Published in final edited form as:
Med Phys. 2022 October ; 49(10): 6666–6683. doi:10.1002/mp.15913.

Virtual Particle Monte Carlo (VPMC), a new concept to avoid simulating secondary particles in proton therapy dose calculation

Jie Shan, MS[1,*], **Hongying Feng, PhD**[1,*], **Danairis Hernandez Morales, MS**[1], **Samir H. Patel, MD**[1], **William W. Wong, MD**[1], **Mirek Fatyga, PhD**[1], **Martin Bues, PhD**[1], **Steven E. Schild, MD**[1], **Robert L. Foote, MD**[2], **Wei Liu, PhD**[1]

[1]Department of Radiation Oncology, Mayo Clinic, Phoenix, AZ 85054, USA

[2]Department of Radiation Oncology, Mayo Clinic, Rochester, MN, 55902, USA

Abstract

Background: In proton therapy dose calculation, Monte Carlo (MC) simulations are superior in accuracy but more time consuming, compared to analytical calculations. Graphic Processing Units (GPUs) are effective in accelerating MC simulations but may suffer thread divergence and racing condition in GPU threads that degrades the computing performance, due to the generation of secondary particles during nuclear reactions.

Purpose: A novel concept of virtual particle (VP) MC (VPMC) is proposed to avoid simulating secondary particles in GPU-accelerated proton MC dose calculation and take full advantage of the computing power of GPU.

Methods: Neutrons and gamma rays were ignored as escaping from the human body; doses of electrons, heavy ions, and nuclear fragments were locally deposited; the tracks of deuterons were converted into tracks of protons. These particles, together with primary and secondary protons, are considered to be the realistic particles. Histories of primary and secondary protons were replaced by histories of multiple VPs. Each VP corresponded to one proton (either primary or secondary). A continuous-slowing-down-approximation (CSDA) model, an ionization model, and a large angle scattering event (LAE) model corresponding to nuclear interactions were developed for VPs by generating probability distribution functions (PDFs) based on simulation results of realistic particles using MCsquare. For efficient calculations, these PDFs were stored in the Compute Unified Device Architecture (CUDA) textures. VPMC was benchmarked with TOPAS and MCsquare in phantoms and with MCsquare in thirteen representative patient geometries. Comparisons between the VPMC calculated dose and dose measured in water during

Corresponding author: Wei Liu, PhD, Professor of Radiation Oncology, Department of Radiation Oncology, Mayo Clinic Arizona, 5777 E. Mayo Boulevard, Phoenix, AZ 85054; Liu.Wei@mayo.edu.

[*]Co-first authors who contributed to this work equally

Conflicts of Interest Notification
The authors have no conflicts to disclose.

Ethical considerations
This research was approved by the Mayo Clinic Arizona institutional review board (IRB, 13–005709). The informed consent was waived by IRB protocol. Only image and dose-volume data were used in this study. All patient-related health information was removed prior to the analysis and also publication of the study.

patient-specific quality assurance (PSQA) of the selected 13 patients were also carried out. Gamma analysis was used to compare the doses derived from different methods and calculation efficiencies were also compared.

Results: Integrated-depth dose and lateral-dose profiles in both homogeneous and inhomogeneous phantoms all matched well among VPMC, TOPAS, and MCsquare calculations. The 3D-3D Gamma passing rates with a criterion of 2%/2mm and a threshold of 10% was 98.49% between MCsquare and TOPAS, and 98.31% between VPMC and TOPAS in homogeneous phantoms, and 99.18% between MCsquare and TOPAS and 98.49% between VPMC and TOPAS in inhomogeneous phantoms, respectively. In patient geometries, the 3D-3D Gamma passing rates with 2%/2mm/10% between dose distributions from VPMC and MCsquare were 98.56±1.09% in patient geometries. The 2D-3D Gamma analysis with 3%/2mm/10% between the VPMC calculated dose distributions and the 2D measured planar dose distributions during PSQA was 98.91±0.88%. VPMC calculation was highly efficient and took 2.84±2.44 seconds to finish for the selected 13 patients running on four NVIDIA Ampere GPUs in patient geometries.

Conclusion: VPMC was found to achieve high accuracy and efficiency in proton therapy dose calculation.

Keywords

Monte Carlo; intensity-modulated proton therapy; secondary particles; GPU acceleration; Real time adaptive treatment planning

Introduction

Pencil beam scanning proton therapy (PBS) is the most advanced form of proton therapy. It has distinct advantages of high conformality of target coverage and superior protection of nearby organs-at-risk (OARs) when compared with the photon-based therapy and passive scattering proton therapy, which was brought by its exceedingly high flexibility at the beamlet level in treatment planning and dose delivery[1–4]. A sufficiently accurate dose calculation engine is the foundation for plan optimization and evaluation in PBS to achieve “what you see is what you get”. Meanwhile, a large amount of calculation is required due to the large number of voxels and the large number of beamlets with multiple perturbed scenarios considered in robust optimization[5–32]. Thus, the dose calculation engine should be fast enough and meet the efficiency requirement in clinical practices of robust PBS treatment planning.

Analytical dose engines[33–35] are widely used in the commercial and in-house treatment planning systems (TPSs), which are fast, but with less accuracy. On the other hand, Monte Carlo (MC) simulations are considered to be the gold standard for dose calculation in proton therapy. Clinically significant differences between the dose distributions from analytical and MC dose calculation engines have been reported, especially in heterogeneous geometries such as the lung and head and neck (H&N) sites[36,37] with bones, air cavities, and dental implants, *etc*. The resulting inaccurate dose distributions might mislead healthcare practitioners and cause unexpected adverse events (AEs) and local failure in cancer patients treated with PBS.

General-purpose MC codes (e.g. Geant4[38], MCNPX[39], FLUKA[40,41], and TOPAS[42]) have been proven accurate, but are time-consuming (several hours or days for a single plan dose calculation), and are therefore considered inefficient for clinical use. Whilst fast MC codes (e.g. track-repeating[43], macro Monte Carlo[44], gMC[45], gPMC[46], FRED[47], and MCsquare[48]), dedicated to proton dose calculations with simplified physics models and/or graphic processing unit (GPU) acceleration, have significantly reduced the time for one plan dose calculation to several minutes. Recently, some fast MC software have been released for clinical use in mainstream commercial treatment planning systems (TPSs), such as Eclipse™ (Varian Medical Systems, Palo Alto, CA)[49,50] and RayStation™ (RaySearch, Stockholm, Sweden)[50,51].

Nonetheless, the aforementioned fast MC codes are not yet fast enough for advanced optimizations in PBS treatment planning, such as 4D robust optimization[22,52,53] and linear-energy-transfer (LET)-guided robust optimization[13,15,54,55]. In a 4D robust optimization, tens or even hundreds of perturbed scenarios need to be considered due to the additional respiratory phases, while in the LET-guided robust optimization many independent influence matrices related to LET are required. Therefore, it is crucial to reduce the MC-based proton dose calculation time to seconds or even sub-seconds for real-time and adaptive treatment planning and advanced optimizations. Recently, RaySearch reported that its latest fast MC-based proton dose engine could finish plan dose calculation with a median time of 5.2 seconds and a median calculation speed of 8.4 million particles per second.[56] Although still far from ideal, this may be the first MC-based proton dose calculation engine that enables real time and adaptive PBS treatment planning.

All the current fast MC codes use physics models to handle the interactions between particles and mediums. The generated secondary particles are stored in the shared memory buffers. MC simulation must decide the probability, type, and momentum of the generation of various kinds of secondary particles and then access the shared memory buffer accordingly at each MC step. Therefore, sophisticated logic controls and large shared memory buffers are required. Fortunately, this can be effectively handled by central processing units (CPUs), such as in MCsquare[48].

However, unlike CPUs, GPUs have very simple control logic and very limited shared memory buffer for a single thread. Within a block of threads, the threads executed in a group of 32 is called a warp. Thread divergence occurs when different threads in a warp need to execute different tasks. In the worst-case scenario, if one thread in a warp diverges, the 32 threads in this warp might need to stop and revert to the end of the previous step to redo the dose calculation. Therefore, GPU’s computing performance will be degraded and could be 32 times slower in the worst case. Thread divergence is a significant problem in GPU-based proton dose calculation engines if the “one particle per thread” technique (a GPU thread tracks a particle until the end conditions are satisfied) is adopted in the design of parallelization[45].

Racing condition is another significant problem in large scale shared memory parallel computation, where multiple threads happen to access and manipulate the same shared memory address simultaneously, resulting in possible memory conflicts among different threads. The generation of secondary particles is random and leads to sophisticated computation logic and unexpected computation burdens to the corresponding thread, and the randomly increased number of particles per thread due to the generation of secondary particles greatly increases the chance of racing condition. These two problems are inherent to the hardware architecture of GPUs and prevent taking full advantage of the GPU computing power and achieving MC-based sub-second plan dose calculation based on GPUs. All GPU-accelerated fast MC-based proton dose calculation engines (GPU version of track-repeating[57], gMC, gPMC, and FRED) had tried to mitigate the aforementioned thread divergence and racing condition problems.

In this paper, we propose a fast MC-based proton dose calculation engine based on a novel concept of virtual particles (VPs). VPs inherit essential physical properties from realistic particles but are conceptually designed for parallel computing in GPUs by avoiding the simulation of secondary particles. Therefore, simulation of VPs instead of realistic particles (primary protons and secondary particles) can take full advantage of the unique hardware architectures of GPUs (many simultaneous threads but simple control logics and limited shared memory buffers for a single thread), leading to greatly enhanced usage efficiency of the computing power of GPUs.

Methods and Materials

A. Implementation of VP

A.1 Converting realistic particles to VPs—Figure 1 shows the process of converting the histories of realistic particles into equivalent histories of VPs. Three simplifications were made to further speed up the calculations and still meet the clinical accuracy requirements: (1) the doses of electrons, heavy ions, and nuclear fragments were locally deposited (from Fig. 1(a) to Fig. 1(b)), (2) neutrons and gamma rays were considered to escape from the human body and thus ignored as energy loss, (from Fig. 1(b) to Fig. 1(c)) (3) the tracks of deuterons were converted into tracks of protons since deuterons have a low generating possibility in proton dose calculation of clinical proton therapy energies (0~230 MeV) and contribute less than 1% to the final dose[58] while having a range less than but close to protons. After these simplifications, one primary proton could be represented by multiple VPs that started at the same starting position of the primary proton (from Fig. 1(c) to Fig. 1(d)). Assuming the number of secondary protons and deuterons generated from one primary proton in the MC simulation is M, the number of VPs starting at the same starting position of the primary proton would be M+1 for this primary proton.

Please note that VP is a statistical concept. Every VP evolves independently based on the pre-generated corresponding possibility distribution functions (PDFs) of the multiple Coulomb scattering, ionization, and nuclear physics models (please see the following subsections for details). Given a large number of VPs, the simulation results following the pre-generated PDFs will converge to the simulation results of a large number of realistic particles (including secondary particles and the final dose), since the PDFs are generated based on the simulation results of realistic particles. Two VPs were used in Figure 1 only for the demonstration purpose. In Figure 1, we showed in a very simplified and unrealistic way how we could use two VPs to get the equivalent dose distribution from one primary proton and one secondary proton. The path before two VPs to diverge at R2 shown in Figure 1(d) is random and is determined by the calculated possibility of the nuclear reaction model discussed in the following subsection.

After converting the histories of realistic particles into histories of VPs that generate the same dose as realistic particles, statistic models consisting of probability distribution functions (PDFs) of the corresponding parameters were obtained to describe the behaviors of VPs. The models were derived by performing a statistics study of the track histories of realistic protons and their corresponding secondary particles from a conventional MC simulation (please see the rest of Sec. A for details). The open-source fast MC code, MCsquare[35,59–62] has been thoroughly validated against other MC codes and measurements in both phantoms and patient geometries[59,60,62–68]. In addition, MCsquare has been fully commissioned and has been incorporated into our in-house treatment planning system, Shiva[52,69–73], and has been clinically used as the second monitor unit (MU) check system at our proton center for years[60,62] and other proton centers[74]. Therefore, we chose MCsquare to generate all the PDFs used in the VPMC dose engine. To generate the PDFs of physics parameters, 20 million primary monoenergetic protons with infinitesimal spot sizes were used to irradiate normally into a water phantom of 20.1×20.1×40cm, so that all the generated PDFs could reach a statistical uncertainty (please see subsection D. Statistical uncertainty and efficiency for details) as low as 0.093% to guarantee that VPMC calculated dose based on those PDFs could reach a sufficiently low statistical uncertainty as well.

As for how the PDF databases of the physics parameters were generated and how HU-to-materials and HU-to-density conversions were done, please refer to Sec. A and B in Supplementary Materials for details.

A.2 Multiple Coulomb scattering—A continuing-slowing-down-approximation (CSDA) model (condensed class-II) was used to describe the Multiple coulomb scattering (MCS, i.e., soft electromagnetic interactions) for VPs. Rather than an on-the-fly calculation of the related physics parameters (the deposited energy, energy straggling, the deflection angle, and the ionization probability) based on physics models in MCsquare (and other MC codes), VPMC directly calculated such physics parameters using the 2D Compute Unified Device Architecture (CUDA) textures, i.e., look up tables (LUTs), with 2 dimension self-variables as the energy of VPs (E) and the Hounsfield Units (HU) value (HU) of the VP location for each MC step, which greatly sped up the calculation[75].

To generate such CUDA textures, the energy and HU value were scaled (0 to 230 MeV for energy and −1050 to 29000 for HU values, respectively) in MCsquare, and the calculated physics parameters were recorded and used to obtain PDFs describing the relationship between physics parameters and the energy and HU value, i.e., X(E, HU), where X indicates either of the deposited energy, energy straggling, the deflection angle, or the ionization probability. Since X(E, HU) was not linear, further tuning (please refer to Sec. A in Supplementary Materials for details) work was done to render VPMC achieve the same dose accuracy as MCsquare.

A.3 Ionization—Based on the ionization cross section interpolated from the CUDA texture generated in the CSDA step, the ionization for VPs was modelled in VPMC. If ionization happened, the status of VPs was updated based on the same analytical formula used in MCsquare.

A.4 Nuclear reaction—A large angle event (LAE) model was used to handle the nuclear reaction. During the nuclear reactions, possible secondary particles were generated, and resulted in an increased number of realistic particles. In conventional MC simulations, such secondary particles would randomly appear in the middle of the calculation (Fig. 1(a)) and need to be tracked thereafter as primary particles, which caused the possible thread divergence and higher chance of racing condition in GPU threads. However, in VPMC, these realistic particles (including both primary and secondary particles) were replaced by VPs, which appeared from the very beginning of the MC simulation (Fig. 1(d)).

Therefore, the total number of VPs was pre-determined (please refer to Sec. C in Supplementary Materials for details) before the MC simulation and did not change during the MC simulation. Thus, besides the conventional physics parameters of the nuclear reaction probability, the deposited energy, the energy loss (energies from neutrons and gamma rays escaped from human bodies), and the deflection angle, we introduced another parameter called “weight gain”. Weight gain is the fraction of the dose contributed by one certain VP at a certain step. This parameter is unique to VPs and does not exist in the MC simulation of realistic particles. During the dose scoring stage of the MC simulation, the weight gain would be multiplied to the dose contribution of the corresponding VP within the corresponding step to get the correct dose distribution as generated by realistic particles. Taking Figure 1(d) as an example, the first two steps of the MC simulation before the bifurcation at R2 were shared by two VPs, thus these two VPs in these two steps would be assigned a weight gain of 0.5 for the dose scoring. While after the bifurcation at R2, both VPs (A and B) exclusively possessed their own third step, therefore, the weight gain would be changed to 1.0 for both VPs for the corresponding third step during the dose scoring stage.

Similar to the CSDA model, the LAE model was not calculated on-the-fly either but was derived using all the related parameters (i.e., the nuclear reaction probability, the deposited energy, the energy loss, the deflected angle, and the weight gain) obtained based on the pre-generated database using MCsquare[48,60] with 20 million monoenergetic primary protons with infinitesimal spot sizes irradiating normally into phantoms stored in the CUDA textures for efficient calculation.

A.5 Range shifter—Two institution-specific range shifters, range shifter positioned at the exit of nozzle (labeled as RS) and extended range shifter positioned at an extended position (labelled as ERS), were also included in the VPMC dose engine. The range shifters were made of ABS Resin composed of hydrogen, carbon, nitrogen, and oxygen. The water equivalent thickness was 4.5 cm for both range shifters. RS was positioned at 42.5 cm from isocenter, while ERS was positioned 30 cm from isocenter. Similar to the track histories of VPs in patient geometries, the track histories of VPs in the range shifter were also modeled by various PDFs. However, unlike the highly heterogenous patient geometries, the range shifter was a homogenous medium with a fixed HU value. Therefore, the 2D CUDA textures for the CSDA model in patient geometries was reduced to a 1D CUDA textures with the fixed HU value corresponding to the range shifter material.

B. Workflow of VPMC

Figure 2 shows the workflow of VPMC. At the very beginning, all the pre-calculated databases of the CSDA model in range shifter (energy deposit, energy straggling, deflection angle, and ionization probability), the CSDA model in patient geometry (energy deposit, energy straggling, deflection angle, and ionization probability), and the LAE model (energy deposit, energy loss, deflection angle, weight gain, and nuclear reaction probability) were loaded. Once loaded, such databases were stored in shared memories as CUDA textures for repeated and efficient access from the subsequent VP simulations. For the VP simulations, a number of VPs were sampled at first with parameters of energy, position, and momentum and each VP was assigned to a certain GPU thread.

As for the model used for particle sampling, we commissioned VPMC to get a single set of phase space parameters suitable for all three machines (one without range shifter, one with RS, and the other with ERS). In order to speed up the calculation we did not simulate the particle transport in the nozzle.[76,77] Instead, we derived the phase space at the exit of the nozzle (but before any beam modifiers such as range shifters) using integrated depth dose (IDD) curves and in-air lateral profiles at five positions of proton beams. Hence, the phase space has a large emittance due to the scattering of the beamline components in the nozzle and we chose to use double Gaussian lateral profile to model the beam source more accurately[48,60]. For our synchrotron-based system we commissioned 97 discrete energies, rather than a selected number of energies as are typical for a cyclotron-based system.[46] The MU calibration curve and the corresponding CT calibration curve were also commissioned accordingly as in MCsquare[48,60].

If a range shifter (either RS or ERS) was used, VPs were first simulated in the range shifter (green box in Fig. 2), and then simulated in the patient geometry (blue box in Fig. 2). In the range shifter, by comparing the ionization probability (P_{ion}) calculated based on the VP parameters and the sampled ionization probability (sP_{ion}), the ionization process was considered based on the same analytical formula as used in MCsquare if needed ($sP_{ion} < P_{ion}$). Then the nuclear reaction probability (P_{LAE}) calculated based on the LAE model was compared to the sampled nuclear reaction probability (sP_{LAE}). If a nuclear reaction happened ($sP_{LAE} < P_{LAE}$), the status of VPs would be updated based on the LAE model. This stage ended when VPs exited the range shifter.

The calculation in the patient geometry was similar to the calculation in the range shifter with some minor differences: (1) the CSDA model in patient geometry was used instead of the CSDA model in range shifter, (2) the CT DICOM coordinate was used instead of the beam eye view coordinate, (3) dose scoring was done to each voxel, which was globally shared among all threads in memory buffers, in the patient geometry, while no dose scoring was done in the calculation in the range shifter. The calculation in the patient geometry ended when VPs were absorbed in the patient geometry or exited the patient geometry. The whole simulation ended when the calculation in the patient geometry ended.

The dose scoring to each voxel from each VP simulated by a certain thread took place once the VP status was updated. Atomic addition function[46] was used to guarantee "thread-safe", i.e., operations of the per-voxel dose, which was globally shared among all threads in memory buffers, by one thread was not interfered by other threads. This would avoid possible racing conditions with some compromise of the computing performance. The avoidance of the simulation of the secondary particles would further greatly minimize the chance of the possible racing conditions in VPMC.

C. Validation

The validation was first done in homogeneous/inhomogeneous phantoms and then in patient geometries. For phantoms, a proton beam with a nominal energy of 228.8 MeV and a weight of 1 MU was normally irradiated to the phantom surface. Two phantoms were used: (1) a homogeneous water phantom and (2) an inhomogeneous phantom with a cube of HU=1000 in the middle of water blocking part of the beamlet path (Fig. 3). Dose distributions for each phantom were generated from three different dose engines, respectively: (1) VPMC, (2) MCsquare, and (3) TOPAS (ver. 3.7)[78,79], which is a standard MC code. TOPAS simulation results were considered as golden standard. 3D-3D Gamma analysis[80,81] between MCsquare and TOPAS, and between VPMC and TOPAS was used to assess the agreement between the dose distributions in these phantoms, respectively. The size and resolution of the phantoms used for benchmark were 400×201×201, and 1.0 mm×1.0 mm×1.0 mm, respectively. 1×10^7 primary protons were simulated in MCsquare, while 2×10^7 VPs were simulated in VPMC to guarantee low statistical uncertainty (<0.2%, Table S-1). More VPs in VPMC than primary protons in MCsquare were used because primary protons would generate secondary particles, thus resulting in more realistic particles than the number of the initial primary protons in a MCsquare simulation.

For patient geometries, clinically approved PBS plans for thirteen patients with different disease sites and a wide range of total spot number were selected to representatively cover most clinical scenarios. Range shifter was used in 6 plans. Detailed characteristics of the selected plans were included in Table 1. Three dose distributions were generated for each plan from three different dose engines, respectively: (1) VPMC, (2) MCsquare, and (3) the analytical dose engine in our commercial treatment planning system, Eclipse™ ver. 15.6 (Varian Medical Systems, Palo Alto, CA). 3D-3D Gamma analysis between VPMC and MCsquare and between Eclipse™ and MCsquare were performed, respectively. The same dose grid resolution of 2.5 mm in all three directions was used for all dose calculations included in the study.

Comparisons between the VPMC-calculated in-water dose and the measured in-water dose during patient-specific quality assurance (PSQA) and comparisons between the Eclipse™-calculated in-water dose and the measured in-water dose during PSQA for the 13 selected plans were also carried out. The PSQA procedure at our institution[81] was done by delivering treatment plans to a water tank measured with a 2D MatrixxPT ionization chamber array (IBA Dosimetry GmbH, Schwarzenbruck, Germany). For each field of the delivered plan, 2D plane doses at two or three representative axial depths were measured. For each measured plane, 1,020 detection points were included. The measured plane doses were then compared to the calculated 3D dose distribution in water from different dose engines. 2D (measurements)-3D (calculated dose) Gamma analysis[81,82] was used for the comparison. The detector, MatriXX PT, is used with a resolution of 7.6 mm * 7.6 mm[81].

American Association of Physicists in Medicine (AAPM) Task Group 218 recommended a criterion of 3%/2mm and a threshold of 10% for the Gamma analysis done in PSQA for intensity-modulated radiation therapy (IMRT)[83]. Thus, in the 2D-3D Gamma analysis between the 3D VPMC calculation and the in-water 2D planar measurements in PSQA, the recommended criterion of 3%/2mm with 10% threshold was used, with the in-water measured 2D planar dose as references. However, one may have possible concerns of the inflated Gamma passing rates derived from comparing the two MC-calculated dose distributions due to the statistical noises inherent in MC-calculated dose distributions. Therefore, in the 3D-3D Gamma analysis in both phantoms and patient geometries, a more stringent criterion of 2%/2mm with a threshold of 10% was used. For the 3D-3D Gamma analysis in both homogeneous and inhomogeneous phantoms, the TOPAS simulation results were chosen as the reference doses, while for the 3D-3D Gamma analysis in patient geometries where there were no TOPAS simulations, simulation results from the comprehensively-benchmarked MCsquare were chosen as the reference doses. For all Gamma analysis, trilinear interpolation was performed through CUDA textures on the evaluation grid. A global criterion was applied.

D. Statistical uncertainty and efficiency

The statistical uncertainty of a certain MC simulation in terms of a chosen number of simulated particles, was calculated using the equation[84]:

$$\sigma = \frac{1}{N_{d_i > 20\% D_{max}}} \sum_{d_i > 20\% D_{max}} \frac{\sigma_i}{D_{max}^{i}}$$

where $N_{d_i > 20\% D_{max}}$ is the number of voxels, whose dose is larger than the 20% of the maximum dose D_{max} of the whole dose distribution, σ_i is the statistical uncertainty for voxel i and calculated by the equation:

$$\sigma_i = \sqrt{\frac{\overline{d_\imath^2} - \overline{d_\imath}^2}{N_{run} - 1}}$$

in which $\overline{d_\imath^2}$ is the average square dose for voxel i among a number of repeated simulations, and $\overline{d_\imath}^2$ is the square of average dose for voxel i among a number of repeated simulations. N_{run} is the number of repeated simulations. D_{max}^{i} is the maximum dose among the N_{run} simulations for voxel i. In this study, N_{run} was set to 10 for each simulation scenario, and the number of particles to be simulated in both phantoms and patient geometries was chosen carefully to guarantee that a low statistical uncertainty (< 1%) was achieved in all simulations.

Due to the fundamental difference between the realistic particles tracked in MCsquare and VPs tracked in VPMC, to fairly compare the efficiency between MCsquare and VPMC, the index $e = 1/(\sigma^2 t)$ was used to compare efficiency between MCsquare and VPMC[84]. σ is the statistical uncertainty as defined above, while t is the calculation time for one simulation.

Results

A. Validations in phantoms

Figure 4 shows comparison among dose distributions calculated by TOPAS (magenta), MCsquare (blue), and VPMC (orange). Subpanels in the top row are results calculated in the homogeneous phantom, while subpanels in the bottom row are results calculated in the inhomogeneous phantom. From left to right, the four columns are IDD curves, MCsquare and VPMC IDD differences to TOPAS, log-scale lateral dose profile at the Bragg Peak, and MCsquare and VPMC lateral dose profile differences to TOPAS at the Bragg Peak, respectively. For both homogeneous and inhomogeneous phantoms, excellent agreement was observed among the dose distributions generated by TOPAS, MCsquare, and VPMC. The 3D-3D Gamma passing rates between MCsquare and TOPAS and between VPMC and TOPAS were 98.49% and 98.31% in the homogeneous phantom and 99.18% and 98.49% in the inhomogeneous phantom, respectively.

B. Validations in patient geometries

The validation results in patient geometries were shown in Table 2. The mean 3D-3D Gamma passing rate between VMPC and MCsquare was 98.58±1.09%, while the 3D-3D Gamma passing rate between Eclipse™ and MCsquare was 89.60±7.24% with 2%/2mm/10%. VPMC drastically reduced the dose calculation time for a plan to 2.84±2.33 seconds, compared to 112.56±114.38 seconds for MCsquare, running on AMD EPYC™ 7543 equipped with 4 NVIDIA Ampere A100 GPUs. Patient 12 took the longest dose calculation time of 8.9 seconds for VMPC. This patient had the largest number of spots (67,875) and the largest number of voxels (5,495,896). The average statistical uncertainty of MCsquare was 0.27±0.08%, while the average statistical uncertainty of VPMC was 0.57±0.23%. The average efficiency was $0.62 \pm 1.05 \times 10^4\ s^{-1}$ for MCsquare, and $5.37 \pm 9.96 \times 10^4\ s^{-1}$ for VPMC. VPMC is 8.66 times more efficient than MCsquare. VPMC calculations with more VPs achieved a statistical uncertainty (0.36±0.13%) closer to MCsquare and an enhanced efficiency ($7.86 \pm 15.85 \times 10^4\ s^{-1}$, which is now 12.5 times better than the average efficiency of MCsquare), at a cost of prolonged calculation time (5.41±4.94 seconds) (Table S-5). Three representative patients (one prostate patient without range shifter, one H&N patient with ERS, and one chest wall patient with RS) were selected to compare the dose profile on typical transverse planes between VPMC and MCsquare (Fig. 5–7). We found that VPMC agreed well with MCsquare in patient geometry. Please note that the MCsquare used in this comparison was the modified version, which had been successfully enhanced to be about 10 times as fast as the original MCsquare[60].

C. Comparisons with PSQA measurements

Table 3 shows the comparison of the 2D-3D gamma analysis passing rates, with a criterion of 3%/2mm and a threshold of 10%, between Eclipse™ and VPMC calculation results with the measured 2D plane dose during PSQA, respectively. The average passing rate for Eclipse™ was 98.87% ± 1.12%, while the average passing rate for VPMC was 98.91% ± 0.88%. 2D-3D Gamma passing rates at different measurement locations of each field for all patients included in this study comparing the Eclipse™ and VPMC calculation results with the 2D measured plane doses are reported in Table S-4. Figure 8–10 shows the comparisons of dose maps between Eclipse and VPMC calculation results with the measured 2D plane dose during PSQA in three selected typical patients (without range shifter, with RS, and with ERS). We found that the VPMC calculation results agreed well with the measured results and the Eclipse calculation results in PSQA.

Discussion

In this study, we report the successful development of a GPU-accelerated fast MC dose engine, based on the novel concept of VP to avoid the simulation of the secondary particles generated during nuclear reactions. Pre-calculated PDF databases of the needed parameters were generated using the well-established open-source fast MC dose engine, MCsquare[48,60]. The PDFs were stored as CUDA texture LUTs for efficient calculation in VMPC. After fine-tuning of parameters needed in VPMC, VPMC agreed well with MCsquare. Moreover, VPMC significantly reduced the dose calculation time of a plan to 2.84±2.33 seconds, which was about 40 times faster and 9 times more efficient than the enhanced version of MCsquare, which is about 10 times faster than the original MCsquare[60].

Pre-calculated parameter database is a commonly used technique to increase the calculation efficiency in MC simulation[43,85–87], by using database querying instead of on-the-fly calculation. Such technique can perform even better on GPUs than CPUs, since the possible interpolation can be done efficiently using the unique characteristics of the GPU textures. However, there is an inherent drawback for using pre-calculated parameter databases. Once the system is updated, for example with a new CT calibration curve[88], the pre-calculated parameter database must be updated accordingly. Fortunately, the corresponding pre-calculated parameter databases only need to be generated once and can then be stored as the CUDA texture LUTs for future calculation.

Compared to other CPU-based fast MC codes using the pre-calculated databases, the proposed VPMC had some unique features. In Macro MC (MMC)[85,86], the pre-calculated databases were generated based on macro blocks (such as slab, cylinder, or sphere). And realistic particles, including both primary and secondary particles, were simulated in the conventional way. In the track-repeating[43,87] algorithm, a database of proton trajectories, including secondary protons, was generated in water with discrete steps, where the step length, angles relative to the previous step, energy loss, and energy deposit were stored. Then the extrapolation from water to other materials was achieved by scaling the path length of each step and the angle between steps. When a tracked history was selected for a proton, the proton was then transported as if it followed this assigned track. By re-tracing the proton track from the database of the pre-calculated-histories, dose distribution could be calculated in heterogeneous mediums.

VPMC's pre-calculated database was similar to how MMC used the pre-calculated database, where the parameter PDFs of each step were stored. It is thus different from how the track-repeating algorithm used the pre-calculated database, where proton histories were stored. At each step, VPMC and MMC required sampling based on the parameter PDFs, while track-repeating needed scaling the corresponding step length and the corresponding angle. However, despite the similar use of the pre-calculated parameter databases, VPMC used realistic small steps between two events in MC instead of macro blocks (such as slab, cylinder, or sphere) in MMC. In addition, VPMC was specifically developed for GPU-acceleration and introduced a GPU-friendly concept of VP to avoid the simulation of the secondary particles generated during the nuclear reaction, while MMC was originally developed for CPU-based computing platforms with conventional methods to handle secondary particles. Therefore, if one attempts to adapt the MMC in the GPU-based computing platforms, the aforementioned thread divergence problem needs to be considered properly.

GPU has been widely used to accelerate calculations in MC-based proton dose calculation. To help mitigate the thread divergence problem in GPU threads, various efforts have been made. In gMC[45], two loops of particle simulation were used. In the first loop of primary particles simulation, the secondary particles generated in nonelastic nuclear interactions were grouped and stored for further processing. After the first loop of primary particles simulation was completed, the second loop of the stored secondary particles simulation would take place. The daughter secondary particles generated in the second loop would be handled accordingly as well. Such loops would repeat until all secondary particles were simulated. In gPMC[46], protons were simulated in batches. In each batch, a number (M) of protons that could be transported by GPU simultaneously would be generated and simulated. At the same time, a special stack was created to store secondary particles, which would be continuously monitored. When the number of stored secondary particles reached or surpassed the number of M, the stack would pop up M secondary particles to GPU to be simulated in the following batch. FRED[47] employed a similar idea to address the thread divergence as gMC and gPMC, in which secondary particles were queued for later simulation. In the GPU version of track-repeating[57], the same proton history was used within a block, but with different starting positions in the normal direction of the beam. Such technique was very GPU-friendly since each GPU thread essentially performed the same operations all the time.

VPMC used a novel concept of VPs to avoid the simulation of secondary particles generated in nuclear reactions to take full advantage of the computing capacity of GPUs. The concept of VPs theoretically avoided the possible thread divergence in GPU threads. Every VP is simulated equivalently with the same control logic and the same memory buffer, which is very friendly to the GPU hardware architecture. Therefore, by assigning the simulation of a VP to a GPU thread, every thread within a warp is equally executing the same computation task. Therefore, the thread divergence can be effectively avoided.

The concept of VP was originally developed from the perspective of statistics, rather than the perspective of physics models. Therefore, a descriptive LAE model was used to describe the nuclear reaction, where secondary particles could be generated, with five parameters of VP: nuclear reaction probability, reflected angle, energy loss, energy deposit, and weight gain. Weight gain was an arbitrary parameter specifically introduced to get the correct final dose during the scoring stage. And the PDF databases of those parameters were pre-calculated and stored using the CUDA texture LUTs for efficient calculation. The fast CPU-based MC code, MCsquare, was used for generating these databases by the statistical analysis of the pre-calculated 20 million primary particles in all materials defined during the commissioning of MCsquare based on our proton machine and CT simulator[60]. Theoretically, any fully-blown MC codes can be used for this purpose accordingly.

In the future, other auxiliary methods will be exploited to further accelerate the calculation speed and improve the calculation accuracy of VPMC. We will try to enhance the current CT resampling of a fixed resolution by introducing the more advanced adaptive resolution resampling method to reduce the number of the voxels, thus accelerating the calculation speed. We will also integrate the bias sampling method[89] in particle generation, in which the central high dose region of a beamlet will be down-sampled, while the lateral low dose region of a beamlet will be up-sampled to reduce the total number of particles for simulation to accelerate the calculation speed without compromising the calculation accuracy. Particle splitting[90] is another well recognized method in MC simulations to speed up calculations in which fewer numbers of particles are initially generated, but more particles are dynamically generated in the regions with high statistical noise. Essentially the total number of particles can be reduced without compromising the calculation efficiency, thus leading to faster calculation. We will also try the dynamic step size method, such as random-hinge[91,92] to enhance the calculation efficiency.

Conclusion

We have developed a GPU-friendly MC dose engine based on the novel concept of VP to avoid the simulation of secondary particles generated during the nuclear reactions. The concept of VP avoided the thread divergence problem caused by secondary particles in GPU threads to take full advantage of the computing power of GPUs. We found that the proposed VPMC can calculate proton dose distributions efficiently and accurately in PBS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Cancer Institute (NCI) Career Developmental Award K25CA168984, Arizona Biomedical Research Commission Investigator Award, the Lawrence W. and Marilyn W. Matteson Fund for Cancer Research, and the Kemper Marley Foundation.

Data availability statement

Data available on request from the authors.

Reference

1. van de Water TA, Bijl HP, Schilstra C, Pijls-Johannesma M, Langendijk JA. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: a systematic review of literature [published online ahead of print 2011/02/26]. Oncologist. 2011;16(3):366–377. [PubMed: 21349950]
2. Lin A, Swisher-McClure S, Millar LB, et al. Proton therapy for head and neck cancer: current applications and future directions. Translational Cancer Research. 2012;1(4):255–263.
3. Frank SJ, Cox JD, Gillin M, et al. Multifield optimization intensity modulated proton therapy for head and neck tumors: a translation to practice [published online ahead of print 2014/05/29]. International journal of radiation oncology, biology, physics. 2014;89(4):846–853. [PubMed: 24867532]
4. Schild SE, Rule WG, Ashman JB, et al. Proton beam therapy for locally advanced lung cancer: A review [published online ahead of print 2014/10/11]. World J Clin Oncol. 2014;5(4):568–575. [PubMed: 25302161]
5. Pflugfelder D, Wilkens JJ, Oelfke U. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy [published online ahead of print 2008/03/28]. Phys Med Biol. 2008;53(6):1689–1700. [PubMed: 18367797]
6. Unkelbach J, Bortfeld T, Martin BC, Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning [published online ahead of print 2009/02/25]. Med Phys. 2009;36(1):149–163. [PubMed: 19235384]
7. Fredriksson A, Forsgren A, Hardemark B. Minimax optimization for handling range and setup uncertainties in proton therapy [published online ahead of print 2011/04/28]. Med Phys. 2011;38(3):1672–1684. [PubMed: 21520880]
8. Liu W, Zhang X, Li Y, Mohan R. Robust optimization in intensity-modulated proton therapy. Med Phys. 2012;39:1079–1091. [PubMed: 22320818]
9. Liu W, Li Y, Li X, Cao W, Zhang X. Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques [published online ahead of print 2012/07/05]. Med Phys. 2012;39(6):3089–3101. [PubMed: 22755694]
10. Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy [published online ahead of print 2012/01/10]. Phys Med Biol. 2012;57(3):591–608. [PubMed: 22222720]
11. Liu W, Liao Z, Schild SE, et al. Impact of respiratory motion on worst-case scenario optimized intensity modulated proton therapy for lung cancers [published online ahead of print 2014/11/22]. Pract Radiat Oncol. 2015;5(2):e77–86. [PubMed: 25413400]
12. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning [published online ahead of print 2018/11/13]. Phys Med Biol. 2018;63(22):22TR02.
13. An Y, Liang J, Schild SE, Bues M, Liu W. Robust treatment planning with conditional value at risk chance constraints in intensity-modulated proton therapy [published online ahead of print 2017/01/04]. Med Phys. 2017;44(1):28–36. [PubMed: 28044325]
14. An Y, Shan J, Patel SH, et al. Robust intensity-modulated proton therapy to reduce high linear energy transfer in organs at risk [published online ahead of print 2017/10/05]. Med Phys. 2017;44(12):6138–6147. [PubMed: 28976574]
15. Liu C, Patel SH, Shan J, et al. Robust Optimization for Intensity Modulated Proton Therapy to Redistribute High Linear Energy Transfer from Nearby Critical Organs to Tumors in Head and Neck Cancer [published online ahead of print 2020/01/29]. International journal of radiation oncology, biology, physics. 2020;107(1):181–193. [PubMed: 31987967]
16. Liu C, Yu NY, Shan J, et al. Technical Note: Treatment planning system (TPS) approximations matter - comparing intensity-modulated proton therapy (IMPT) plan quality and robustness between a commercial and an in-house developed TPS for nonsmall cell lung cancer (NSCLC) [published online ahead of print 2019/09/10]. Med Phys. 2019;46(11):4755–4762. [PubMed: 31498885]
17. Liu C, Schild SE, Chang JY, et al. Impact of Spot Size and Spacing on the Quality of Robustly Optimized Intensity Modulated Proton Therapy Plans for Lung Cancer [published online ahead of

print 2018/03/20]. International journal of radiation oncology, biology, physics. 2018;101(2):479–489. [PubMed: 29550033]

18. Liu W, Inventor. System and Method For Robust Intensity-modulated Proton Therapy Planning. 09/02/2014, 2014.
19. Liu W, ed Robustness quantification and robust optimization in intensity-modulated proton therapy. Springer; 2015. Rath A, Sahoo N, eds. Particle Radiotherapy: Emerging Technology for Treatment of Cancer.
20. Liu W, Frank SJ, Li X, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers [published online ahead of print 2013/05/03]. Med Phys. 2013;40(5):051711. [PubMed: 23635259]
21. Liu W, Frank SJ, Li X, Li Y, Zhu RX, Mohan R. PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization [published online ahead of print 2013/02/08]. Med Phys. 2013;40(2):021709. [PubMed: 23387732]
22. Liu W, Schild SE, Chang JY, et al. Exploratory Study of 4D versus 3D Robust Optimization in Intensity Modulated Proton Therapy for Lung Cancer [published online ahead of print 2016/01/05]. International journal of radiation oncology, biology, physics. 2016;95(1):523–533. [PubMed: 26725727]
23. Shan J, Sio TT, Liu C, Schild SE, Bues M, Liu W. A novel and individualized robust optimization method using normalized dose interval volume constraints (NDIVC) for intensity-modulated proton radiotherapy [published online ahead of print 2018/11/06]. Med Phys. 2018. doi: 10.1002/mp.13276.
24. Shan J, An Y, Bues M, Schild SE, Liu W. Robust optimization in IMPT using quadratic objective functions to account for the minimum MU constraint [published online ahead of print 2017/11/18]. Med Phys. 2018;45(1):460–469. [PubMed: 29148570]
25. Feng H, Shan J, Ashman JB, et al. 4D robust optimization in small spot intensity-modulated proton therapy (IMPT) for distal esophageal carcinoma. Medical Physics.
26. Feng H, Sio TT, Rule WG, et al. Beam angle comparison for distal esophageal carcinoma patients treated with intensity-modulated proton therapy [published online ahead of print 2020/10/16]. J Appl Clin Med Phys. 2020;21(11):141–152.
27. Li Y, Liu W, Li X, Quan E, Zhang X. Toward a thorough Evaluation of IMPT Plan Sensitivity to Uncertainties: Revisit the Worst-Case Analysis with An Exhaustively Sampling Approach. Medical Physics. 2011;38(6):3853-+.
28. Liu C, Bhangoo RS, Sio TT, et al. Dosimetric comparison of distal esophageal carcinoma plans for patients treated with small-spot intensity-modulated proton versus volumetric-modulated arc therapies [published online ahead of print 2019/05/22]. J Appl Clin Med Phys. 2019;20(7):15–27. [PubMed: 31112371]
29. Liu C, Sio TT, Deng W, et al. Small-spot intensity-modulated proton therapy and volumetric-modulated arc therapies for patients with locally advanced non-small-cell lung cancer: A dosimetric comparative study [published online ahead of print 2018/10/18]. J Appl Clin Med Phys. 2018;19(6):140–148. [PubMed: 30328674]
30. Tryggestad EJ, Liu W, Pepin MD, Hallemeier CL, Sio TT. Managing treatment-related uncertainties in proton beam radiotherapy for gastrointestinal cancers [published online ahead of print 2020/03/17]. J Gastrointest Oncol. 2020;11(1):212–224. [PubMed: 32175124]
31. Zaghian M, Cao W, Liu W, et al. Comparison of linear and nonlinear programming approaches for “worst case dose” and “minmax” robust optimization of intensity-modulated proton therapy dose distributions [published online ahead of print 2017/03/17]. J Appl Clin Med Phys. 2017;18(2):15–25. [PubMed: 28300378]
32. Zaghian M, Lim G, Liu W, Mohan R. An Automatic Approach for Satisfying Dose-Volume Constraints in Linear Fluence Map Optimization for IMPT [published online ahead of print 2014/12/17]. J Cancer Ther. 2014;5(2):198–207. [PubMed: 25506501]
33. Hong L, Goitein M, Bucciolini M, et al. A pencil beam algorithm for proton dose calculations [published online ahead of print 1996/08/01]. Phys Med Biol. 1996;41(8):1305–1330. [PubMed: 8858722]

34. Schaffner B, Pedroni E, Lomax A. Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation [published online ahead of print 1999/03/11]. Phys Med Biol. 1999;44(1):27–41. [PubMed: 10071873]

35. Younkin JE, Morales DH, Shen J, et al. Clinical Validation of a Ray-Casting Analytical Dose Engine for Spot Scanning Proton Delivery Systems [published online ahead of print 2019/11/23]. Technol Cancer Res Treat. 2019;18:1533033819887182. [PubMed: 31755362]

36. Taylor PA, Kry SF, Followill DS. Pencil beam algorithms are unsuitable for proton dose calculations in lung. International Journal of Radiation Oncology* Biology* Physics. 2017;99(3):750–756.

37. Sasidharan BK, Aljabab S, Saini J, et al. Clinical Monte Carlo versus Pencil Beam Treatment Planning in Nasopharyngeal Patients Receiving IMPT [published online ahead of print 2019/11/28]. Int J Part Ther. 2019;5(4):32–40. [PubMed: 31773039]

38. Agostinelli S, Allison J, Amako Ka, et al. GEANT4—a simulation toolkit. Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2003;506(3):250–303.

39. Waters LS. MCNPX user's manual. Los Alamos National Laboratory. 2002;124.

40. Battistoni G, Bauer J, Boehlen TT, et al. The FLUKA code: an accurate simulation tool for particle therapy. Frontiers in oncology. 2016;6:116. [PubMed: 27242956]

41. Kozłowska WS, Böhlen TT, Cuccagna C, et al. FLUKA particle therapy tool for Monte Carlo independent calculation of scanned proton and carbon ion beam therapy. Physics in Medicine & Biology. 2019;64(7):075012. [PubMed: 30695766]

42. Perl J, Shin J, Schümann J, Faddegon B, Paganetti H. TOPAS: an innovative proton Monte Carlo platform for research and clinical applications. Medical physics. 2012;39(11):6818–6837. [PubMed: 23127075]

43. Yepes P, Randeniya S, Taddei PJ, Newhauser WD. Monte Carlo fast dose calculator for proton radiotherapy: application to a voxelized geometry representing a patient with prostate cancer. Physics in Medicine & Biology. 2008;54(1):N21. [PubMed: 19075361]

44. Fix MK, Frei D, Volken W, Born EJ, Aebersold DM, Manser P. Macro Monte Carlo for dose calculation of proton beams. Physics in Medicine and Biology. 2013;58(7):2027–2044. [PubMed: 23458969]

45. Wan Chan Tseung H, Ma J, Beltran C. A fast GPU-based Monte Carlo simulation of proton transport with detailed modeling of nonelastic interactions. Medical physics. 2015;42(6Part1):2967–2978. [PubMed: 26127050]

46. Jia X, Schümann J, Paganetti H, Jiang SB. GPU-based fast Monte Carlo dose calculation for proton therapy. Physics in Medicine & Biology. 2012;57(23):7783. [PubMed: 23128424]

47. Schiavi A, Senzacqua M, Pioli S, et al. Fred: a GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy. Physics in Medicine & Biology. 2017;62(18):7482. [PubMed: 28873069]

48. Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi-and many-core CPU architectures. Medical physics. 2016;43(4):1700–1712. [PubMed: 27036568]

49. Lin L, Huang S, Kang M, et al. A benchmarking method to evaluate the accuracy of a commercial proton monte carlo pencil beam scanning treatment planning system. Journal of applied clinical medical physics. 2017;18(2):44–49.

50. Chang CW, Huang S, Harms J, et al. A standardized commissioning framework of Monte Carlo dose calculation algorithms for proton pencil beam scanning treatment planning systems. Medical physics. 2020;47(4):1545–1557. [PubMed: 31945191]

51. Saini J, Maes D, Egan A, et al. Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations. Physics in Medicine & Biology. 2017;62(19):7659. [PubMed: 28749373]

52. Feng H, Shan J, Ashman JB, et al. Technical Note: 4D robust optimization in small spot intensity-modulated proton therapy (IMPT) for distal esophageal carcinoma [published online ahead of print 2021/06/01]. Med Phys. 2021;48(8):4636–4647. [PubMed: 34058026]

53. Yu J, Zhang X, Liao L, et al. Motion-robust intensity-modulated proton therapy for distal esophageal cancer [published online ahead of print 2016/03/05]. Med Phys. 2016;43(3):1111–1118. [PubMed: 26936698]

54. Bai X, Lim G, Grosshans D, Mohan R, Cao W. Robust optimization to reduce the impact of biological effect variation from physical uncertainties in intensity-modulated proton therapy [published online ahead of print 2018/12/14]. Phys Med Biol. 2019;64(2):025004. [PubMed: 30523932]

55. Cao W, Khabazian A, Yepes PP, et al. Linear energy transfer incorporated intensity modulated proton therapy optimization [published online ahead of print 2017/11/14]. Phys Med Biol. 2017;63(1):015013. [PubMed: 29131808]

56. Fracchiolla F, Engwall E, Janson M, et al. Clinical validation of a GPU-based Monte Carlo dose engine of a commercial treatment planning system for pencil beam scanning proton therapy. Physica Medica. 2021;88:226–234. [PubMed: 34311160]

57. Yepes PP, Mirkovic D, Taddei PJ. A GPU implementation of a track-repeating algorithm for proton radiotherapy dose calculations. Physics in Medicine & Biology. 2010;55(23):7107. [PubMed: 21076192]

58. Newhauser WD, Zhang R. The physics of proton therapy [published online ahead of print 2015/03/24]. Physics in medicine and biology. 2015;60(8):R155–R209. [PubMed: 25803097]

59. Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures [published online ahead of print 2016/04/03]. Med Phys. 2016;43(4):1700. [PubMed: 27036568]

60. Deng W, Younkin JE, Souris K, et al. Technical Note: Integrating an open source Monte Carlo code “MCsquare” for clinical use in intensity-modulated proton therapy [published online ahead of print 2020/03/11]. Med Phys. 2020;47(6):2558–2574. [PubMed: 32153029]

61. Deng W, Ding X, Younkin JE, et al. Hybrid 3D analytical linear energy transfer calculation algorithm based on precalculated data from Monte Carlo simulations [published online ahead of print 2019/11/24]. Med Phys. 2020;47(2):745–752. [PubMed: 31758864]

62. Holmes J, Shen J, Shan J, et al. Evaluation and second check of a commercial Monte Carlo dose engine for small-field apertures in pencil beam scanning proton therapy [published online ahead of print 2022/03/20]. Med Phys. 2022. doi: 10.1002/mp.15604.

63. Wagenaar D, Tran LT, Meijers A, et al. Validation of linear energy transfer computed in a Monte Carlo dose engine of a commercial treatment planning system. Physics in Medicine & Biology. 2020;65(2):025006. [PubMed: 31801119]

64. Huang S, Souris K, Li S, et al. Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy. Medical Physics. 2018;45(12):5631–5642. [PubMed: 30295950]

65. Huang S, Kang M, Souris K, et al. Validation and clinical implementation of an accurate Monte Carlo code for pencil beam scanning proton therapy. Journal of Applied Clinical Medical Physics. 2018;19(5):558–572. [PubMed: 30058170]

66. Sorriaux J, Testa M, Paganetti H, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. Physica Medica. 2017;38:10–15. [PubMed: 28610689]

67. Huang S, Souris K, Li S, et al. Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy [published online ahead of print 2018/10/09]. Med Phys. 2018;45(12):5631–5642. [PubMed: 30295950]

68. Huang S, Kang M, Souris K, et al. Validation and clinical implementation of an accurate Monte Carlo code for pencil beam scanning proton therapy [published online ahead of print 2018/07/31]. J Appl Clin Med Phys. 2018;19(5):558–572. [PubMed: 30058170]

69. Shan J, Yang Y, Schild SE, et al. Intensity-modulated proton therapy (IMPT) interplay effect evaluation of asymmetric breathing with simultaneous uncertainty considerations in patients with non-small cell lung cancer [published online ahead of print 2020/09/24]. Med Phys. 2020;47(11):5428–5440. [PubMed: 32964474]

70. Feng H, Shan J, Anderson JD, et al. Per-voxel constraints to minimize hot spots in linear energy transfer-guided robust optimization for base of skull head and neck cancer patients in

IMPT [published online ahead of print 2021/11/30]. Med Phys. 2022;49(1):632–647. [PubMed: 34843119]

71. Yang Y, Muller OM, Shiraishi S, et al. Empirical Relative Biological Effectiveness (RBE) for Mandible Osteoradionecrosis (ORN) in Head and Neck Cancer Patients Treated With Pencil-Beam-Scanning Proton Therapy (PBSPT): A Retrospective, Case-Matched Cohort Study. Front Oncol. 2022;12.
72. Yang Y, Vargas CE, Bhangoo RS, et al. Exploratory Investigation of Dose-Linear Energy Transfer (LET) Volume Histogram (DLVH) for Adverse Events Study in Intensity Modulated Proton Therapy (IMPT) [published online ahead of print 2021/02/24]. Int J Radiat Oncol Biol Phys. 2021;110(4):1189–1199. [PubMed: 33621660]
73. Feng H PS, Wong WW, Younkin JE, Penoncello GP, Hernandez Morales D, Stoker JB, Robertson DG, Fatyga M, Bues M, Schild SE, Foote RL, Liu W. GPU-Accelerated Monte Carlo-based Online Adaptive Proton Therapy - A Feasibility Study. Medical Physics. 2022.
74. Liu C, Ho MW, Park J, et al. Fast MCsquare-Based Independent Dose Verification Platform for Pencil Beam Scanning Proton Therapy. Technology in cancer research & treatment. 2021;20:15330338211033076–15330338211033076. [PubMed: 34338058]
75. Persoon LCGG, Podesta M, van Elmpt WJC, MaNijsten SMJJG, Verhaegen F. A fast three-dimensional gamma evaluation using a GPU utilizing texture memory for on-the-fly interpolations. Medical Physics. 2011;38(7):4032–4035. [PubMed: 21859001]
76. Ding X, Liu W, Shen J, et al. Use of a radial projection to reduce the statistical uncertainty of spot lateral profiles generated by Monte Carlo simulation. Journal of applied clinical medical physics. 2017;18(6):88–96.
77. Anand A, Sahoo N, Zhu XR, et al. A procedure to determine the planar integral spot dose values of proton pencil beam spots. Medical physics. 2012;39(2):891–900. [PubMed: 22320798]
78. Perl J, Shin J, Schumann J, Faddegon B, Paganetti H. TOPAS: an innovative proton Monte Carlo platform for research and clinical applications [published online ahead of print 2012/11/07]. Med Phys. 2012;39(11):6818–6837. [PubMed: 23127075]
79. Faddegon B, Ramos-Méndez J, Schuemann J, et al. The TOPAS tool for particle simulation, a Monte Carlo simulation tool for physics, biology and clinical research [published online ahead of print 2020/04/06]. Phys Med. 2020;72:114–121. [PubMed: 32247964]
80. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Medical Physics. 1998;25(5):656–661. [PubMed: 9608475]
81. Hernandez Morales D, Shan J, Liu W, et al. Automation of routine elements for spot-scanning proton patient-specific quality assurance [published online ahead of print 2018/10/20]. Med Phys. 2019;46(1):5–14. [PubMed: 30339270]
82. Wendling M, Zijp LJ, McDermott LN, et al. A fast algorithm for gamma evaluation in 3D. Medical Physics. 2007;34(5):1647–1654. [PubMed: 17555246]
83. Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218. Medical physics. 2018;45(4):e53–e83. [PubMed: 29443390]
84. Li Y, Sun W-Z, Liu H, et al. Development of a GPU-superposition Monte Carlo code for fast dose calculation in magnetic fields. Physics in Medicine & Biology. 2022.
85. Neuenschwander H, Born EJ. A macro Monte Carlo method for electron beam dose calculations. Physics in Medicine & Biology. 1992;37(1):107.
86. Fix MK, Frei D, Volken W, Born EJ, Aebersold DM, Manser P. Macro Monte Carlo for dose calculation of proton beams. Physics in Medicine & Biology. 2013;58(7):2027. [PubMed: 23458969]
87. Yepes P, Randeniya S, Taddei PJ, Newhauser WD. A track-repeating algorithm for fast Monte Carlo dose calculations of proton radiotherapy. Nuclear technology. 2009;168(3):736–740. [PubMed: 20865140]
88. Kang Y, Shen J, Bues M, Hu Y, Liu W, Ding X. Clinical modeling and validation of breast tissue expander metallic ports in a commercial treatment planning system for proton therapy. Medical Physics. 2021.

89. Frenkel D Speed-up of Monte Carlo simulations by sampling of rejected states. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(51):17571. [PubMed: 15591337]

90. Ramos-Méndez J, Perl J, Faddegon B, Schümann J, Paganetti H. Geometrical splitting technique to improve the computational efficiency in Monte Carlo calculations for proton therapy. Medical physics. 2013;40(4):041718–041718. [PubMed: 23556888]

91. Fernández-Varea J, Mayol R, Baró J, Salvat F. On the theory and simulation of multiple elastic scattering of electrons. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms. 1993;73(4):447–473.

92. Salvat F, Fernández-Varea JM, Sempau J. PENELOPE-2006: A code system for Monte Carlo simulation of electron and photon transport. Paper presented at: Workshop proceedings2006.

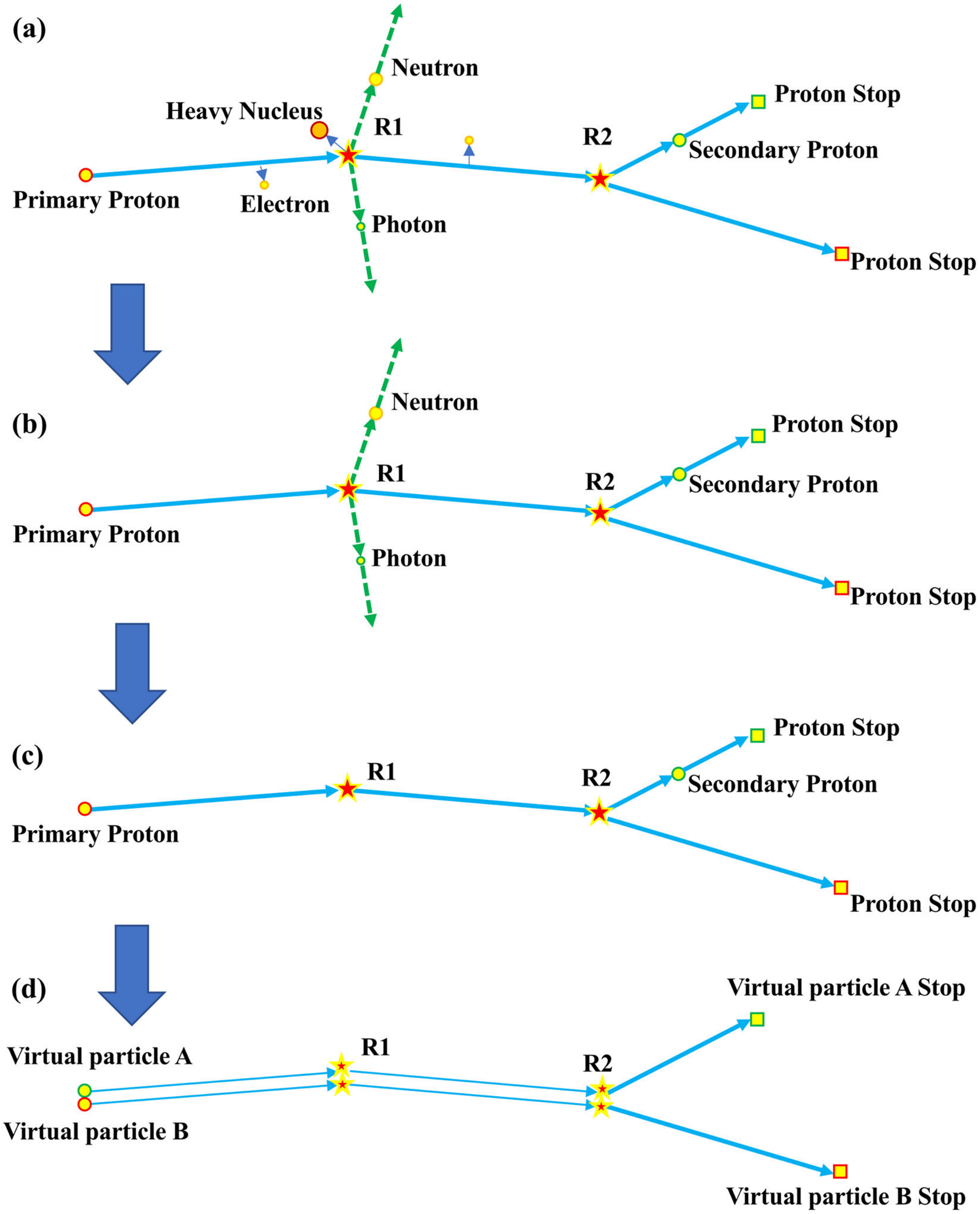

Figure 1.

Diagram of converting the track histories of a realistic proton and its secondaries in a conventional Monte Carlo simulation into two virtual particles (a) Track history of a realistic proton. (b) the doses of electrons, heavy ions, and nuclear fragments were locally deposited. (c) neutrons and gamma rays were considered to escape from human bodies and thus ignored. (d) Track history of two virtual particles to have the same path from the starting point to R2. These two VPs are independent from each other. This figure is only used to demonstrate the concept of VP.

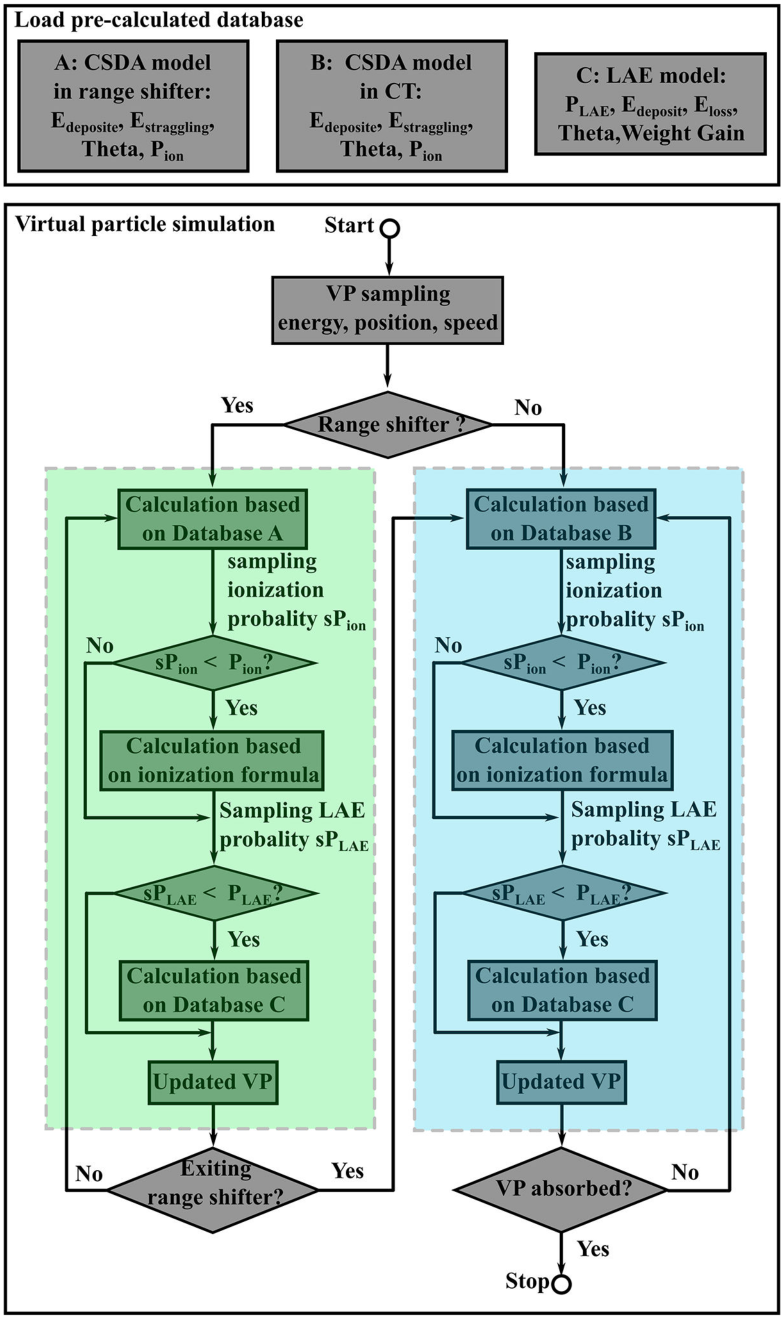
Load pre-calculated database
A: CSDA model in range shifter: $E_{deposite}$, $E_{straggling}$, Theta, P_{ion}
B: CSDA model in CT: $E_{deposite}$, $E_{straggling}$, Theta, P_{ion}
C: LAE model: P_{LAE}, $E_{deposit}$, E_{loss}, Theta,Weight Gain
Virtual particle simulation
Start
VP sampling energy, position, speed
Range shifter ?
Yes
No
Calculation based on Database A
sampling ionization probality sP_{ion}
sP_{ion} < P_{ion}?
No
Yes
Calculation based on ionization formula
Sampling LAE probality sP_{LAE}
sP_{LAE} < P_{LAE}?
Yes
Calculation based on Database C
Updated VP
Calculation based on Database B
sampling ionization probality sP_{ion}
sP_{ion} < P_{ion}?
No
Yes
Calculation based on ionization formula
Sampling LAE probality sP_{LAE}
sP_{LAE} < P_{LAE}?
Yes
Calculation based on Database C
Updated VP
No
Exiting range shifter?
Yes
VP absorbed?
No
Yes
Stop

Figure 2.

Workflow of VPMC. The upper box shows the loading process of the pre-calculated databases of the CSDA model in range shifter, the CSDA model in patient geometry, and the LAE model. The lower box shows the simulation process of VPs. The steps in the green box indicates the simulations in range shifter, while the steps in the blue box indicates the simulations in patient geometry.

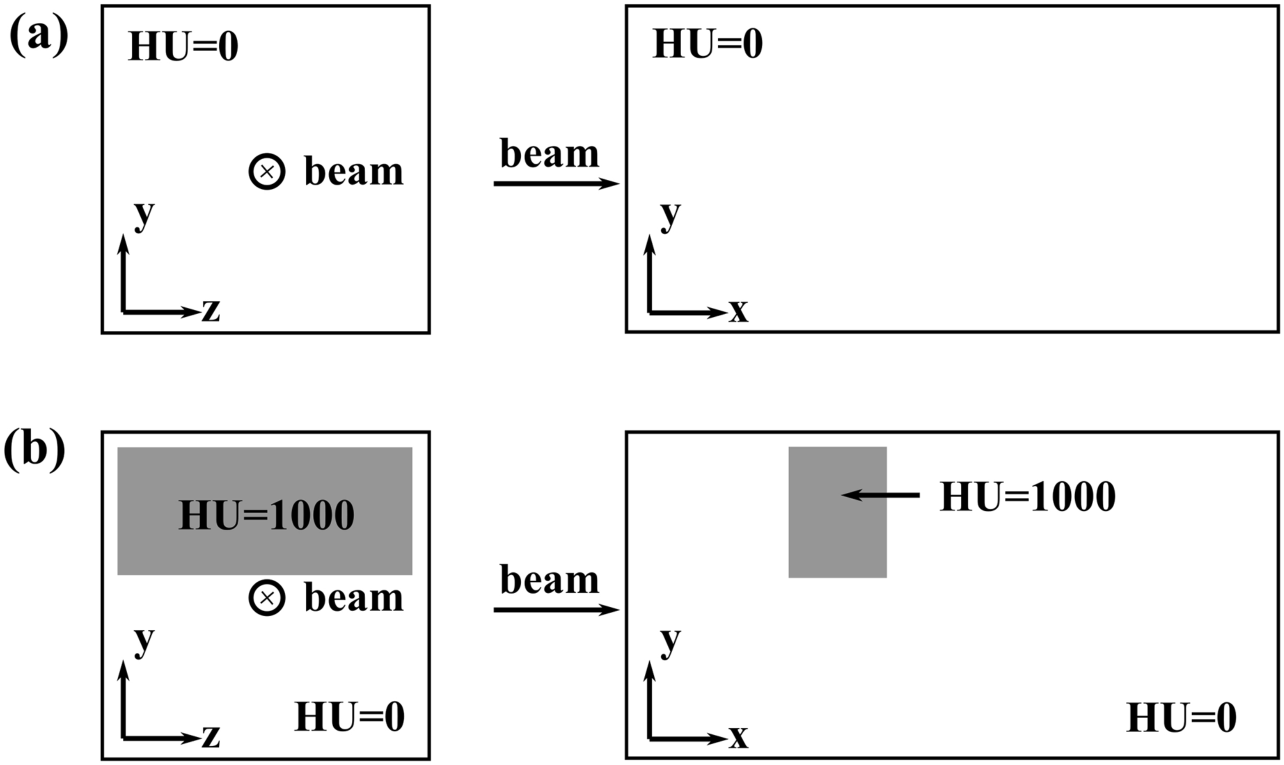

Figure 3.

Illustration of the homogenous phantom (a) and inhomogeneous phantom. The beam enters the phantoms at the center of the y-z plane. The larger white boxes are composed of the water (HU = 0), the smaller grey boxes are composed of the blocking material (HU = 1000).

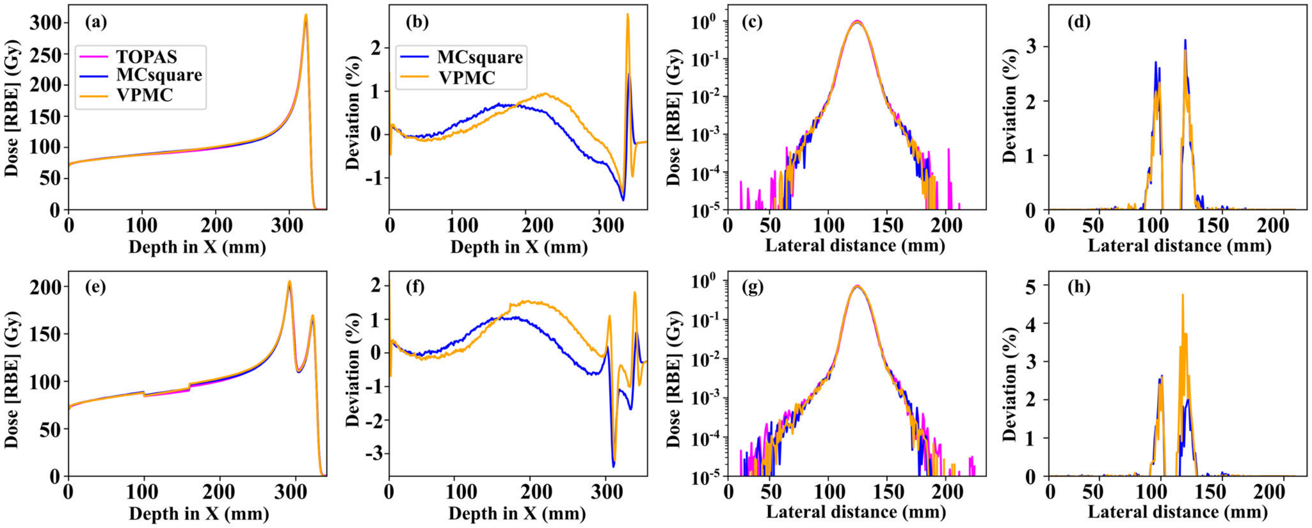

Figure 4.

The IDD curves (a)(e), log-scale lateral profiles at the Bragg peak (c)(g) of the dose distribution generated by TOPAS (magenta), MCsquare (blue), and VPMC (orange), and the corresponding IDD (b)(f) and lateral profile (d)(h) differences between MCsquare and TOPAS (blue) and the differences between VPMC and TOPAS (orange), in a homogeneous phantom (top row) and in an inhomogeneous phantom (bottom row). RBE = 1.1.

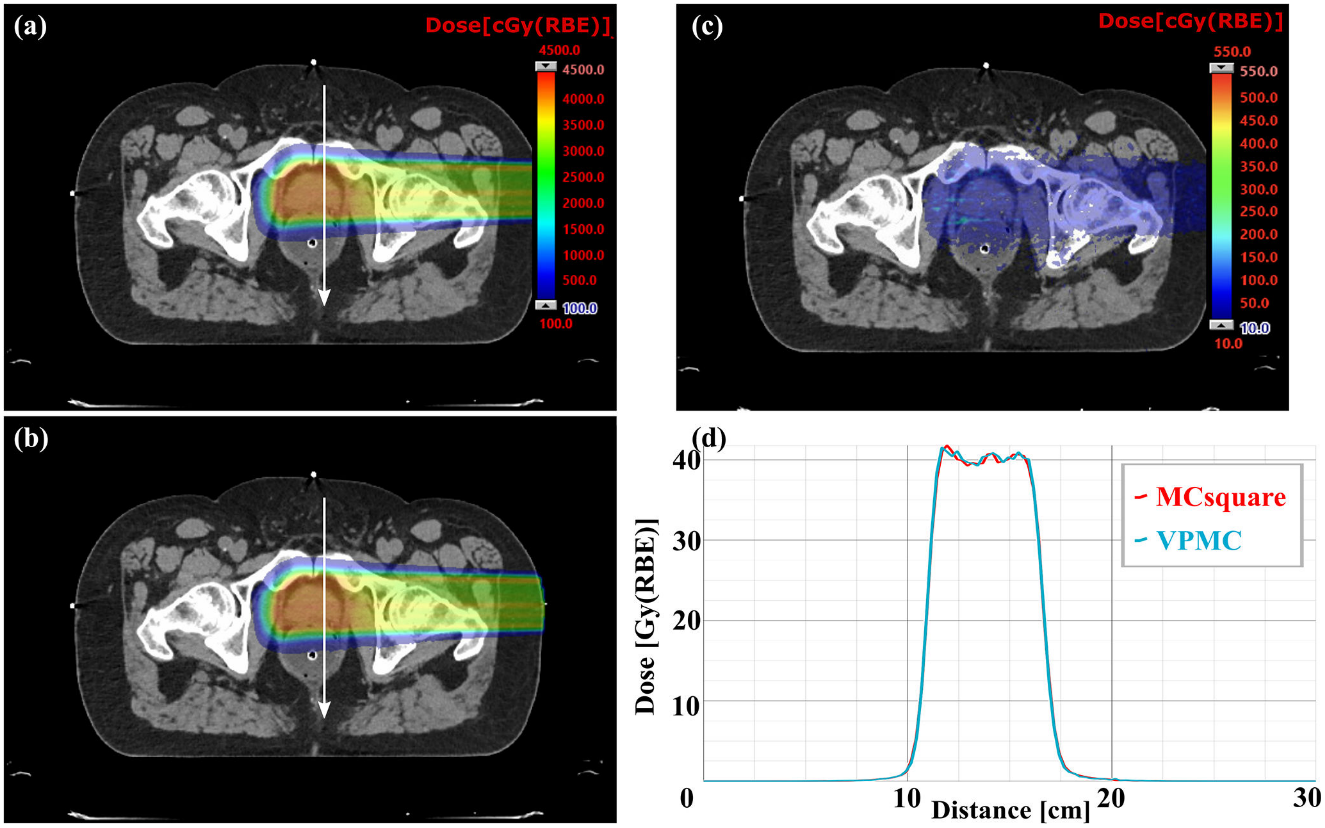

Figure 5.

Comparison of dose profiles on typical transverse planes of a prostate patient without range shifter between VPMC and MCsquare. (a) Dose map from MCsquare, (b) Dose map from VPMC, (c) Absolute dose difference map between the MCsquare calculated dose and the VPMC calculated dose. The white arrow indicates the position and direction of the dose profiles in (c). The red curve in (d) is the dose profile from MCsquare, while the blue curve is the one from VPMC. RBE = 1.1.

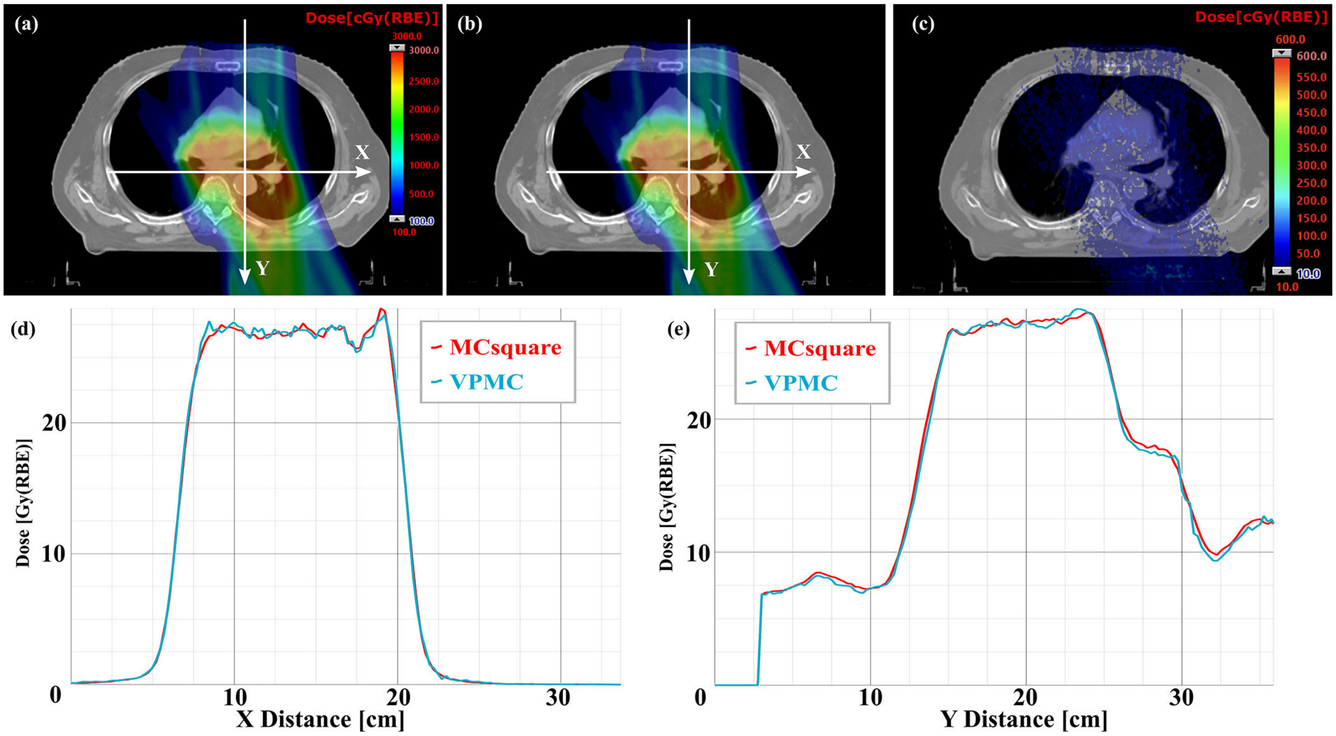

Figure 6.

Comparison of dose profiles on typical transverse planes of a chest wall patient with ERS between VPMC and MCsquare. (a) Dose map from MCsquare, (b) Dose map from VPMC, (c) Absolute dose difference map between the MCsquare calculated dose and the VPMC calculated dose. The white arrows indicate the position and directions of the dose profiles in (d) dose profile in X direction and (e) dose profile in Y direction. The red curves in (d) and (e) are the dose profile from MCsquare, while the blue curves are the ones from VPMC. RBE = 1.1.

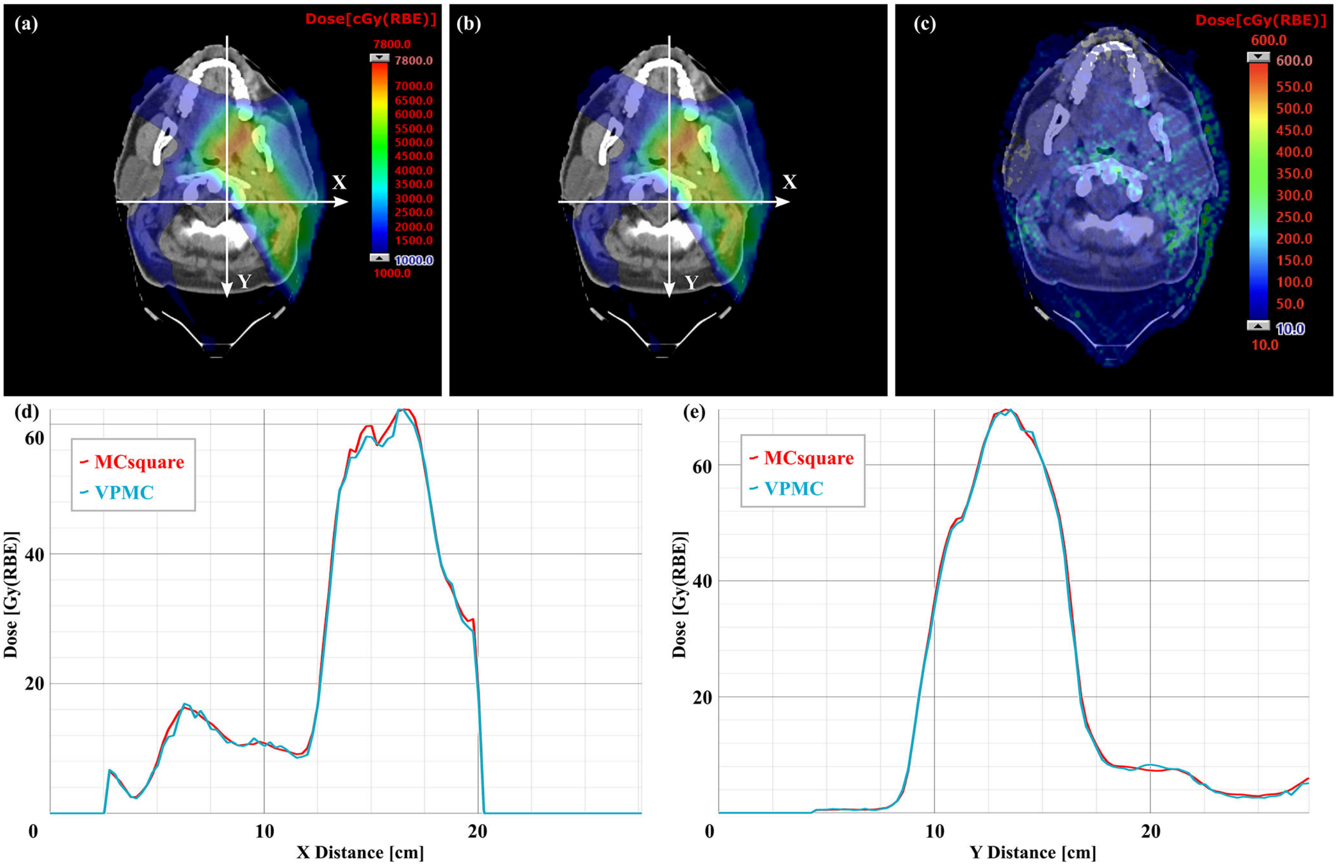

Figure 7.

Comparison of dose profiles on typical transverse planes of a H&N patient with RS between VPMC and MCsquare. (a) Dose map from MCsquare, (b) Dose map from VPMC, (c) Absolute dose difference map between the MCsquare calculated dose and the VPMC calculated dose. The white arrows indicate the position and directions of the dose profile in (d) dose profile in X direction and (e) dose profile in Y direction. The red curves in (d) and (e) are the dose profile from MCsquare, while the blue curves are the ones from VPMC. RBE = 1.1.

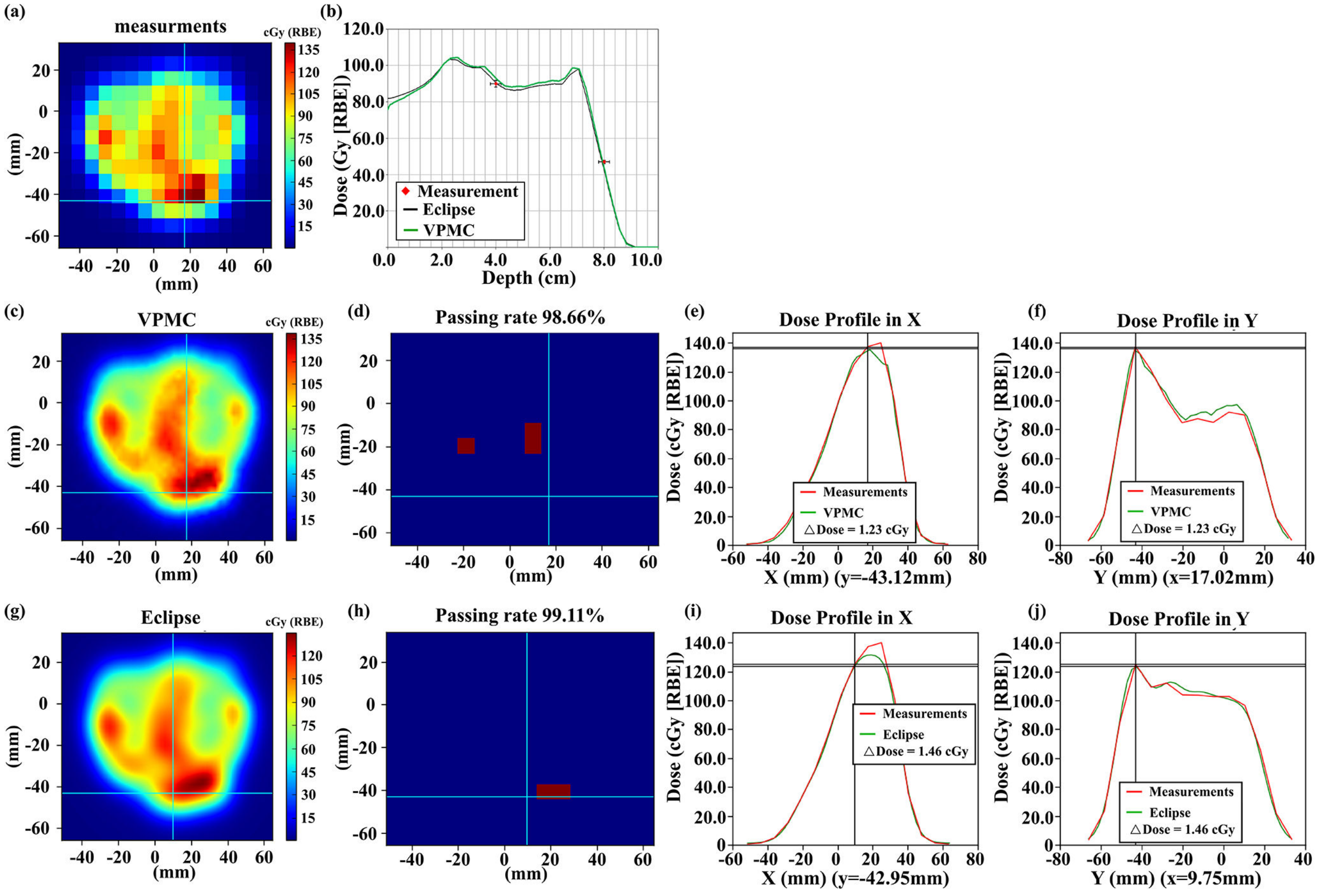

Figure 8.

Comparisons of the measured 2D plane dose during PSQA (a) with VPMC calculation result (c), and Eclipse™ calculation result (g) at a depth of 4.0 cm for a lung cancer patient without range shifter. The corresponding 2D-3D Gamma analysis pass/fail maps are shown in (d) and (h) with 3%/2mm/10%. Subpanel (b) displays the dose profiles from VPMC and Eclipse™ in the beam direction. The black line is from the Eclipse™ calculation result, while the green line is from the VPMC calculation result. Red points are the measured results with an error bar of 2%/2mm. Subpanel (e) (f) displays the dose profile comparison between the VPMC calculated dose and the measured dose in the X direction at the Y position indicated by the horizonal line in (c) and in the Y direction at the X position indicated by the vertical line in (c), respectively. Subpanel (i) (j) displays the dose profile comparison between the Eclipse™ calculated dose and measured dose in the X direction at the Y position indicated by the horizonal line in (g) and in the Y direction at the X position indicated by the vertical line in (g), respectively. RBE = 1.1.

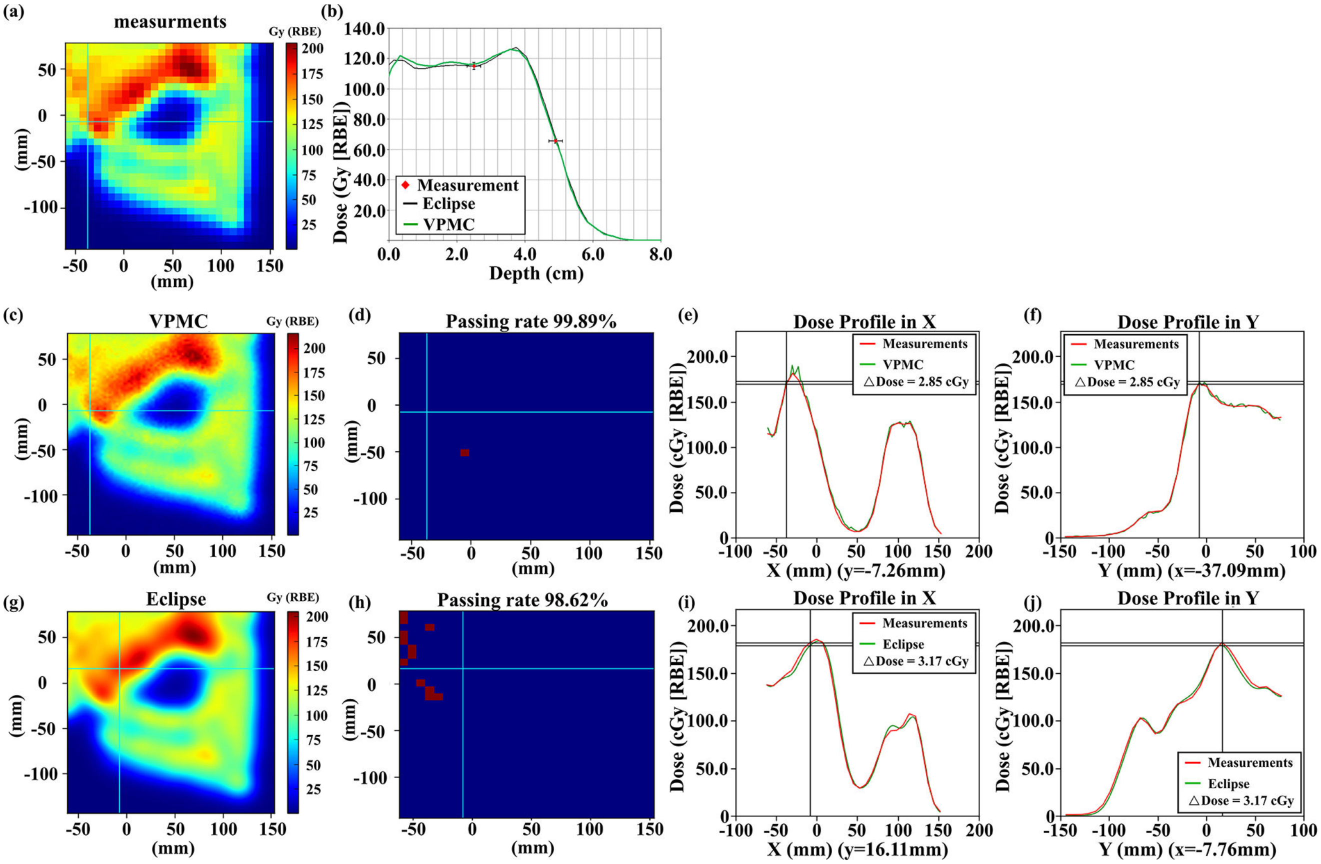

Figure 9.

Comparisons of the measured 2D plane dose during PSQA (a) with VPMC calculation results (c) and Eclipse™ calculation results (g) at a depth of 2.0 cm for a chest wall patient with RS. The corresponding 2D-3D Gamma analysis pass/fail maps are shown in (d) and (h) with 3%/2mm/10%. Subpanel (b) displays the dose profiles from VPMC and Eclipse™ in the beam direction. The black line is from the Eclipse™ calculation result, while the green line is from the VPMC calculation result. Red points are the measured results with an error bar of 2%/2mm. Subpanel (e) (f) displays the dose profile comparison between the VPMC calculated dose and the measured dose in the X direction at the Y position indicated by the horizonal line in (c) and in the Y direction at the X position indicated by the vertical line in (c), respectively. Subpanel (i) (j) displays the dose profile comparison between the Eclipse™ calculated dose and measured dose in the X direction at the Y position indicated by the horizonal line in (g) and in the Y direction at the X position indicated by the vertical line in (g), respectively. RBE = 1.1.

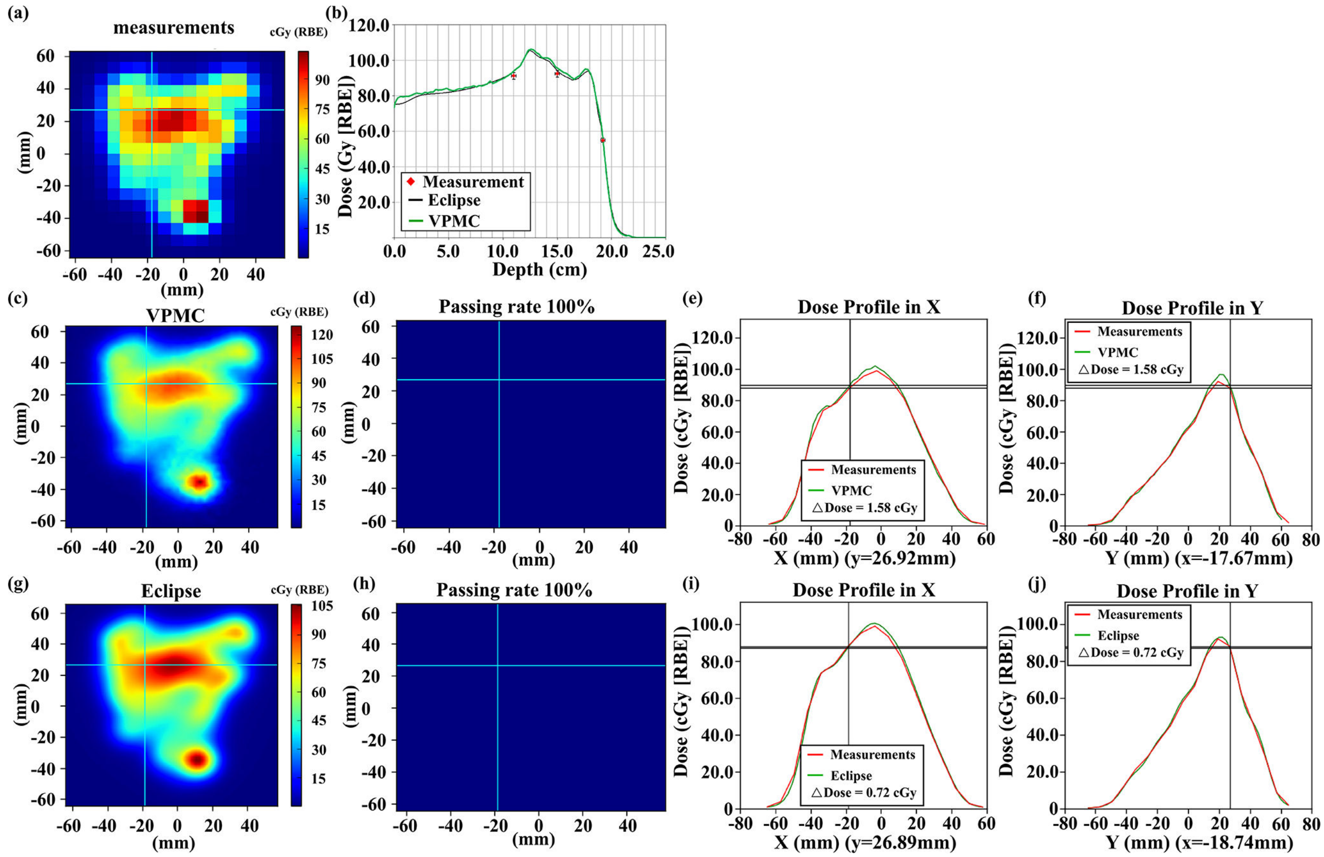

Figure 10.

Comparisons of the measured 2D plane dose during PSQA (a) with VPMC calculation results (c) and Eclipse™ calculation results (g) at a depth of 15.0 cm for a H&N patient with ERS. The corresponding 2D-3D Gamma analysis results were shown in (d) and (h) with 3%/2mm/10%. Subpanel (b) displays the dose profiles from VPMC and Eclipse™ in the beam direction. The black line is from the Eclipse™ calculation result, while the green line is from the VPMC calculation result. Red points are the measured results with an error bar of 2%/2mm. Subpanel (e) (f) displays the dose profile comparison between the VPMC calculated dose and the measured dose in the X direction at the Y position indicated by the horizonal line in (c) and in the Y direction at the X position indicated by the vertical line in (c), respectively. Subpanel (i) (j) displays the dose profile comparison between the Eclipse™ calculated dose and measured dose in the X direction at the Y position indicated by the horizonal line in (g) and in the Y direction at the X position indicated by the vertical line in (g), respectively. RBE = 1.1.

Table. 1

Characteristics of selected plans.

Patient #	Disease site	Range shifter	Total spot	Dose grid size
1	Prostate	No	2,235	168*115*216
2	Prostate	No	1,786	149*109*171
3	Head & Neck	ERS	6,501	123*89*105
4	Head & Neck	ERS	17,224	226*119*154
5	Brain	No	5,035	223*164*161
6	Orbit	No	1,644	213*162*171
7	Pelvis	No	26,850	200*125*189
8	Thorax	No	3,802	256*104*125
9	Chest wall	ERS	60,690	260*129*188
10	Lung	ERS	20,403	229*184*174
11	Liver	No	9,758	217*151*144
12	Chest wall	RS	67,875	257*164*204
13	Chest wall	RS	51,965	251*136*161

abbreviations: RS for range shifter located at a regular position, ERS for range shifter located at an extended position.

Table. 2

Comparison between VPMC and MCsquare in the 13 selected patients. 3D-3D Gamma analysis was done with 2%/2mm/10%.

Patient #	3D-3D Gamma referenced to MC2 (%)		Time (s)		Statistical Uncertainty (%)		Efficiency (10e4*1/s)	
	ECL	VPMC	MC2	VPMC	MC2	VPMC	MC2	VPMC
1	75.80	98.97	9.84	0.91	0.34	0.40	0.88	6.86
2	72.40	99.19	8.63	0.82	0.31	0.34	1.23	10.84
3	96.43	98.93	70.22	1.23	0.20	0.68	0.36	1.74
4	94.05	97.35	92.25	2.46	0.38	0.91	0.08	0.49
5	91.23	99.58	100.19	1.65	0.22	0.45	0.20	3.05
6	90.00	99.63	23.71	1.25	0.24	0.46	0.72	3.82
7	96.27	99.84	171.01	3.28	0.19	0.35	0.17	2.55
8	95.22	98.12	19.13	0.84	0.38	0.76	0.37	2.06
9	91.19	97.62	188.12	6.02	0.26	0.72	0.08	0.32
10	93.73	99.25	133.01	3.98	0.34	0.59	0.07	0.73
11	92.99	99.44	23.13	1.17	0.11	0.15	3.88	36.91
12	90.19	96.74	410.89	7.89	0.33	0.84	0.02	0.18
13	85.25	96.64	213.10	5.46	0.29	0.80	0.06	0.29
Mean	89.60	98.56	112.56	2.84	0.27	0.57	0.62	5.37
SD	7.24	1.09	114.38	2.33	0.08	0.23	1.05	9.96

abbreviations: MC2 stands for MCsquare, ECL stands for the analytical dose engine of Eclipse™ver. 15.6, SD stands for standard Deviation.

Table 3.

2D-3D Gamma passing rates comparing Ecplise™ and VPMC calculation results with the 2D measured plane doses during PSQA with a criterion of 3%/2mm and a threshold of 10%.

Patient #	2D-3D Gamma analysis (%)	
	ECL vs Measurements	**VPMC vs Measurements**
1	98.93	98.50
2	96.94	97.83
3	99.97	100.00
4	99.93	100.00
5	98.42	97.47
6	99.48	98.44
7	99.60	99.47
8	99.57	99.67
9	97.60	97.83
10	100.00	99.37
11	98.91	98.56
12	99.22	99.89
13	96.80	98.79
Mean	98.87	98.91
SD	1.12	0.88

abbreviations: ECL for the analytical dose engine of Eclipse™, SD stands for standard deviation.