on the hemisphere with angles of 0.4°, corresponding to the size of the detector used in the experiment. The energy distributions of fragments were calculated based on the time of flight of fragments scored in the 0.4° arcs.

Electromagnetic interactions were modeled using *Opt4*, with a global production cut of 10 mm for electrons, the production cut was set this high since the electron production did not affect the production of fragments. The *QGSP_BIC_HP* hadronic physics component was used with alternative physics models describing the hadronic inelastic scattering of heavy ions: *G4IonBinaryCascadePhysics* (*BIC*), *G4IonQMDPhysics* (*QMD*), and *G4IonINCLXXPhysics* (*INCL*). All default configurations were used for all models investigated in this work.

### 5.D.2. Results and discussion

Fragment yields scored within a 10° cone are shown in Fig. [19]. It can be seen that the yield varies significantly between the fragmentation models, up to ∼20% for particular fragment species. For the lighter fragments H and He, which represent the most abundant fragments, all alternative models agreed within ∼20% of experimental measurements. For this particular version of Geant4 (10.5), *INCL* gives the best agreement with experiment for H fragments while *QMD* agrees the best for the yield of He fragment. Sample angular and energy distributions are shown in Fig. 20, the simulation distributions shown here are normalized to the same area under the curve as the experiment. This is done to compare just the distributions produced by the simulation against the experimental data. *INCL* was seen to reproduce experimental angular distributions better than the other two models with the exception of H, where *QMD* reproduced the best. For kinetic energy distributions, *INCL* was seen to perform the poorest of the three models, with its energy distributions systematically being shifted ∼10% to lower energies compared to the other models for most measurements. *BIC* and *QMD* were seen to perform very similarly to one another for both the angular and energy distributions, with *QMD* performing slightly better out of the two. Due to different fragmentation models giving better agreement with experimental measurements for angular and energy distributions of fragments, the user should carefully consider their needs when selecting a model for fragmentation. The simulations were also performed with the Geant4 EM constructor *Opt3*; however no significant differences were found between the two compared to *Opt4*.

## 5.E. Test on cell survival curve modeling for proton therapy

This test is focused on evaluating the capability of Geant4 to reproduce *in-vitro* cell survival curves, also referred to as surviving fraction (SF), against experimental measurements,[120] performed at the CATANA facility radiating the radio-responsive prostate DU145 cells and breast cancer cell line MDA-MB-231,[27] at 20 mm depth in water, corresponding to the mid of a clinical 62 MeV modulated clinical proton beam. The test is called *Hadrontherapy* in the *geant-val* web interface.[12]

The SF is calculated with the linear quadratic model:

$$SF(D) = e^{-(\alpha D+\beta D^2)},$$

where SF(D) is the fraction of cells surviving when delivering a dose D; $\alpha$ and $\beta$ are constants describing the linear and quadratic components of cell killing, respectively.

Precompiled lookup tables (LUTs) contain the values of $\alpha_i$ and $\beta_i$ for a set of monoenergetic ions (with Z spanning from 1 to 8). $\alpha_i$ and $\beta_i$ are calculated with the local effect model (LEM)[121,122] and are tabulated as a function of the ion kinetic energy $E_i$ and delivered dose $D_i$ in a specific voxel of the

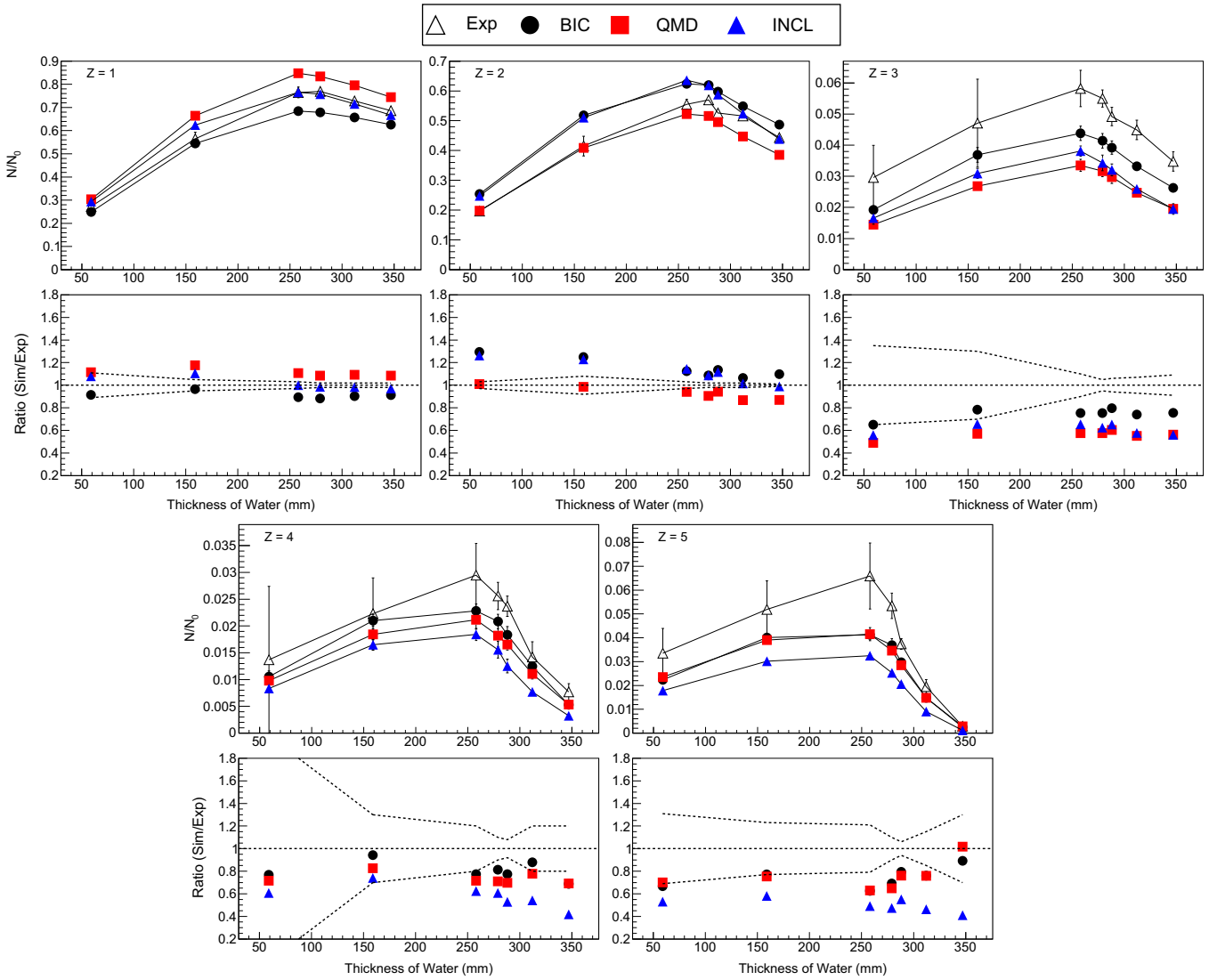


FIG. 19. The fragment yields, $N$, produced from $N_0$ monoenergetic 400 MeV/u $^{12}$C ions incident upon different thicknesses of water. The fragments were scored within a forward angle of 10° from the center of the water targets. Experimental measurements of fragments with different atomic number ranging from 1 to 5 are compared against alternative models available in Geant4. The dashed lines on the bottom ratio plots indicate the experimental uncertainty $\sigma_{\text{ref}}$.

simulated material. $E_i$ and $D_i$ can be derived from the Geant4 simulations and then used as input of an algorithm able to compute the average values of $\langle\alpha\rangle$ and $\langle\beta\rangle$ in a mixed radiation field, interpolating the best $\alpha_i$ and $\beta_i$ and using the following weighted formulas:

$$\langle\alpha\rangle = \frac{\sum_i \alpha_i \cdot D_i}{\sum_i D_i},$$

$$\langle\beta\rangle = \left(\frac{\sum_i \sqrt{\beta_i} \cdot D_i}{\sum_i D_i}\right)^2,$$

where the index $i$ refers to the different incident ions interacting in the voxel. This method is referred here as LEM III-weighted method. The LUTs are created taking into account the total LET at the corresponding position of the cell layer irradiation.[123] The input parameters of the LEM models were optimized referring to the radiobiological results. The threshold dose was fixed to 5 and 9 Gy for the MDA-MB-231 and DU-145, respectively, the nucleus radius was 4 μm in both cases.

### 5.E.1. Simulation setup

The experimental setup (transport beamline elements, detectors, irradiation device, etc.) used for cells irradiation has been modeled in detail using the *Hadrontherapy* advanced example.[124] The source consists of a 62 MeV proton beam characterized by a gaussian energy and angular distribution with a standard deviation of 0.3 MeV and 0.028°, respectively. The beam spot size was circular, with a gaussian spatial distribution (standard deviation of 3 mm). A plastic rotating modulator wheel[125] routinely used for real treatments

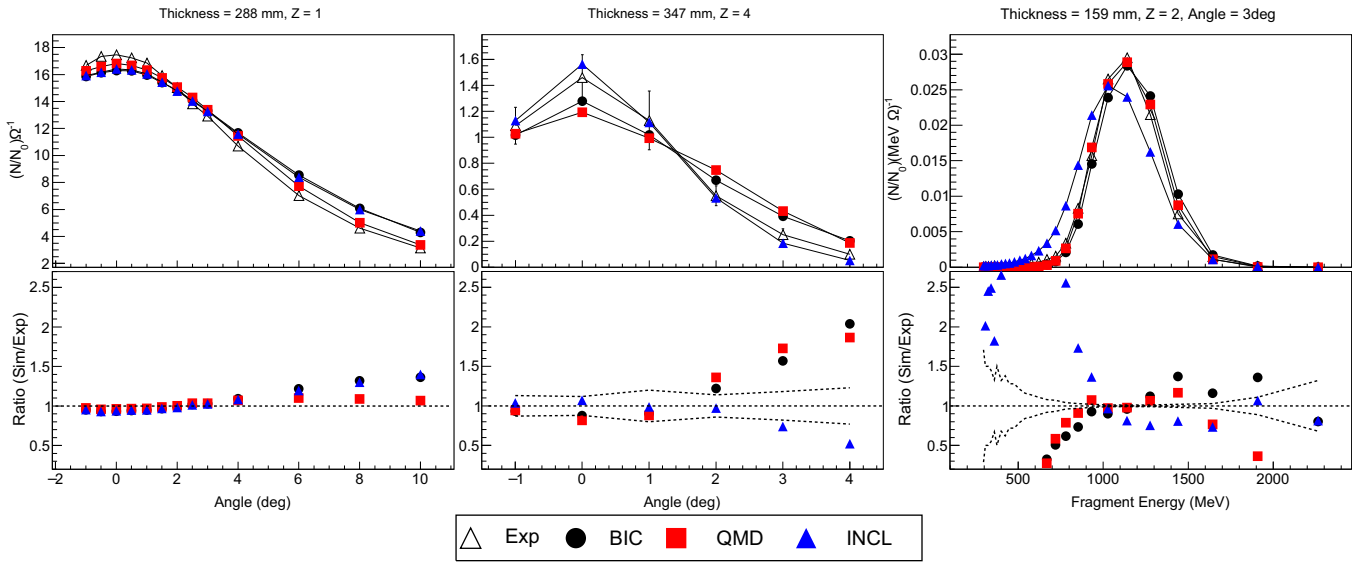


FIG. 20. Example angular and energy fragment distributions comparing experimental measurements against alternative models available in Geant4 for a 400 MeV/u mono-energetic $^{12}$C ion incident upon water. Left and Middle: Angular distribution of H (left) and Be (middle) produced with a water thickness of 288 mm and 347 mm, respectively. Right: An example energy distribution of He fragments recorded at an angle of 3° from the center of the $^{12}$C ion beam when incident upon 159 mm of water. The dashed lines on the bottom ratio plots indicate the experimental uncertainty $\sigma_{\text{ref}}$.

was simulated for the correct reconstruction of the modulated beam (clinical Spread Out Bragg Peak or SOBP). A total of $3.6 \cdot 10^8$ histories were simulated in each simulation. Dose as a function of depth is calculated in a voxelized water phantom positioned at the end of the beam line. The water tank was divided into $400 \times 400 \times 0.01$ mm$^3$ segments to reproduce the real experimental conditions. The water material is modeled with G4_WATER . Dose, fluence, and dose-averaged LET of the primary proton beam and generated secondaries were calculated in the simulations; $\alpha$ and $\beta$ values were determined using equations (3) and (4), along with the LUTs. The QGSP_BIC_HP physics list was used. The dose-averaged LET at the cells was 4.5 keV/$\mu$m. The global production cut was set to 0.01 mm.

### 5.E.2. Results and discussion

Figure 21 shows the SF curves as derived from the experimental measurements and the Geant4 simulation.

The agreement between experimental data and Geant4 simulations data was evaluated applying the $\chi^2$ test in the case of two independent distributions.

We obtained a $\chi^2$ value of 4.23 and 3.64 for the DU145 and MDA cell lines, respectively.

The $\chi^2$ test confirmed the good agreement between the two distributions even if, as expected, at doses >3 Gy the differences become more evident (well above the 20%).[27]

## 6. SUMMARY AND CONCLUSIONS

*G4-Med* is a group of currently 18 tests, regularly executed on the CERN computing infrastructure via the *geant-val* platform, to benchmark Geant4 against reference data for medical physics applications. *G4-Med* can also be used for regression testing purposes of Geant4. The results of the tests are provided to the Geant4 user community on-line via the *geant-val* web interface.[12]

This work was focused on the description of the tests of *G4-Med* and on the results of the benchmarking of Geant4 10.5 against reference data.

Table XI is intended to provide a summary of the main results of the electromagnetic physics tests, subject of Section III. We encourage the readers to refer to the specific tests of interest, for a more detailed analysis of the results in specific scenarios of application of Geant4.

The agreement is defined based on the discussion of the results of each individual test in Section III. *SS* and *GS* are not included in the Table as they have been used in a very limited set of tests.

In the case of the test concerning photon attenuation coefficients in water (*PhotonAttenuation*), slightly better agreement with respect to reference data was found for *Opt4*. The test includes one material only at this stage and we will expand it to other materials in the near future. We will also consider to extend the comparison of Geant4 to other reference data.

The electronic stopping power test (*ElectDEDX*) showed an agreement with ESTAR within the reference data uncertainty (1 $\sigma_{\text{ref}}$). This test is useful for regression testing purposes as it monitors a physical quantity important for the calculation of energy deposition in a target.

The electron backscattering test shows an agreement with respect to the available experimental data that clearly depends on the electron energy range and the Geant4 EM physics constructor. In particular below 5 keV the results should be taken with caution for any constructor, including the suggested

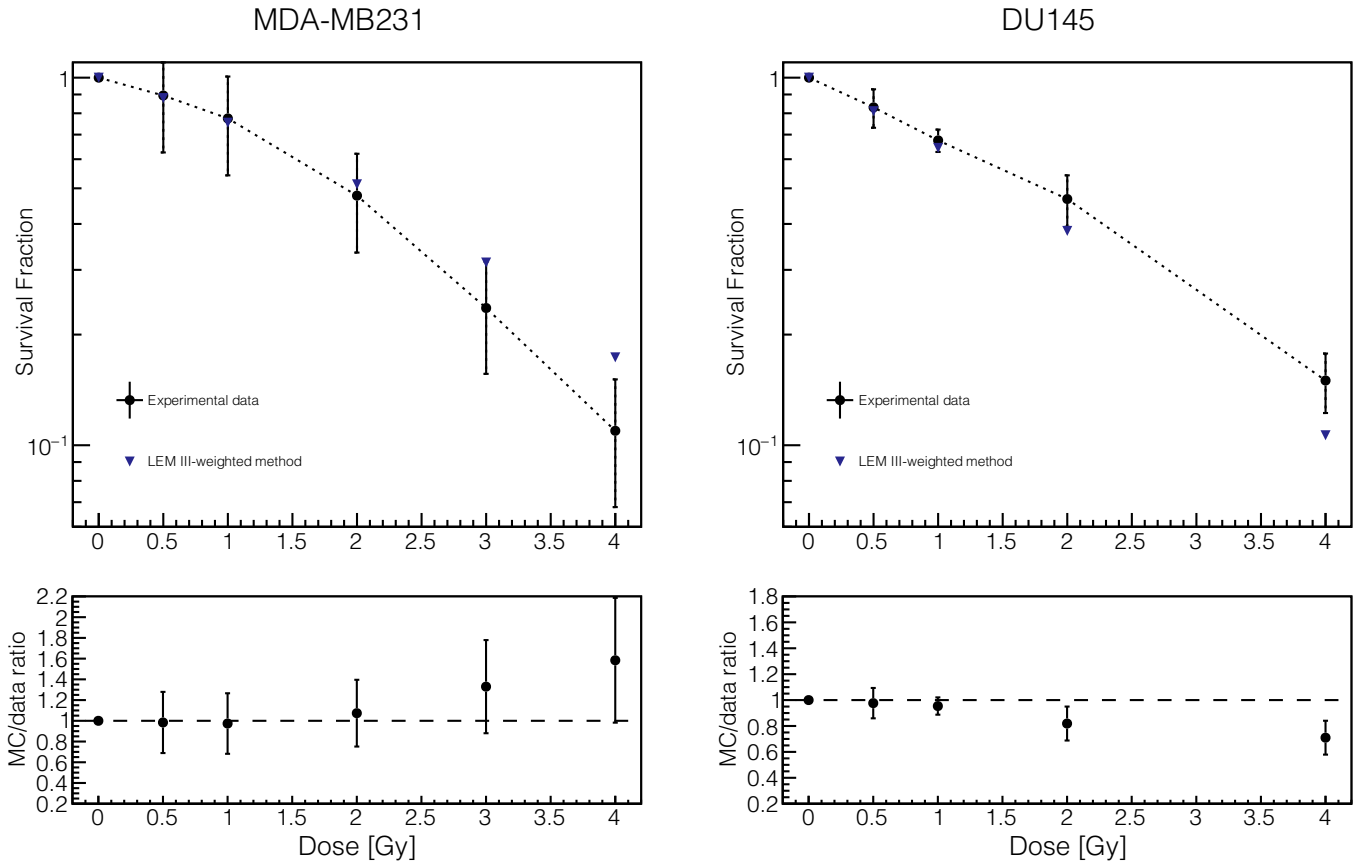


FIG. 21. Experimental cell surviving fraction (SF) measured irradiating DU145 cells and MDA-MB231 (dotted points) and the corresponding curves calculated with the LEM III-weighted method (triangles). Ratios of Geant4 simulations to experimental data are also plotted.

ones. The best EM physics constructors to model the electron backscattering were found to be *Opt4*, *SS*, *Livermore* and *Penelope*.

For electron forward scatter, the angular distribution of electron fluence at 1.18 m calculated with *Opt4* was significantly underestimated (> 1 standard deviation) in comparison to experimental data for foils of higher $Z$ than Be by 2–5%, with calculated results comparable to or better than the comparison with the other physics constructors.

For bremsstrahlung from thick Be, Al, and Pb targets at radiotherapy energies, bremsstrahlung yields (≤1–2 standard deviations) were in better agreement with experimental data when using *Opt3* and *Opt4*. For these constructors, the differences were significant (> 1 standard deviation) at larger angles for the higher $Z$ targets (≥60°).

The Fano cavity test (*FanoCavity*) was designed specifically to check the accuracy of the condensed history electron transport and this test demonstrates that the GS multiple scattering model (included in *Opt4*, *Livermore*, and *Penelope*) provides a high level of accuracy and stability. Similar results were obtained for the *WVI** constructor.

The Dose Point Kernel test (*LowEElecDPK*) demonstrates that the *Opt4* constructor provides the best agreement with the selected standard (EGSnrc). This test is currently used as a regression testing benchmark for Geant4 allowing to verify the stability of DPK profiles along with Geant4 updates, especially when electron multiple scattering models undergo changes or updates. It remains impossible to validate Geant4 in terms of DPK in the absence of experimental data in water in the low electron energy range.

The results of the microdosimetry test (*microyz*) demonstrate that, when using Geant4 condensed history approach in EM simulations of electron transport in microdosimetric applications, it is important to limit the step size. It is shown that 1/10 of the sensitive volume diameter is adequate.

The Brachytherapy test (*Brachy-Ir*) has a regression testing purpose. Geant4 is compared against previous simulation results in calculating the radial dose rate distribution of a $^{192}$Ir source. This test is of interest to monitor the evolution of Geant4 and will be extended in the near future to other brachytherapy sources. We intend also to compare the results of the simulation to dosimetric experimental measurements available in the literature.

The test on monoenergetic x-ray internal breast dosimetry (*Mammo*) showed an agreement between Geant4 and experimental data within the experimental uncertainty. It was observed that *Opt0* provided the worst performance among the EM constructors under study.

In summary, the EM tests showed that *Opt*0 in general is the constructor with the least agreement with respect to the

TABLE XI. Summary of the results of the electromagnetic physics tests of Section 3.

| Test | Reference | Difference with reference data |
|---|---|---|
| *Photon attenuation* | NIST XCOM (Berger et al[41]) | $\leq 1\ \sigma_{ref}$ for the total $\mu/\rho$, more details in Table IV |
| *ElectDEDX* | ESTAR (Berger et al[127]) | $< 1\ \sigma_{ref}$ |
| *ElectBackScat* | Exp data in Table V | $< 3\ \sigma_{ref}$ (< 5%) for $E_{e^-} > 0.2$ MeV < 15% for 5 keV <$E_{e^-}$ < 0.2 MeV and < 40% for E <5 keV for *Opt0* and *Opt3*, < 30% for E <5 keV for *Opt4*, *Penelope* and *Livermore* |
| *ElecForwScat* | Ross et al[61] Exp data | Characteristic angle ≤ 3% for Be, Al, Cu, Ta, and Au foils, 3%-5% for Ti and C foils, for all EM constructors |
| *Bremsstrahlung* | Faddegon et al[63] Exp data | $< 2\ \sigma_{ref}$ for *Opt0*, *Penelope*, and *Livermore* $< 1\ \sigma_{ref}$ for emission angles < 60° for *Opt3* and *Opt4* $< 2\ \sigma_{ref}$ for emission angles > 60° for *Opt3* and *Opt4* |
| *Fano cavity* | NA | < 2% for *Opt0*, 0.03% for *Opt3*, 0.02% for *Opt4* for *dRoverRange* equal to 0.2 (default value, see Table III) |
| *LowEElecDPK* | EGSnrc | < 10% for *Opt3* (same for *Opt0*) < 3% for *Opt4*, *Livermore* and *Penelope* |
| *Brachy-Ir* | Granero et al[83], Geant4 7.1 | $< 1\ \sigma_{ref}$ |
| *Mammo* | Fedon et al[21,22] Exp data | < 7.7% for *Opt0* and < 6% ($\sigma_{ref} = 5.7\%$) for the other EM constructors |

The second column reports the type of reference data, either from Monte Carlo simulations, theoretical or experimental (Exp) data. $\sigma_{ref}$ is the uncertainty of the reference data.

reference data. For the other physics constructors, the results of the tests should support users to select the appropriate EM physics constructor for their medical physics application. It was observed that *Opt4* had good agreement with the reference data for all the tests. This result was expected as this constructor was developed to provide the most accurate Geant4 physics models, irrespective of CPU performance.

Table XII summarizes the results obtained for the hadronic tests subject of Section 4.

The test of nucleus–nucleus hadronic inelastic scattering cross sections (*NucNucInelXS*) considered a limited number of projectiles of interest for hadron therapy (protons and $^{12}$C ions), incident on a limited number of targets. The choice of targets was performed based on the availability of experimental measurements in the EXFOR data library.[96] In general, an agreement within approximately 10% was found between Geant4 and the experimental data for the cases under study. The next step of this test is to find more reference experimental measurements to be able to benchmark the Geant4 inelastic scattering cross sections for more targets of interest for medical physics applications.

The 62 MeV/u $^{12}$C fragmentation test (*LowEC12Frag*) showed that both the Geant4 Ion Binary Cascade and the

TABLE XII. Summary of the results of the hadronic physics tests of Section 4.

| Test | Reference | Difference with reference data |
|---|---|---|
| *NucNucInelXS* | Zerkin et al[97] | < 10% overall |
| *C12FragCC* | Toshito et al[24] | $< 1\ \sigma_{ref}$ for total and Li cross sections |
| | | 18% for Be isotopes, 34% for B isotopes for *QMD* |
| | | 26% for Be isotopes, 36% for B isotopes for *BIC* |
| | | 38% for Be isotopes, 48% for B isotopes for *INCL* |

The second column reports the reference data, which are all experimental and published in the literature. $\sigma_{ref}$ is the uncertainty of the reference data.

*INCL* models have limitations in reproducing doubly differential cross sections of fragments produced in the interaction of $^{12}$C with a thin carbon target. The results of this test show a possible domain of improvement of Geant4 of interest for carbon ion therapy.

The test on the 300 MeV/u $^{12}$C ion charge-changing cross section (*C12FragCC*) demonstrated that all hadronic ion fragmentation models, *G4IonQMDPhysics* (*QMD*), *G4IonBinaryCascade*(*BIC*), and *G4IonINCLXXPhysics* (*INCL*), reproduce the total and Li partial charge-changing cross sections. However, they all underestimate the reference experimental data for Be and B fragments.

Table 13 summarizes the results obtained for the hadronic tests subject of Section 5, where the EM constructor is varied while keeping the hadronic physics component modeled with *QGSP_BIC_HP*.

The test modeling a 67.5 MeV proton Bragg Peak in water (*LowEProtonBraggPeak*) showed that Geant4 can reproduce the Bragg peak spread within the uncertainty of the experimental measurements and the range within 0.5 mm accuracy. This result was confirmed for higher proton energies (125–200 MeV/u) by the light ion Bragg curve test (*LightIonBraggPeak*). In this case, an agreement within 0.1%

TABLE XIII. Summary of the results of the hadronic physics tests of Section 5.

| Test | Reference | Physical quantity | Difference with reference data |
|---|---|---|---|
| *LowEProton BraggPeak* | Faddegon et al[25] | 67.5 MeV proton Bragg Peak | |
| | | Spread Range | $< 1\ \sigma_{ref}$ (within 0.5 mm) $< 2\ \sigma_{ref}$ (within 0.5 mm) for all EM constructors |
| *LightIon BraggPeak* | Schardt et al[14] | 125–200 MeV/μ protons, $R_{d82}$ | < 0.1% (< 0.2 mm) < 1.1% (< 0.8 mm) for all cases apart from *Opt0* |
| | | 100–400 MeV/μ $^{12}CC$, $R_{d82}$ | |

The EM constructor is tested (*Opt0*, *Opt3*, *Opt4*, *Livermore* and *Penelope*) while keeping the same hadronic physics component (*QGSP_BIC_HP*). All the reference data are experimental measurements. $\sigma_{ref}$ is the uncertainty of the reference data.

TABLE XIV. Summary of the results of the hadronic physics tests of Section 5.

| Test | Reference | Physical quantity | Difference with respect to reference data |
|---|---|---|---|
| *ProtonC12 NeutronYield* | Meier et al[117] Meier et al[118] | Neutron yield of 113 MeV and 256 MeV p | $< 2\ \sigma_{ref}$ for *QGSP_BIC_HP*, $< 4\ \sigma_{ref}$ for *QGSP_BERT_HP* |
| *ProtonC12 NeutronYield* | Satoh et al[119] | Neutron yield 290 MeV/μ $^{12}$C | < 11% against PHITS < 50% against exp data for both for *QGSP_BIC_HP* and *QGSP_BERT_HP* |
| *FragTest* | Haettner et al[120] | 400 MeV/μ $^{12}$C fragmentation yield | *QGSP_BIC_HP* for hadrons *BIC* and *QMD*: < 25% , *INCL*:< 40% |
| *Hadrontherapy* | Petringa et al[27] | Cell survival test 62 MeV proton beam | *QGSP_BIC_HP* < 20% for DU145 > 20% for MDA-MB-231 |

The EM constructor is kept constant (*Opt4*) while changing the hadronic physics component. All reference data are experimental. $\sigma_{ref}$ is the uncertainty of the reference data.

(<0.2 mm) was observed for $R_{d82}$. In the case of incident 100–400 MeV/u $^{12}$C ions, an agreement within 1.1% was found for all EM constructors apart from *Opt0* (Table XIII).

Table XIV summarizes the results obtained for the hadronic tests subject of Section 5, where the EM constructor is kept constant while changing the hadronic physics component.

The neutron yield produced by 113 and 256 MeV protons showed a marginally closer agreement when using *QGSP_BIC_HP* rather than *QGSP_BERT_HP*.

The 400 MeV/u carbon ion fragmentation test (*FragTest*) showed that *BIC* and *QMD* produce similar yields of light fragments (Z = 1–5), with similar kinetic energy and angular distribution. Worst agreement in terms of fragments yield was observed for *INCL*, however this model produces the best description in terms of angular distributions of the fragments. The results of this test are consistent with the outcome of the charge-changing cross section test which demonstrated a better agreement between *BIC* and *QMD* to the reference data than *INCL*.

The test on the cell survival curves (*Hadrontherapy*), of interest for protontherapy, is intended to validate Geant4 against *in-vitro* radiobiological measurements. In particular, surviving fraction curves calculated with *Hadrontherapy* were compared with experimental data for different cell lines. A good agreement was found for low doses (up to 3 Gy), while significant discrepancies were evident at higher delivered doses. Provided the current stage of research in *in silico* radiobiological modeling (e.g., Abolfath et al[127]), the overall agreement was deemed satisfactory.

For hadron therapy, it is recommended not to use the EM constructor *Opt0*. The hadronic tests showed that *QGSP_BIC_HP* physics list (which models EM interactions with *Opt4* in Geant4 10.5) provides an overall adequate description of the physics involved in hadron therapy, including proton and carbon ion therapy. However, it should be noted that only few physical quantities (position and spread of the Bragg Peak, neutron yields, fragment yields) have been subject of testing. New tests should be included in the next stage of the project to validate Geant4 in terms of dosimetry against experimental measurements.

Concerning the carbon ion fragmentation tests, the results show a slightly better performance when using *QMD* and *BIC* ion fragmentation models. The charge-changing cross section and fragmentation tests show that ion fragmentation modeling is a domain of possible improvement of Geant4.

In conclusion, we encourage Geant4 users to use benchmarked Geant4 physics constructors and lists. The results presented and discussed in this paper will aid users tailoring physics lists to their particular application.

The next phase of the project will proceed in two main directions. The first one is the benchmarking of the computing performance of the prebuilt Geant4 physics constructors and lists. We will use the regression testing to investigate the evolution of the Geant4 physics component for medical applications in terms of both accuracy and computing performance. The second main direction will be to extend the current tests and include new ones to cover domains which are currently not tested, for example, x-ray radiotherapy and nuclear medicine.

## ACKNOWLEDGMENTS

The authors thank Krzysztof Genser and Daren Sawkey for their feedback aimed to improve the quality of the paper. D. Bolst acknowledges the support of the Australian Government Research Training Program Scholarship. M. A. Cortés-Giraldo, A. Perales, and J. M. Quesada acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness under grant FPA2016-77689-C2-1-R. B. Faddegon and J. Ramos-Méndez acknowledge partial financial support from the NIH grant U24CA215123. D. Bolst, S. Guatelli, D. Sakata, S. Incerti, and I. Kyriakou acknowledge financial support from the Australian Research Council, ARC DP170100967. S. Incerti acknowledges the financial support of CNRS through the IN2P3/MOVI Master Project and through the France-Greece PICS 8235 funding scheme. I. Kyriakou acknowledges additional financial support from the European Space Agency (Contract No. 4000126645/19/NL/BW). E. C. Simpson acknowledges financial support from the Australian Research Council under grant DP170102423. I. Sechopoulos and C. Fedon acknowledge financial support from the Susan G Komen Foundation for the Cure grant IIR13262248. S. Guatelli and D. Bolst acknowlege the use of computing resources of the Australian National Computing Infrastructure (NCI), through the NCMAS 2020 grant scheme.

[a)]Author to whom correspondence should be addressed. Electronic mail: susanna@uow.edu.au.

## REFERENCES


1. Agostinelli S, Allison J, Amako K, et al. Geant4 -a simulation toolkit. *Nucl Instrum Methods Phys Res Sect A*. 2003;506:250–303.
2. Allison J, Amako K, Apostolakis J, et al. Geant4 developments and applications. *IEEE Trans Nucl Sci*. 2006;53:270–278.
3. Allison J, Amako K, Apostolakis J, et al. Recent developments in Geant4. *Nucl Instrum Methods Phys Res Sect A*. 2016;835:186–225.
4. Guatelli S, Incerti S. Monte carlo simulations for medical physics: from fundamental physics to cancer treatment; 2017.
5. Archambault L, Beaulieu L, Carrier J-F, et al. Overview of Geant4 applications in medical physics. In: *2003 IEEE Nuclear Science Symposium Conference Record (IEEE Cat. No. 03CH37515)*, vol. 3, pp. 1743–1745. IEEE; 2003.
6. Arce P, Ignacio Lagares J, Harkness L, et al. Gamos: a framework to do Geant4 simulations in different physics fields with an user-friendly interface. *Nucl Instrum Methods Phys Res Sect A*. 2014;735:304–313.
7. Jan S, Benoit D, Becheva E, et al. GATE v6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Phys Med Biol*. 2011;56:881–901.
8. Akagi T, Aso T, Faddegon B, et al. The PTSim and TOPAS projects, bringing Geant4 to the particle therapy clinic. *Progr Nucl Sci Technol*. 2011;2:912–917.
9. Akagi T, Aso T, Iwai G, et al. Geant4-based particle therapy simulation framework for verification of dose distributions in proton therapy facilities. *Prog Nucl Sci Technol*. 2014;4:896–900.
10. Perl J, Shin J, Schümann J, Faddegon B, Paganetti H. Topas: an innovative proton Monte carlo platform for research and clinical applications. *Med Phys*. 2012;39:6818–6837.
11. https://twiki.cern.ch/twiki/bin/view/Geant4/G4MSBG.
12. https://geant-val.cern.ch/layouts. *geant-val*web interface.
13. https://www.geant4.org/geant4. Geant4 Collaboration web page.
14. Freyermuth L, Konstantinov D, Latyshev G, Razumov I, Pokorski W, Ribon A. Geant-val: a web application for validation of detector simulations. *EPJ Web Conf*. 2019;214:05002.
15. Amako K, Guatelli S, Ivanchenko V, et al. Validation of Geant4 electromagnetic physics versus the NIST databases. *IEEE Trans Nucl Sci*. 2005;52:910.
16. Berger MJ, Hubbell JH, Seltzer SM, et al.NIST XCOM: Photon Cross Sections Database; 2010. https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html
17. Dondero P, Mantero A, Ivanchencko V, Lotti S, Mineo T, Fioretti V. Electron backscattering simulation in Geant4. *Nucl Instrum Methods Phys Res Sect B*. 2018;425:18–25.
18. Faddegon BA, Kawrakow I, Kubyshin Y, Perl J, Sempau J, Urban L. The accuracy of EGSnrc, Geant4 and PENELOPE Monte Carlo systems for the simulation of electron scatter in external beam radiotherapy. *Phys Med Biol*. 2009;54:6151.
19. Faddegon BA, Asai M, Perl J, et al. Benchmarking of Monte Carlo simulation of bremsstrahlung from thick targets at radiotherapy energies. *Med Phys*. 2008;35:4308–4317.
20. Kyriakou I, Emfietzoglou D, Ivanchenko V, et al. Microdosimetry of electrons in liquid water using the low-energy models of Geant4. *J Appl Phys*. 2017;122:024303.
21. Fedon C, Caballo M, Longo R, Trianni A, Sechopoulos I. Internal breast dosimetry in mammography: experimental methods and Monte carlo validation with a monoenergetic x-ray beam. *Med Phys*. 2018a;45:1724–1737.
22. Fedon C, Caballo M, Sechopoulos I. Internal breast dosimetry in mammography: Monte carlo validation in homogeneous and anthropomorphic breast phantoms with a clinical mammography system. *Med Phys*. 2018b;45:3950–3961.
23. Mancini-Terracciano C, Caccia B, Colonna M, et al. Validation of Geant4 nuclear reaction models for hadron therapy and preliminary results with BLOB. In: *World Congress on Medical Physics and Biomedical Engineering 2018*. Berlin: Springer; 2019:675–685.
24. Toshito T, Kodama K, Sihver L, et al. Measurements of total and partial charge-changing cross sections for 200-to 400-mev/nucleon c 12 on water and polycarbonate. *Phys Rev C*. 2007;75:054606.
25. Faddegon BA, Shin J, Castenada CM, Ramos-Méndez J, Daftari IK. Experimental depth dose curves of a 67.5 mev proton beam for benchmarking and validation of Monte Carlo simulation. *Med Phys*. 2015;42:4199–4210.
26. Bolst D, Cirrone GA, Cuttone G, et al. Validation of Geant4 fragmentation for heavy ion therapy. *Nucl Instrum Methods Phys Res Sect A*. 2017;869:68–75.
27. Petringa G, Romano F, Manti L, et al. Radiobiological quantities in proton-therapy: estimation and validation using Geant4-based Monte Carlo simulations. *Phys Medica*. 2019;58:72–80.
28. https://geant4.web.cern.ch/support/user_documentation. Geant4 Physics Reference Manual.
29. Apostolakis J, Asai A, Bogdanov A, et al. Geometry and physics of the Geant4 toolkit for high and medium energy applications. *Radiat Phys Chem*. 2009;78:859–873.
30. Ivanchenko V, Incerti S, Allison J, et al. Geant4 electromagnetic physics: improving simulation performance and accuracy. In: *SNA+ MC 2013-Joint International Conference on Supercomputing in Nuclear Applications+ Monte Carlo*: 03101. EDP Sciences; 2014.
31. Allison J, Apostolakis J, Bagulya A, et al. Geant4 electromagnetic physics for high statistic simulation of LHC experiments. *J Phys*. 2012;396:022013.
32. Chauvie S, Guatelli S, Ivanchenko V, et al. Geant4 low energy electromagnetic physics. In: *IEEE Symposium Conference Record Nuclear Science 2004*, vol. 3:1881–1885. IEEE; 2004.
33. Baro J, Sempau J, Fernández-Varea J, Salvat F. Penelope: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter. *Nucl Instrum Meth Phys Res Sect B*. 1995;100:31–46.
34. Ivanchenko V, Bagulya A, Bakr S, et al. Progress of Geant4 electromagnetic physics developments and applications. In: *EPJ Web of Conferences*, vol. 214:02046. EDP Sciences; 2019.
35. Sigmund P, Schinner A, Paul H. Errata and addenda for ICRU report 73, stopping of ions heavier than helium. *J ICRU*. 2009;5:1–10.
36. Brown JMC, Dimmock MR, Gillam JE, Paganin D. A low energy bound atomic electron Compton scattering model for Geant4. *Nucl Instrum Methods Phys Res Sect B*. 2014;338:77–88.
37. Bernard D. Polarimetry of cosmic gamma-ray sources above e+ e- pair creation threshold. *Nucl Instrum Methods Phys Res Sect A*. 2013;729:765–780
38. Incerti S, Ivanchenko V, Novak M. Recent progress of Geant4 electromagnetic physics for calorimeter simulation. *J Instrum*. 2018a;13:C02054.
39. Ivanchenko V, Kadri O, Maire M, Urban L. Geant4 models for simulation of multiple scattering. *J Phys*. 2010;219:032045.
40. Geant4 Application Developer Manual. available online at: https://www.geant4.org/geant4.
41. Berger MJ, Hubbell J. Xcom: Photon cross sections on a personal computer. Technical Report, National Bureau of Standards, Washington, DC (USA). Center for Radiation; 1987.
42. Hubbell JH, Gimm HA. Pair, triplet, and total atomic cross sections (and mass attenuation coefficients) for 1 Mev-100 GeV photons in elements Z=1 to 100. *J Phys Chem Ref Data*. 1980;9:1023.
43. Cirrone G, Cuttone G, Di Rosa F, Pandola L, Romano F, Zhang Q. Validation of the Geant4 electromagnetic photon cross sections for elements and compounds. *Nucl Instrum Methods Phys Res Sect A*. 2010;618:315–322.
44. Berger MJ, Coursey JS, Zucker MA, Chang J. NIST Standard Reference Database, 124; 2017https://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html
45. Thomas DJ. ICRU report 85: fundamental quantities and units for ionizing radiation; 2012.
46. Sakata D, Incerti S, Bordage MC, et al. An implementation of discrete electron transport models for gold in the Geant4 simulation toolkit. *J Appl Phys*. 2016;120:244901.
47. Lockwood GJ, Ruggles LE, Miller GH, Halbleib J. Calorimetric measurement of electron energy deposition in extended media. Theory vs experiment. Technical Report, Sandia Labs., Albuquerque, NM (USA); 1980.
48. Lockwood GJ, Ruggles LE, Miller GH, Halbleib J. Electron energy and charge albedos-calorimetric measurement vs Monte Carlo theory. Technical Report, Sandia National Labs., Albuquerque, NM (USA); 1981.

49. Reimer L, Tollkamp C. Measuring the backscattering coefficient and secondary electron yield inside a scanning electron microscope. *Scanning*. 1980;3:35–39.
50. Hunger H-J, Küchler L. Measurements of the electron backscattering coefficient for quantitative EPMA in the energy range of 4 to 40 kev. *Physica Status Solidi (a)*. 1979;56:K45–K48.
51. Cosslett V, Thomas R. Multiple scattering of 5-30 kev electrons in evaporated metal films iii: backscattering and absorption. *Br J Appl Phys*. 1965;16:779.
52. Sim KS, Kamel N. A new setup and measurement technique for SNR. *Phys Lett A*. 2003:317–378.
53. Böngeler R, Golla U, Kässens M, et al. Electron-specimen interactions in low-voltage scanning electron microscopy. *Scanning*. 1993;15:1–18.
54. Bishop H. Electron scattering and x-ray production. PhD thesis, University of Cambridge; 1966.
55. Heinrich K. Electron probe microanalysis by specimen current measurement. Proceedings of the Fourth International Congress X-Ray Optics and Microanalysis. Hermann, Paris. 1966:159–167.
56. Neubert G, Rogaschewski S. Backscattering coefficient measurements of 15 to 60 kev electrons for solids at various angles of incidence. *Phys Status Solidi (a)*. 1980;59:35–41.
57. Drescher H, Reimer L, Seidel H. Backscattering and secondary electron emission of 10-100 kev electrons in scanning electron microscopy. *Z. ANGEW PHYSIK*. 1970;29:331–336.
58. Wittry D. Secondary electron emission in the electron probe. In: *Proc. 4th Int. Conf. on X-ray Optics and X-ray Microanalysis*: 168–180; 1966.
59. Bronshtein I, Fraiman BS. Vtorichnaya elektronnaya emissiya.(secondary electron emission); 1969.
60. Assa'd A, El Gomati M. Backscattering coefficients for low energy electrons. *Scann Microsc*. 1998;12:185–192.
61. Ross C, McEwen M, McDonald A, Cojocaru C, Faddegon B. Measurement of multiple scattering of 13 and electrons by thin foils. *Med Phys*. 2008;35:4121–4131.
62. Vilches M, García-Pareja S, Guerrero R, Anguiano M, Lallena A. Multiple scattering of 13 and 20 mev electrons by thin foils: A Monte Carlo study with Geant, Geant4, and Penelope. *Med Phys*. 2009;36:3964–3970.
63. Faddegon BA, Ross C, Rogers D. Forward-directed bremsstrahlung of 10- to 30-Mev electrons incident on thick targets of Al and Pb. *Med Phys*. 1990;17:773–785.
64. Faddegon BA, Ross C, Rogers D. Angular distribution of bremsstrahlung from 15-mev electrons incident on thick targets of Be, Al, and Pb. *Med Phys*. 1991;18:727–739.
65. Poon E, Seuntjens J, Verhaegen F. Consistency test of the electron transport algorithm in the Geant4 Monte Carlo code. *Phys Med Biol*. 2005;50:681.
66. Kawrakow I. Accurate condensed history Monte Carlo simulation of electron transport. ii. Application to ion chamber response simulations. *Med Phys*. 2000;27:499–513.
67. Fano U. Note on the Bragg-Gray cavity principle for measuring energy dissipation. *Radiat Res*. 1954;1:237–240.
68. Simiele E, DeWerd L. On the accuracy and efficiency of condensed history transport in magnetic fields in Geant4. *Phys Med Biol*. 2018;63:235012.
69. Prestwich WV, Nunes J, Kwok CS. Beta dose point kernels for radionuclides. *J Nucl Med*. 1989;30:1036–1046.
70. Simpkin DJ, Mackie TR. Egs4 Monte Carlo determination of the beta dose kernel in water. *Med Phys*. 1990;17:179–186.
71. Mainegra-Hing E, Rogers D, Kawrakow I. Calculation of photon energy deposition kernels and electron dose point kernels in water. *Med Phys*. 2005;32:685–699.
72. Perrot Y. Evaluation de la dose déposée par des faisceaux d'électrons en radiothérapie dans des fantômes voxélisés en utilisant la plateforme de simulation Monte Carlo GATE fondée sur Geant4 dans un environnement de grille. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II; 2011.
73. ICRU Report 40–The quality factor in radiation protection. *International Commission on Radiation Units and Measurements*; 1986.
74. ICRU Report 36-Microdosimetry. *International Commission on Radiation Units and Measurements*; 1983.
75. Amols H, Lagueux B, Cagna D. Radiobiological effectiveness (RBE) of megavoltage x-ray and electron beams in radiotherapy. *Radiat Res*. 1986;105:58–67.
76. Lindborg L, Hultqvist M, Tedgren ÅC, Nikjoo H. Lineal energy and radiation quality in radiation therapy: model calculations and comparison with experiment. *Phys Med Biol*. 2013;58:3089.
77. Incerti S, Kyriakou I, Bernal MA, et al. Geant4-DNA example applications for track structure simulations in liquid water: a report from the Geant4-DNA project. *Med Phys*. 2018b;45:e722–e739.
78. Bernal MA, Bordage MC, Brown J, et al. Track structure modeling in liquid water: a review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. *Phys Medica*. 2015;31:861–874.
79. Incerti S, Ivanchenko A, Karamitros M, et al. Comparison of Geant4 very low energy cross section models with experimental data in water. *Med Phys*. 2010a;37:4692–4708.
80. Incerti S, Baldacchino G, Bernal M, et al. The Geant4-DNA project. *Int J Model Simulat Scient Comput*. 2010b;1:157–178.
81. Lazarakis P, Incerti S, Ivanchenko V, et al. Investigation of track structure and condensed history physics models for applications in radiation dosimetry on a micro and nano scale in Geant4. *Biomed Phys Eng Express*. 2018;4:024001.
82. Kyriakou I, Ivanchenko V, Sakata D, et al. Influence of track structure and condensed history physics models of Geant4 to nanoscale electron transport in liquid water. *Phys Medica*. 2019;58:149–154.
83. Granero D, Pérez-Calatayud J, Casal E, Ballester F, Venselaar J. A dosimetric study on the high dose rate flexisource. *Med Phys*. 2006;33:4578–4582.
84. Perry N, Broeders M, De Wolf C, et al. European guidelines for quality assurance in breast cancer screening and diagnosis fourth edition. *Luxembourg: Office for Official Publications of the European Communities*; 2006.
85. Dance DR, Sechopoulos I. Dosimetry in x-ray-based breast imaging. *Phys Med Biol*. 2016;61:R271.
86. Castelli E, Tonutti M, Arfelli F, et al. Mammography with synchrotron radiation: first clinical experience with phase-detection technique. *Radiology*. 2011;259:684–694.
87. Sempau J, Sanchez-Reyes A, Salvat F, ben Tahar HO, Jiang S, Fernández-Varea J. Monte carlo simulation of electron beams from an accelerator head using penelope. *Phys Med Biol*. 2001;46:1163.
88. Fedon C, Longo F, Mettivier G, Longo R. Geant4 for breast dosimetry: parameters optimization study. *Phys Med Biol*. 2015;60:N311.
89. Paganetti H. Nuclear interactions in proton therapy: dose and relative biological effect distributions originating from primary and secondary particles. *Phys Med Biol*. 2002;47:747.
90. Tommasino F, Durante M. Proton radiobiology. *Cancers*. 2015;7:353–381.
91. Böhlen TT, Cerutti F, Dosanjh M, et al. Benchmarking nuclear models of Fluka and Geant4 for carbon ion therapy. *Phys Med Biol*. 2010;55:5833.
92. Folger G, Ivanchenko V, Wellisch J. The binary cascade. *Eur Phys J A-Hadrons Nuclei*. 2004;21:407–417.
93. Koi T. New native QMD code in Geant4. In: *Proceedings of the MC2010 Monte Carlo Conference*, vol. 5, p. 7; 2010.
94. Boudard A, Cugnon J, David J-C, Leray S, Mancusi D. New potentialities of the liege intranuclear cascade model for reactions induced by nucleons and light charged particles. *Phys Rev C*. 2013;87:014606.
95. Mancusi D, Boudard A, Cugnon J, David J-C, Kaitaniemi P, Leray S. Extension of the liège intranuclear-cascade model to reactions induced by light nuclei. *Phys Rev C*. 2014;90:054602.
96. Yarba J. Recent developments and validation of Geant4 hadronic physics. *J Phys*. 2012;396:022060.
97. Zerkin V, Pritychenko B. The experimental nuclear reaction data (EXFOR): extended computer database and web retrieval system. *Nucl Instrum Methods Phys Res Sect A*. 2018;888:31–43.
98. Kopeliovich B. Transparent nuclei and deuteron-gold collisions at relativistic energies. *Phys Rev C*. 2003;68:044906.
99. Fesefeldt H. The simulation of hadronic showers. Physics and Application, Technical Report PITHA, pp. 85–02; 1985.
100. Grichine V. A simple model for integral hadron–nucleus and nucleus–-nucleus cross-sections. *Nucl Instrum Methods Phys Res Sect B*. 2009;267:2460–2462.

101. Dyer P, Bodansky D, Seamster A, Norman EB, Maxson D. Cross sections relevant to gamma-ray astronomy: proton induced reactions. *Phys Rev C*. 1981;23:1865.
102. Davids CN, Bonner T. Enhancement of the 3^{4} he->^{12} c reaction rate by inelastic proton scattering. *Astrophys J*. 1971;166:405.
103. De Napoli M, Agodi C, Battistoni G, et al. Carbon fragmentation measurements and validation of the Geant4 nuclear reaction models for hadrontherapy. *Phys Med Biol* 2012;57:7651.
104. Colonna M, Di Toro M, Guarnera A. Mean field instabilities in dissipative heavy ion collisions. *Nucl Phys A*. 1995;589:160–174.
105. Di Toro M, Olmi A, Roy R. Neck dynamics. In: Chomaz P, Gulminelli F, Trautmann W, Yennello SJ, eds. *Dynamics and Thermodynamics with Nuclear Degrees of Freedom*. Berlin, Heidelberg: Springer; 2006.
106. Mancini-Terracciano C, Asai M, Caccia B, et al. Preliminary results coupling "stochastic mean field and Boltzmann-Langevin one body models with Geant4. *Phys Medica*. 2019;67:116–122.
107. Ciardiello A, Asai M, Caccia B, et al. Preliminary results in using deep learning to emulate BLOB, a nuclear interaction model. *Phys Medica*. 2020;73:65–72.
108. Kossov M. Chiral-invariant phase space model. *Eur Phys J A-Hadrons Nuclei*. 2002;14:265–269.
109. Grichine V. Geant4 hadron elastic diffuse model. *Comput Phys Commun*. 2010;181:921–927.
110. Wright D, Kelsey M. The Geant4 Bertini cascade. *Nucl Instrum Methods Phys Res Sect A*. 2015;804:175–188.
111. Seltzer S, Fernández-Varea J, Andreo P, et al. Key data for ionizing-radiation dosimetry: measurement standards and applications, ICRU report 90; 2016.
112. Bortfeld T. An analytical approximation of the Bragg curve for therapeutic proton beams. *Med Phys*. 1997;24:2024–2033.
113. Testa M, Schümann J, Lu H-M, et al. Experimental validation of the TOPAS Monte Carlo system for passive scattering proton therapy. *Med Phys*. 2013;40:121719.
114. Schardt D, Steidl P, Krämer M, Weber U, Parodi K, Brons S. Precision Bragg-curve measurements for light-ion beams in water. *GSI Sci Report 2007*. 2008;2008-1:373–373.
115. Bichsel H, Hiraoka T, Omata K. Aspects of fast-ion dosimetry. *Radiat Res*. 2000;153:208–219.
116. Lechner A, Ivanchenko V, Knobloch J. Validation of recent Geant4 physics models for application in carbon ion therapy. *Nucl Instrum Methods Phys Res Sect B*. 2010;268:2343–2354.
117. Meier M, Clark D, Goulding C, et al. Differential neutron production cross sections and neutron yields from stopping-length targets for 113-mev protons. *Nucl Sci Eng*. 1989;102:310–321.
118. Meier M, Goulding C, Morgan G, Ullmann J. Neutron yields from stopping-and near-stopping-length targets for 256-mev protons. *Nucl Sci Eng*. 1990;104:339–363.
119. Satoh D, Kajimoto T, Shigyo N, et al. Distributions of neutron yields and doses around a water phantom bombarded with 290-mev/nucleon and 430-mev/nucleon carbon ions. *Nucl Instrum Methods Phys Res Sect B*. 2016;387:10–19.
120. Haettner E, Iwase H, Krämer M, Kraft G, Schardt D. Experimental study of nuclear fragmentation of 200 and 400 mev/u 12c ions in water for applications in particle therapy. *Phys Med Biol*. 2013;58:8265.
121. Chaudhary P, Marshall TI, Perozziello FM, et al. Relative biological effectiveness variation along monoenergetic and modulated Bragg peaks of a 62-mev therapeutic proton beam: a preclinical assessment. *Int J Radiat Oncol Biol Phys*. 2014;90:27–35.
122. Elsässer T, Weyrather WK, Friedrich T, et al. Quantification of the relative biological effectiveness for ion beam radiotherapy: direct experimental comparison of proton and carbon ion beams and a novel approach for treatment planning. *Int J Radiat Oncol Biol Phys*. 2010;78:1177–1183.
123. Friedrich T, Scholz U, Elsässer T, Durante M, Scholz M. Calculation of the biological effects of ion beams based on the microscopic spatial damage distribution pattern. *Int J Radiat Biol*. 2012;88:103–107.
124. Cortés-Giraldo M, Carabe A. A critical study of different Monte Carlo scoring methods of dose average linear-energy-transfer maps calculated in voxelized geometries irradiated with clinical proton beams. *Phys Med Biol*. 2015;60:2645.
125. Cirrone GA, Cuttone G, Raffaele L, et al. Clinical and research activities at the CATANA facility of INFN-LNS: from the conventional hadrontherapy to the laser-driven approach. *Front Oncol*. 2017;7:223.
126. Jia SB, Romano F, Cirrone GA, et al. Designing a range modulator wheel to spread-out the Bragg peak for a passive proton therapy facility. *Nucl Instrum Methods Phys Res Sect*. 2016;806:101–108.
127. Abolfath R, Peeler CR, Newpower M, Bronk L, Grosshans D, Mohan R. A model for relative biological effectiveness of therapeutic proton beams based on a global fit of cell survival data. *Sci Rep*. 2017;7:8340.