A GPU-accelerated Monte Carlo dose calculation platform and its application toward validating an MRI-guided radiation therapy beam model

Yuhe Wang, Thomas R. Mazur, Olga Green, Yanle Hu, Hua Li, Vivian Rodriguez,
H. Omar Wooten, Deshan Yang, Tianyu Zhao, Sasa Mutic, and H. Harold Li[a)]
*Department of Radiation Oncology, Washington University School of Medicine, 4921 Parkview Place,
Campus Box 8224, St. Louis, Missouri 63110*

(Received 9 December 2015; revised 5 May 2016; accepted for publication 19 May 2016;
published 10 June 2016)

Purpose: The clinical commissioning of IMRT subject to a magnetic field is challenging. The purpose of this work is to develop a GPU-accelerated Monte Carlo dose calculation platform based on PENELOPE and then use the platform to validate a vendor-provided MRIdian head model toward quality assurance of clinical IMRT treatment plans subject to a 0.35 T magnetic field.

Methods: PENELOPE was first translated from FORTRAN to C++ and the result was confirmed to produce equivalent results to the original code. The C++ code was then adapted to CUDA in a workflow optimized for GPU architecture. The original code was expanded to include voxelized transport with Woodcock tracking, faster electron/positron propagation in a magnetic field, and several features that make GPENELOPE highly user-friendly. Moreover, the vendor-provided MRIdian head model was incorporated into the code in an effort to apply GPENELOPE as both an accurate and rapid dose validation system. A set of experimental measurements were performed on the MRIdian system to examine the accuracy of both the head model and GPENELOPE. Ultimately, GPENELOPE was applied toward independent validation of patient doses calculated by MRIdian's KMC.

Results: An acceleration factor of 152 was achieved in comparison to the original single-thread FORTRAN implementation with the original accuracy being preserved. For 16 treatment plans including stomach (4), lung (2), liver (3), adrenal gland (2), pancreas (2), spleen(1), mediastinum (1), and breast (1), the MRIdian dose calculation engine agrees with GPENELOPE with a mean gamma passing rate of 99.1% ± 0.6% (2%/2 mm).

Conclusions: A Monte Carlo simulation platform was developed based on a GPU- accelerated version of PENELOPE. This platform was used to validate that both the vendor-provided head model and fast Monte Carlo engine used by the MRIdian system are accurate in modeling radiation transport in a patient using 2%/2 mm gamma criteria. Future applications of this platform will include dose validation and accumulation, IMRT optimization, and dosimetry system modeling for next generation MR-IGRT systems. © *2016 American Association of Physicists in Medicine.* [http://dx.doi.org/10.1118/1.4953198]

Key words: Monte Carlo, GPU, CUDA, PENELOPE, MRIdian

1. INTRODUCTION

Monte Carlo radiation transport simulation is generally considered to be the most accurate method for dose calculation in radiation therapy.[1,2] Well-known Monte Carlo packages such as MCNP,[3] GEANT4,[4] EGS4/EGSnrc,[5,6] and PENELOPE (Refs. 7 and 8) have been demonstrated to agree excellently with experimental data under a wide range of conditions. For example, EGSnrc was shown to pass the Fano cavity test at the 0.1% level.[2] Here, we categorize these platforms as "accuracy-oriented." While these packages are highly accurate, they typically require long simulation time to finish a sufficient number of histories in order to achieve adequate statistical uncertainty.

Three approaches have been considered for accelerating Monte Carlo calculations: (1) simplifying particle transport mechanisms, thus reducing the necessary time for each particle history; (2) using variance reduction techniques such as particle splitting, Russian roulette, and interaction forcing to reduce the total history number required to achieve a given uncertainty; and (3) enhancing the computational capability by parallelizing the simulation on multiple CPU or GPU threads.[9] Packages like VMC (Refs. 10–12) and DPM (Ref. 13) applied approaches (1) and (2) to achieve clinically desired speeds, but sacrificed generality and absolute accuracy by dropping simulation of positrons and using simpler cross section profiles, among other simplifications. gDPM (Refs. 14 and 15) further utilized approach (3) (i.e., GPU parallelism) to obtain higher efficiency compared to the original DPM, while GPUMCD (Ref. 16) performed similar simplification to DPM and was directly oriented to GPU implementation. GMC (Ref. 17) was developed based on GEANT4 but it results in larger discrepancy from GEANT4 than expected (2%/2 mm gamma passing rate is 91.74% for IMRT plans). Accuracy was possibly compromised for GMC by the fact that GEANT4 uses a lot of virtual functions and class inheritances that make implementing a faithful adaptation from C++ to CUDA difficult. Here, we categorize these implementations as "efficiency-oriented."

Fast Monte Carlo implementations like VMC and gDPM perform admirably for applications that require quick response to user changes such as treatment planning. Most recent developments include online IMRT planning using GPUMCD for the MRI-linear accelerator[18] and the clinical use of KMC on the MRIdian system (ViewRay, Inc., Cleveland, OH). The MRIdian system integrates a 0.35 T whole-body MR imaging system into an RT delivery system consisting of a rotating gantry with three ^{60}Co heads spaced 120° apart that can provide a maximum combined dose rate of 550 cGy/min at the isocenter. Its treatment planning system uses an optimized Monte Carlo code based on VMC to achieve clinically acceptable speed at the expense of accuracy to some extent. KMC is able to complete an IMRT plan calculation subject to a magnetic field within a few minutes, thus rendering online adaptive treatment a clinical reality. Commissioning such a clinical system, however, is challenging and relies largely upon complex experimental validations. The limited availability of quantitative, MRI-compatible, water-equivalent dosimeters makes accurate, multidimensional measurements quite difficult.[19,20] Therefore, a computational system needs to be developed to complement the experimental approach that enables dose comparison in the patient geometry and dosimeter modeling, among other things. Without a viable alternative on a parallel platform, physicists to date have relied on the stalwart, accuracy-oriented Monte Carlo implementations,[21,22] mostly without taking magnetic fields into account, to accomplish these tasks in their practices. In certain cases, such as dose validation for adaptive treatments and dose accumulation using large volumetric data over the treatment course, calculation speed is an important factor to be considered. These applications requiring both high accuracy and clinically acceptable computation time motivated us to develop a GPU-accelerated Monte Carlo engine derived from PENELOPE by applying approach (3) only.

PENELOPE is an experimentally well-validated Monte Carlo platform for simulating photon, electron, and positron transport in various supported materials.[23–28] Its kernel (version 2006) is relatively compact including roughly 3000 lines of FORTRAN code. The code is well-documented with extensive detail on the theory underlying all scattering processes, thus readily enabling its adaptation on a GPU, namely, our implementation GPENLELOPE presented in this work. Beyond just implementing the kernel on a GPU, we expanded upon the original code to include voxelized tracking, charged particle propagation in magnetic fields, and several key features that make GPENLELOPE highly user-friendly.

Our primary application of GPENLELOPE in this work is to perform an accurate and independent check of clinical treatment plans within a clinically desired time. First, we verified that the source head is correctly modeled by integrating it into GPENLELOPE (noting that a dose calculation system includes both a head model and calculation engine) and comparing simulation results with experiments. Our measurements indicate that the head model provided by the vendor describes the hardware accurately. Then, the head model together with GPENELOPE can serve as an accurate dose calculation system for the MRIdian platform. We therefore can use this system to check clinical treatment plans generated by MRIdian's treatment planning system (which uses the same head model and the KMC engine) independently and quickly.

2. METHODS

2.A. PENELOPE in C++

We first translated the PENELOPE kernel from FORTRAN to C++. PENELOPE was implemented in FORTRAN 77 with an archaic programming style (e.g., many antiquated "GOTO" statements). Although GPU programming for FORTRAN has been enabled by the PGI (Ref. 29) compiler (via collaboration with NVIDIA), it lacks good object-oriented programming support and some general libraries, so convenient features like batch work and file compression cannot be implemented easily. Moreover, as the MRIdian head model was developed in C++, rewriting the PENELOPE kernel in C++ first will more readily enable its application to MRIdian validation.

To build PENELOPE in C++, we first extracted all relevant material data tables to a class called Material and assigned shared data to global variables. All "jump" and "knock" functions were rewritten as member functions of this class in an optimized logic sequence. The lengthy code for generating data tables (over 7000 lines) did not need to be translated to C++; instead, we added an interface function in PENELOPE that we compiled to a DLL module. We can call this DLL in C++ to preprocess materials and export the relevant data table to a file that will be addressed to memory by the C++ code later.

The original PENELOPE only supports single-thread processes, while multithreading through OpenMP is necessary to fully exploit modern multicore CPUs. We ensure that all kernel functions are thread-safe by managing thread-related variables accordingly in a single class. We additionally exploit MPI (with a set of workload balancing functions for optimizing overall performance) to enable parallel simulation on a distributed network. The random number generators (RNGs) are kept thread-private and are initiated with independent seeds, which are provided by a different type of RNG [e.g., 16 807 RNG (Ref. 30)] in our implementation.

2.B. Validating the C++ version of PENELOPE

Before adapting the C++ implementation onto a GPU, we validated that the C++ version produced identical results to the original FORTRAN code. We set up a simple cone beam incident on a cubic water phantom, ran the two versions with various incident energies, angles, and cutoff energies in single-thread mode, and then exported the particle status of 10^8 serial steps for comparison. Considering the possible runtime library differences between C++ and FORTRAN, we set the allowed error of position, direction, and energy for each step to be 10^{-10} cm, 10^{-10}, and 10^{-10} keV, respectively. We obtained 100% identical step-status outputs, suggesting that our C++ code is completely equivalent to the original FORTRAN code.

2.C. User-friendly features

The original PENELOPE configuration file has strict formatting restrictions that consequently require changes to the source code when adding or deleting certain configuration items. We thus developed an elegant script module that supports "c"-style free writing, declaring nested cells, and macrodefinitions for ease of use. A powerful log module was also developed to manage file records, run batch tasks, implement dose reuse, and provide e-mail notifications. We also developed a binary file manager module powered by a real-time compressing/decompressing algorithm[31] to handle large phantom and dose output files that would otherwise cost a lot of storage space and bandwidth for synchronizing remote simulations.

2.D. Adapting to GPU

Upon confirming the integrity of our c++ code, we proceeded to port the c++ code to CUDA, which is a c-extended GPU programming language that was introduced by NVIDIA in 2007.[32] Though CUDA greatly simplified parallel programming on GPUs, it suffers from two main restrictions in comparison to CPU programming.

First, modern GPUs are designed in single instruction multiple data (SIMD) architecture instead of multiple instruction multiple data (MIMD) due to efficiency and complexity reasons.[33] If threads diverge to different instruction flows through "if" or "switch" statements, the GPU work scheduler will simply execute the instruction flow in series and become less efficient. In general, more deeply nested diverging statements will result in lower efficiency.

Second, the largest device memory (~GB) is only cached by a very small L2 cache, and so cache miss happens frequently and thus causes a long memory accessing latency. Shared and constant memories are hundreds of times faster than device memory, but have a size (~KB) far less than that is necessary for Monte Carlo simulation with large material data tables. To deal with these shortcomings, we designed a dedicated workflow and allocated the scarce fast memory carefully to improve the efficiency of gPENELOPE.

2.D.1. Workflow

In all Monte Carlo codes, instruction divergence is common so we aim to improve efficiency by minimizing the number of nested diverging statements in each CUDA kernel function. Instead of organizing all the simulation codes in one kernel function, we decided to split the code into several independent kernel functions which process different types of scattering events and let the CPU call these kernel functions in a loop within a main function as shown in Fig. 1.

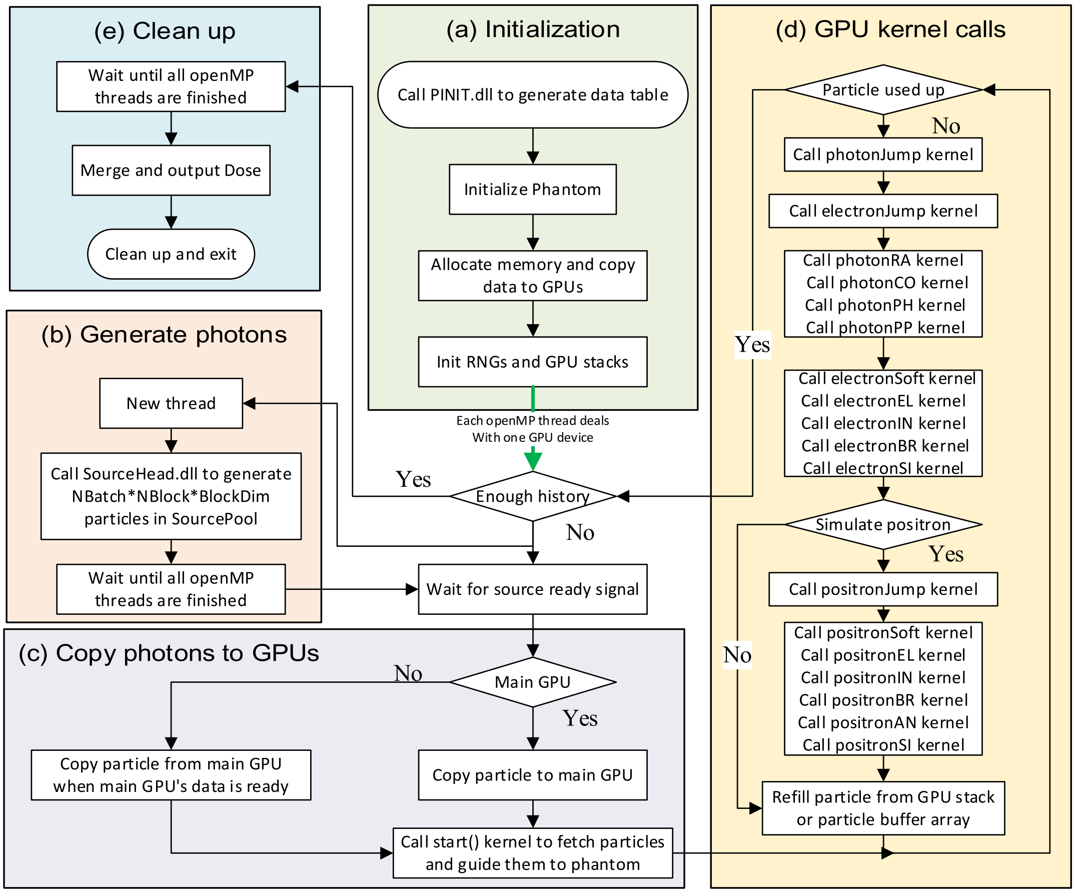


FIG. 1. Workflow of gPENELOPE including (a) initialization, (b) generate photons, (c) copy photons to GPUs, and (d) GPU kernel calls: RA, CO, PH, and PP are short for Rayleigh, Compton, photoelectric, and pair production while EL, IN, BR, SI, and AN are short for elastic, inelastic, bremsstrahlung, shell ionization, and annihilation, respectively. The "kernel-by-kernel" calls in sequence can help reducing instruction divergence. (e) Clean up.

As shown in Fig. 1(a), the program reads and parses the configuration file which includes details regarding the GPU devices, phantom (geometry and materials), and source head. The program then calls a DLL (see Sec. 2.A) to generate necessary data tables for relevant materials. These data and phantom information are then copied to GPU device memory with pointers to large arrays and some small data tables being copied to constant memory on the GPU instead for improving accessing speed. In addition, a random number generator and particle stack are initiated for each GPU thread. As our workstation includes multiple GPU cards, we next launch multiple threads through OpenMP to call GPU kernel functions on each card simultaneously.

Meanwhile, the main thread launches another thread calling the vender-provided head source module to prepare incident photons as shown in Fig. 1(b). As specified in Fig. 1(c), photons are then copied to the main GPU and in turn transferred to other GPUs in order to save I/O time. The GPU kernel function start() is then called to guide photons to the phantom via free propagation. As summarized in Fig. 1(d), distinct "jumping" kernels for photons and charged particles are called to advance relevant particles and label them with the type of interaction that will happen next. These interaction kernel functions are called sequentially such that particles labeled with a different interaction type will simply exit their threads. Though this schedule does not completely resolve the instruction divergence problem, it lowers the level of nested diverging statements and thus reduces the total pausing time. After all interaction kernels finish, we refill the current particle variable either from the stack storing secondary particles or the incident photon buffer array, thus improving the efficiency by enabling constant renewal of particles on all threads.

In addition, we provide an option for toggling positron simulation as the primary photon energy of ^{60}Co is just slightly higher than the threshold for pair production (twice the electron rest mass). We also allow for source particle reuse as occasionally the head model lags the GPU and cannot provide new particles at a sufficient rate. Our simulation comparisons show that the dose differences in "hot areas" ($D > 10\% \times D_{\max}$) caused by reusing source particles are almost totally (99.73%) within the targeted 1% uncertainty for a large (10^9) history number. Moreover, we set up the RNGs to refill their buffers after N loops in order to reduce instruction divergence, where N is the average number of loop iterations when an RNG buffer is exhausted.

2.D.2. Heterogeneous or voxelized tracking

For photon simulation, we added Woodcock tracking[34] to the original PENELOPE to treat a heterogeneous phantom as uniform. In order to obtain an invariant mean free path $\lambda = m_0/(\rho_i\sigma)$ across the whole phantom (m_0 is the molecular mass, ρ_i is the voxel density, and σ is the total scattering cross section), we add a virtual scattering cross section σ_i in each voxel i to maintain $\rho_i(\sigma+\sigma_i) = \rho_{\max}\sigma$ constant everywhere. Then the probability for this virtual scattering to happen is

$$p(\sigma_i) = \frac{\sigma_i}{\sigma+\sigma_i} = 1 - \frac{\rho_i}{\rho_{\max}}. \tag{1}$$

If this virtual interaction is sampled during a knock event, we just continue to propagate the photon without changing direction or losing energy since the virtual event is not real. A shortcoming of this technique is that it could result in low efficiency for a phantom composed mainly of low density material (e.g., lung) because the virtual interaction will most likely be sampled thus wasting random numbers without any energy transfer. We thus instead try to improve the sampling efficiency by forcing the real interactions to always happen, with the secondary particles' weight reduced by factor $\rho_i/\rho_{\max}$ and only $\rho_i/\rho_{\max}$ of the primary photons' status (energy and direction) being changed. This ensures that the probability distribution of deposited energy is unbiased.

For electron and positron simulation, PENELOPE applied the "mixed" condensed history scheme, which treats large energy transfer collisions in an analogue way and uses the continuous slowing down approximation (CSDA)[35] to model small-loss collisions. Since the CSDA range $\bar{s}$ is much smaller than the photon's mean free path λ, we implemented a simple grid detection algorithm to trace the CSDA jump between heterogeneous voxels. Unlike photons, electrons and positrons will cross just a few voxels before being completely stopped. Though soft collisions occur at a high frequency, most of these are determined not to cross the voxel boundary by a rapid test that roughly estimates the nearest distance to the boundary, and so the necessary time for calculating exact crossing points at voxel boundaries is actually not expensive.

2.D.3. Magnetic field

Given that the CSDA range $\bar{s}$ of electrons and positrons is typically very small, the magnetic field in each voxel can be treated as uniform in most applications. The particles will undergo spiral motion in a uniform field **B** at the relativistic angular velocity

$$\vec{\omega} = -\frac{e\vec{B}}{\gamma m_e}, \tag{2}$$

where e denotes the elementary charge, m_e is the electron mass, and γ is the Lorentz factor. The corresponding location after advancing length s in a uniform phantom can be easily evaluated as shown in the PENELOPE user manual to be[1]

$$\vec{r}(s) = \vec{r}_0 + s\hat{v}_0 - \frac{s}{v_0}\vec{v}_{0\perp} + \frac{1}{\omega}[1-\cos(s\omega/v_0)](\hat{\omega}\times\vec{v}_{0\perp}) \\ + \frac{1}{\omega}\sin(s\omega/v_0)\vec{v}_{0\perp}, \tag{3}$$

where r_0 is the initial particle location and v_0 is the particle velocity (with $v_{0\perp}$ being the velocity component perpendicular to **B**). For a heterogeneous voxelized phantom, however, the intersection between the spiral curve and the voxel boundary must be calculated due to the variation of the density in each voxel. Accurate evaluation is messy and inefficient due to many inverse trigonometric function calls. As the CSDA range $\bar{s}$ is relatively small in comparison to the spiral radius

R, we can approximate the spiral motion by small straight line segments that change direction gradually. Taking the allowed error in one segment move to be $\Delta_{\max}$, the maximum segment length s_m is expressed as

$$s_m = \sqrt{(R+\Delta_{\max})^2 - R^2} \approx \sqrt{2R\Delta_{\max}}. \tag{4}$$

If $s > s_m$, we only advance a distance s_m and change direction by angle $\theta \approx s_m/R$ (continuing until s is exhausted). This straight-line advancing procedure uses the same voxel tracking implementation as the situation without magnetic field. While moving a distance, s may cross a voxel boundary, the particle direction may point back to the original voxel when $v \cdot (v + dv) < 0$, and so the current voxel index must be corrected accordingly.

2.E. MRIdian head model

The vendor-provided MRIdian head model provides phase space data including the energy spectrum and flux for a given solid angle for the ^{60}Co source. Each IMRT beam consists of a collection of segments configuring the MLC shape and beam-on time. In our code, each segment is treated as a simulation unit and the history number assigned to each unit is weighted by its beam-on time. The MLC shape determines how many photons will be exported from the ^{60}Co head for each history, which is a nonfixed number due to the patient-specific MLC configuration.

To maximize its efficiency, the GPU should process a fixed number N of photons per batch. Therefore we designed a class to buffer the photons supplied by the head in a multithreading queue such that N photons are fetched in a batch by the GPU when the class is filled with slightly over N photons. The excess photons are then moved to the head of the queue to continue the buffering.

2.F. 3D dose comparisons

For 3D dosimetric evaluation, we considered both gamma indices and statistical histograms to reveal differences between two Monte Carlo systems. The gamma index for each voxel $\vec{r}$ is defined as

$$\gamma(\vec{r}) = \min\{\Gamma(\vec{r},\vec{r}')\}\forall\{\vec{r}'\},$$

$$\Gamma(\vec{r},\vec{r}') = \sqrt{\frac{|\vec{r}-\vec{r}'|^2}{\Delta d^2} + \frac{(D(\vec{r}) - D'(\vec{r}'))^2}{\Delta D^2}}, \tag{5}$$

where $|\vec{r}-\vec{r}'|$ represents the distance between voxels $\vec{r}$ and $\vec{r}'$, Δd is the distance-to-agreement (DTA) value, and ΔD is the dose tolerance value. We label a gamma index at voxel $\vec{r}$ as passing if $\gamma(\vec{r}) \le 1.0$ and count the passing rate for those voxels where $D(\vec{r}) > t \times D_{\max}$, where t is a dose threshold. Higher gamma passing rates for smaller Δd and ΔD tolerances usually suggest stronger agreement between two dose distributions.

A statistical histogram, on the other hand, directly indicates the distribution of dose differences spanning all voxels. Here we define a statistical variable z-score for each voxel as

$$z(\vec{r}) = \frac{D_{\text{test}}(\vec{r}) - D_{\text{ref}}(\vec{r})}{D_{\text{ref}}(\vec{r}) \times \sigma_{\text{tot}}}, \tag{6}$$

where $D_{\text{test}}(\vec{r})$ and $D_{\text{ref}}(\vec{r})$ are the test dose and reference dose at voxel $\vec{r}$, respectively, and σ_{tot} is the standard deviation of the distribution $(D_{\text{test}}(\vec{r}) - D_{\text{ref}}(\vec{r}))/D_{\text{ref}}(\vec{r})$ spanning all voxels. The resultant z-scores are compared to a standard Gaussian distribution in the form of a normalized frequency histogram.

3. RESULTS

3.A. Comparison to C++ PENELOPE

In Sec. 2.B, we showed that C++ PENELOPE performs equivalently to the original code written in FORTRAN by a detailed step-by-step comparison. Here we apply a similar approach to convincingly show that GPENELOPE performs equivalently to C++ PENELOPE in single-thread operation. We simulate a complex lung IMRT plan (shown in Fig. 2) using both platforms and output particle statuses (position, velocity, and energy) in 10^7 knocking steps for comparison. The maximum and average differences are summarized in Table I. The energy difference $|dE|$ is actually negligible in comparison to the incident energy (>1 MeV).

We additionally run 10^6 histories to check differences in dose distributions directly, which turn out to be $\max(|dD|) = 7.63 \times 10^{-5}$ Gy, $\text{mean}(|dD|) = 6.78 \times 10^{-8}$ Gy, $\sigma(dD) = 7.17 \times 10^{-7}$ Gy. The prescription dose for this

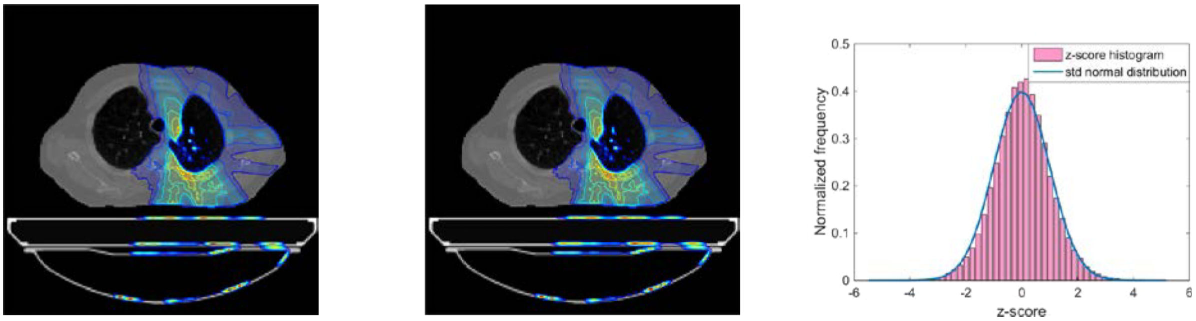


FIG. 2. Dose distributions for a lung case calculated by C++ PENELOPE (left) and GPENELOPE (middle). The two distributions are identical within 1%. The frequency distribution of z-scores in comparison to a standard normal distribution (right).

TABLE I. Particle status differences of 10^7 steps between gPENELOPE and C++ PENELOPE.

$\max(\lvert d\vec{x}\rvert) = 1.42\times10^{-8}$ cm	$\max(\lvert d\hat{v}\rvert) = 1.82\times10^{-9}$	$\max(\lvert dE\rvert) = 2.25\times10^{-5}$ eV
$\mathrm{mean}(\lvert d\vec{x}\rvert) = 4.54\times10^{-14}$ cm	$\mathrm{mean}(\lvert d\hat{v}\rvert) = 1.26\times10^{-14}$	$\mathrm{mean}(\lvert dE\rvert) = 7.29\times10^{-9}$ eV

patient is 50 Gy. That is, the maximum relative dose error is 1.36×10^{-9}. Considering the possible runtime library differences between the GPU and CPU, the status tracking and dose deposition comparisons together show that gPENELOPE and C++ PENELOPE can effectively be considered to be identical in single-thread mode.

Though impractical to compare particle status with gPENELOPE in multithreaded operation (considering thousands of threads simultaneously), we can compare dose distributions generated by gPENELOPE and C++ PENELOPE directly. We thus run a large number of histories (4×10^9) to ensure that the target area ($D > 10\% \times D_{\max}$ region) reaches less than 0.5% uncertainty so that the maximum allowed difference would be less than 1% if gPENELOPE behaves equivalently to C++ PENELOPE. Comparing doses in these voxels, we found that

$$\frac{\max(|dD|)}{\max(D_{\mathrm{ref}})} = 0.93\%, \qquad \frac{\mathrm{mean}(|dD|)}{\max(D_{\mathrm{ref}})} = 0.12\%,$$
$$\frac{\sigma(dD)}{\max(D_{\mathrm{ref}})} = 0.15\%, \tag{7}$$

where D_{ref} is the dose calculated by C++ PENELOPE. The results indicate that the equivalency assertion between gPENELOPE and PENELOPE is valid. In addition, we compared the frequency distribution of z-scores (as defined in Sec. 2.F) to a standard normal distribution as shown in Fig. 2 (right). This comparison indicates that the z-score distribution follows a standard normal distribution.

Since gPENELOPE is effectively equivalent to C++ PENELOPE in single-thread mode, and dose distributions generated by the two agree well within expected statistical uncertainties in multithread operation, we safely deduce that gPENELOPE is a faithful adaptation of PENELOPE that does not compromise accuracy.

The hardware for our tests is a server that includes an Intel Xeon E5 2630 v3 CPU and an NVIDIA Tesla K80 GPU card. The CPU can provide 16 true simultaneous threads, giving an overall processing rate of 1.842×10^5 histories/s. The GPU achieves a simulation rate of 1.756×10^6 histories/s, which is almost 10 times faster than that of the CPU platform. Considering the original PENELOPE engine only supports one thread, our GPU code can actually accelerate PENELOPE by a factor of 152.

For the lung IMRT example above (phantom size $51.3\times50.7\times52.8$ cm^3, voxel size $3\times3\times3$ mm^3), gPENELOPE requires 9.5 min to finish simulating 10^9 histories to achieve 0.5% uncertainty in the $D > 50\% \times D_{\max}$ region. If such a rigorous accuracy is not required, a lot of time could be saved by reducing the history number N as the accuracy is proportional to $\sqrt{N}$. As a platform designed for accurate dose calculation, verification, and accumulation over the treatment course, both the speed and accuracy meet the requirements of clinical applications.

3.B. Magnetic field effects

By integrating an MR scanner into the radiation delivery system, the MRIdian system must consider magnetic field effects on dose distributions. Raaijmakers[36] performed a detailed simulation study using GEANT4 of magnetic field effects on dose distributions for a 6 MV LINAC beam. Although the field strength of the MR scanner on MRIdian is relatively weak (0.35 T), the electron return effect (ERE) might still be nontrivial because the lower energy of a primary photon from the ^{60}Co source tends to result in a smaller spiral radius.

In homogeneous phantoms, dose distortion caused by ERE is generally negligible; however, it will become apparent in heterogeneous phantoms at the interfaces. Here we simulate the radiation delivery for a $10\times10\times16$ cm^3 water–lung–water phantom, where the lung tissue is represented by an 8 cm slab of water with a density of 0.25 g/cm^3 [Fig. 3(a)]. A 4.2×4.2 cm^2 ^{60}Co beam consisting of 10^9 photons was incident on the phantom, and a small dose scoring voxel size was set to $1\times1\times1$ mm^3 to probe for dose distortion. The simulation was repeated with 0.35, 0.75, 1.5, and 3 T magnetic field strengths and the corresponding central axis depth dose profiles were compared as shown in Fig. 3(b). The results are similar to those of Raaijmakers[36] except that the distortion layer is much thinner than for the 6 MV LINAC beam. The dose wash images in the x–z plane are presented in Fig. 3(c). Besides stronger dose accumulation effect, the lateral dose shift will also become more obvious as the magnetic field strength goes up.

The simulation suggests that the 0.35 T magnetic field has a minor effect on the dose distribution in a heterogeneous phantom (spike-shape dose accumulation <3% of max dose within a 3 mm thin layer accompanied by a 1 mm lateral shift), which is consistent with the experimental results of Wooten *et al.* using radiographic film.[20] Wooten *et al.* noted that such perturbation effect would be mitigated by multiple overlapping beams, as in the case of an IMRT plan, for instance. It is interesting that this effect would become almost imperceptible when the voxel size increases to $3\times3\times3$ mm^3, which is the voxel size that most clinics use in treatment planning.

3.C. Experimental measurements

Beyond demonstrating that the gPENELOPE simulation kernel is both fast and accurate, we must confirm that the entire validation system is correctly modeled (especially the ^{60}Co head) in order to ensure safe deployment in the clinic. We

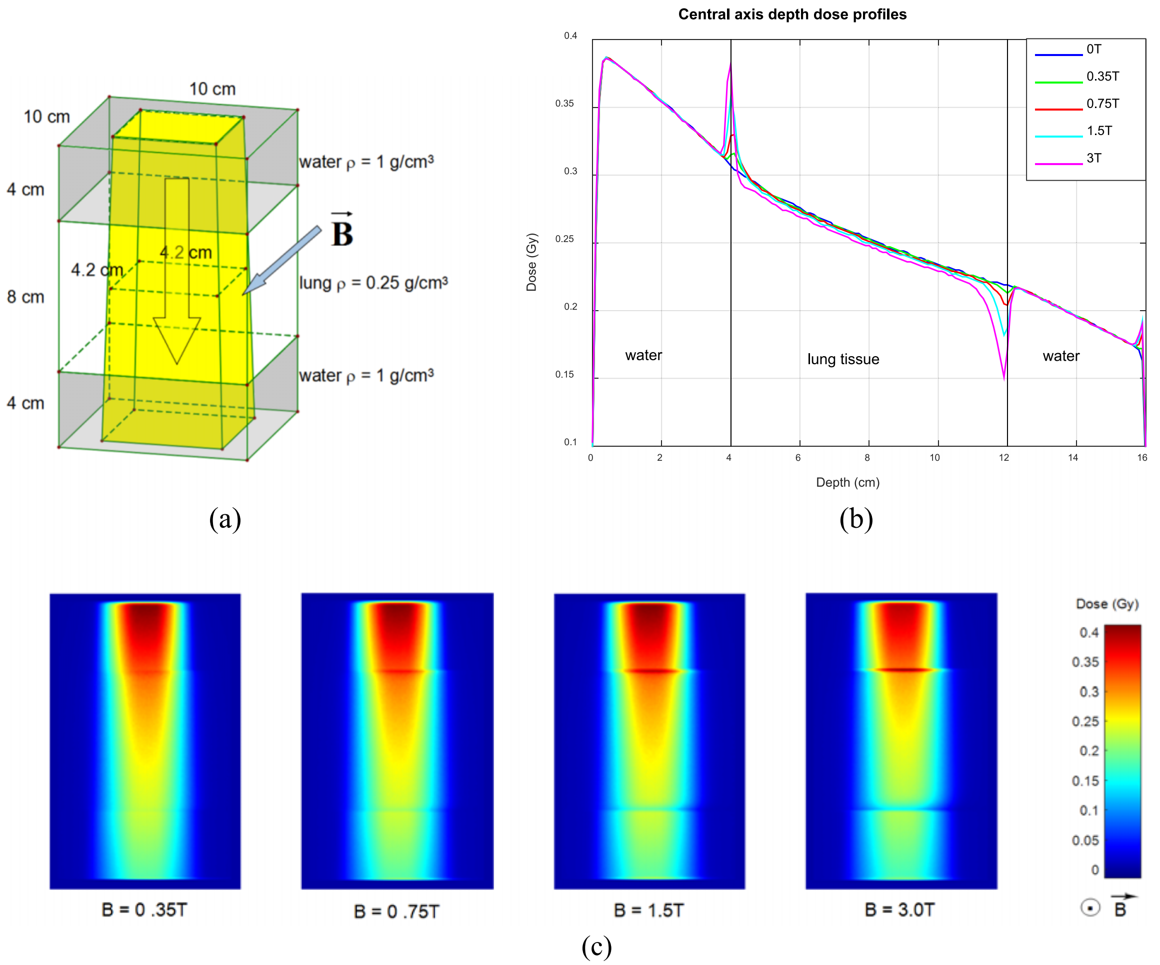


FIG. 3. (a) Water–lung–water phantom, ^{60}Co beam, and magnetic field configuration. Voxel size is 1×1×1 mm^3. (b) Central axis depth dose profiles. Larger magnetic field results in larger dose distortion. (c) Dose distributions for configuration in (a) at indicated magnetic field strengths.

thus investigate several vital comparisons to experimental measurements to validate its overall accuracy.

3.C.1. Depth dose

Measurements were performed in a cubic water phantom (30×30×30 cm^3) placed at SSD = 100 cm using small, medium, and large field sizes (4.2×4.2, 10.5×10.5, and 27.3×27.3 cm^2). The data were collected using an Extradin A18 ion chamber. Note that the chamber is manually positioned at different depths as an MRI compatible beam scanning device is not commercially available now. Considering the cylindrical dimensions of the ion chamber (radius = 2.5 mm, height = 6.4 mm), we set the voxel size of the phantom to be 3×3×3 mm^3 in simulation and run 10^9 histories to ensure sufficiently small statistical uncertainty ($<0.5\%$ for $D > 50\% \times D_{\max}$ region). Figure 4(a) shows comparisons between simulation and experimental data, yielding less than 1% difference.

3.C.2. Off-axis profile

We used EBT2 radiochromic films placed at depths of 5, 10, and 15 cm to sample planar doses for comparisons to simulation results. As shown in Figs. 4(b)–4(d), the simulated dose profiles agree well with measured data to within 2% or 2 mm DTA.

3.C.3. Output factor

Both square and rectangular field output factor measurements were performed in a cubic water phantom (30×30 ×30 cm^3) placed at SSD = 100 cm. The Extradin A18 ion chamber was placed at 5 cm below the surface, i.e., the isocenter. Note that in order to verify the small field output factor, we closed the central ten leaves incrementally from 10.5 to 0.6 cm, as recommended by ViewRay. As shown in Table II, the calculated output factors match well with the experimental data (<2%).

3.C.4. AAPM TG-119

The AAPM Task Group 119 (TG-119)[37] recommends that six cases be considered (two non-IMRT and four IMRT) for IMRT commissioning, including AP-PA, Bands, Multitarget, C-shape, Head and Neck, and Prostate. These treatment plans were planned using MRIdian's inverse treatment planning system and delivered to a 30×30×15 cm^3 water-equivalent

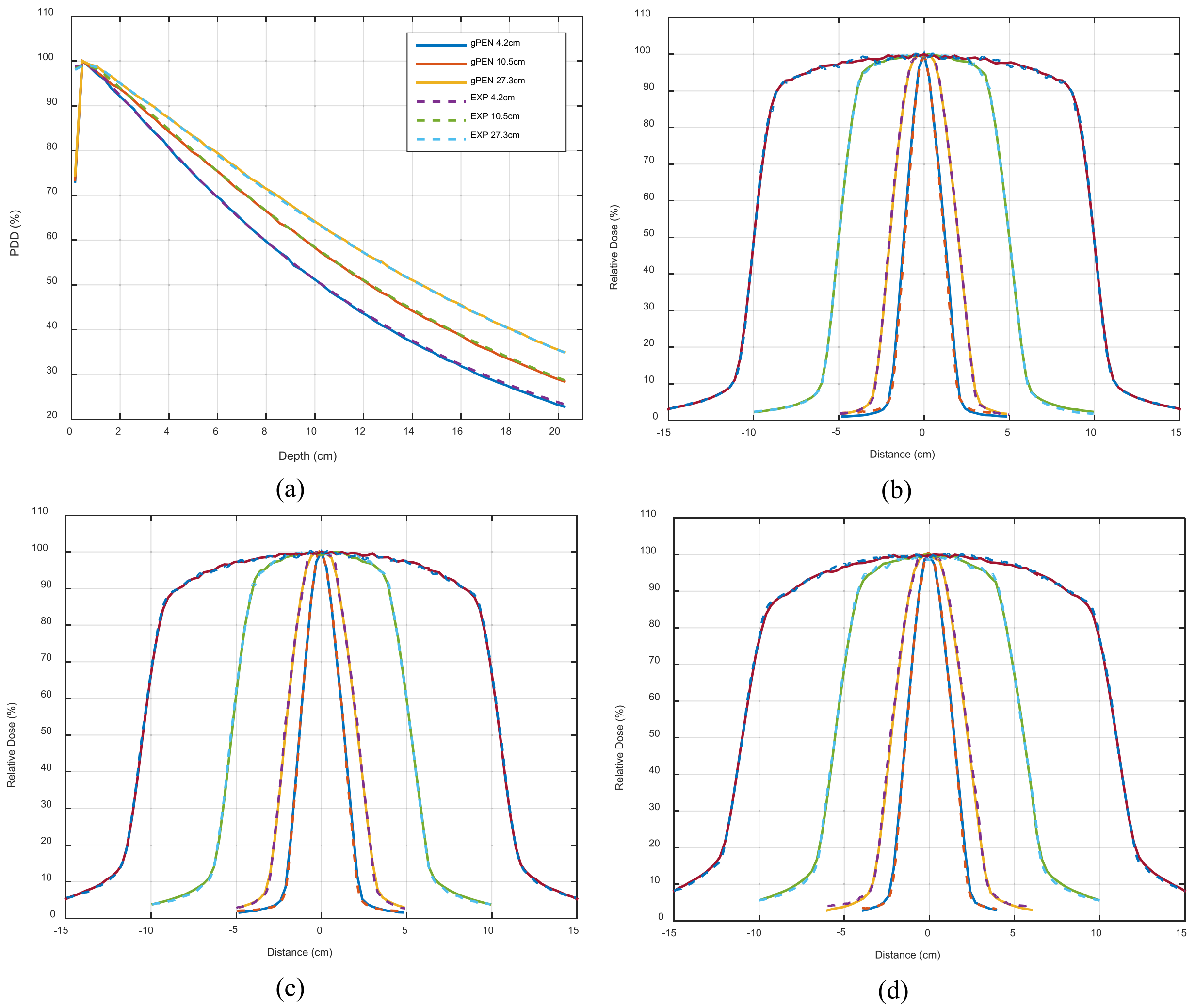


Fig. 4. Percentage-depth-dose comparisons between gPENELOPE and ionization chamber measurement (spline interpolated) for field sizes 4.2×4.2, 10.5×10.5, and 27.3×27.3 cm^2 at 100 cm SSD (a). Off-axis profile comparisons between gPENELOPE and radiochromic film measurement for field sizes 2.1×2.1, 4.2×4.2, 10.5×10.5, and 21.0×21.0 cm^2 at depths of 5 (b), 10 (c), and 15 cm (d) at 100 cm SSD.

plastic phantom containing ionization chambers (ICs) and EDR2 radiographic film.

3.C.4.a. TG-119 point dose. All point dose measurements were made using an ADCL calibrated ionization chamber (Extradin A18). The dose at each plan's isocenter (except C-shape) is measured to evaluate high dose accuracy while a few adjacent points are chosen to examine low dose accuracy.

Table III compares calculated doses from gPENELOPE to experimental measurements. For flat high-dose regions, gPENELOPE gives excellent agreement with measurements (error < 0.31% for non-IMRT plans and error < 2.26% for IMRT plans). All results for the low-dose points are within the TG-119 confidence limit of 4.5%, with Multitarget (4 cm inferior) and C-shape (1 cm posterior) yielding the largest discrepancies. By examining the dose distributions, we find that the two points are located in high-gradient regions [cf. Figs. 5(e) and 5(f)] where small chamber positioning error could induce large measurement difference. For these two cases, we use DTA instead to evaluate gPENELOPE performance where we search around the dose matrix grid with interpolation to find the nearest point that has the exact same dose as measurement, with DTA defined as the distance from this point to the measurement point. Calculated DTAs are less than half of the voxel size (0.91 and 1.42 mm, respectively).

3.C.4.b. TG-119 film dose. For the six plans listed above, radiographic film measurements were made at the isocenter parallel to the coronal plane. Films were digitized and then exported to perform gamma analysis using gamma parameters recommended by TG-119: (a) absolute dose comparison, (b) 3% dose difference threshold, (c) global normalization for percent dose difference, (d) 3 mm DTA threshold, and (e) 10% low dose threshold. The gamma passing rates are 100.0%, 96.2%, 95.5%, 97.7%, 99.9%, and 94.4% for AP-PA, Bands, C-shape, Head and neck, Multitarget, and Prostate cases, respectively, yielding a mean value of 97.3% ± 2.3% (1 SD), which is within the TG-119 recommended confidence limit of 88%. Figure 5 summarizes the agreement between gPENELOPE and experiment using isodose line overlay. For IMRT plans, only the low dose contours (around 10%) show relatively obvious disagreements.

Table II. Output factor comparison for square and rectangular fields.

Field shape	Size (cm^2)	OF (gPEN)	OF (expt.)	Diff. (%)
Square field	4.2 × 4.2	0.8839	0.8780	0.67
	6.3 × 6.3	0.9414	0.9380	0.36
	10.5 × 10.5	1.0000	1.0000	NA
	14.7 × 14.7	1.0293	1.0410	−1.12
	27.3 × 27.3	1.0624	1.0700	−0.71
Rectangular field	0.6 × 10.5	0.2103	0.2070	1.58
	0.8 × 10.5	0.2839	0.2825	0.51
	1.0 × 10.5	0.3607	0.3568	1.08
	1.5 × 10.5	0.5256	0.5246	0.19
	2.0 × 10.5	0.6741	0.6721	0.30
	2.5 × 10.5	0.7953	0.7859	1.19
	3.0 × 10.5	0.8730	0.8583	1.71
	4.0 × 10.5	0.9222	0.9119	1.13
	6.0 × 10.5	0.9636	0.9582	0.56
	8.0 × 10.5	0.9837	0.9822	0.16
	10.5 × 10.5	1.0000	1.0000	NA

3.D. Comparison to MRIdian treatment planning system

The KMC algorithm on the MRIdian TPS adopted many approximations and variance reductions[38] to increase calculation speed. KMC's accuracy should thus be confirmed using a third-party Monte Carlo system devoid of approximations through 3D dose comparisons. Thus we selected 16 recent patient plans (from the ViewRay patient registry at Washington University in St. Louis) created by the MRIdian TPS with treatment sites including stomach (4), lung (2), liver (3), adrenal gland (2), pancreas (2), spleen (1), mediastinum (1), and breast (1). Three-dimensional gamma analysis results (2%/2 mm DTA and 10% threshold criteria) and histograms of z-scores (in comparison to standard normal distributions) are listed in Table IV. The table shows that KMC matches gPENELOPE well (15 out of 16 plans with dose gamma passing rates ≥98% and most closely fitting Gaussian distributions) except that KMC occasionally tends to result in a little higher dose than gPENELOPE. Some z-score distributions (second lung case, first pancreas case, and the breast case) are noticeably offset from the standard distribution, thus indicating that the physical modeling is somewhat affected by the approximations and variance reductions implemented by KMC for calculating a complex ^{60}Co IMRT plan. The statistical gamma passing rates are as high as 99.1% ± 0.6% for the two dose distribution, proving that KMC generally predicts dose consistent with our accuracy-oriented Monte Carlo engine. During MRIdian's commissioning, Wooten *et al.* designed a custom heterogeneity phantom to acquire ionization chamber measurements.[20] They report that the mean ionization chamber measured dose for 27 measurements for 5 plans is within 1% vs KMC.

4. DISCUSSION AND CONCLUSION

The recent clinical use of the MRIdian radiation therapy system represents a significant advance in cancer care, enabling clinicians, for the first time, to deliver highly conformal IMRT with real-time MRI guidance. However, the rapid advances in the technology to deliver such radiation treatments seem to have not been paralleled by corresponding advances in the ability to verify these treatments subject to a permanent magnetic field. For conventional IMRT, despite its widespread utilization at modern radiation therapy clinics,

Table III. TG-119 point dose comparisons: gPENELOPE vs IC.

TG-119 plans	Location	IC (Gy)	gPENELOPE (Gy)	Diff. (%)
AP-PA	Isocenter	1.988	1.991	0.16
Bands	Isocenter	1.422	1.426	0.31
Multitarget	Isocenter	2.085	2.058	−1.29
Multitarget	4 cm superior	1.062	1.038	−2.22
Multitarget	4 cm inferior	0.621	0.593	−4.43
C-shape	2.5 cm anterior	2.152	2.131	−0.96
C-shape	1 cm posterior	0.917	0.882	−3.77
Head and neck	Isocenter	2.215	2.265	2.26
Head and neck	5 cm posterior	0.917	0.919	0.27
Prostate	Isocenter	1.817	1.85	1.82
Prostate	4.5 cm posterior	0.372	0.374	0.62

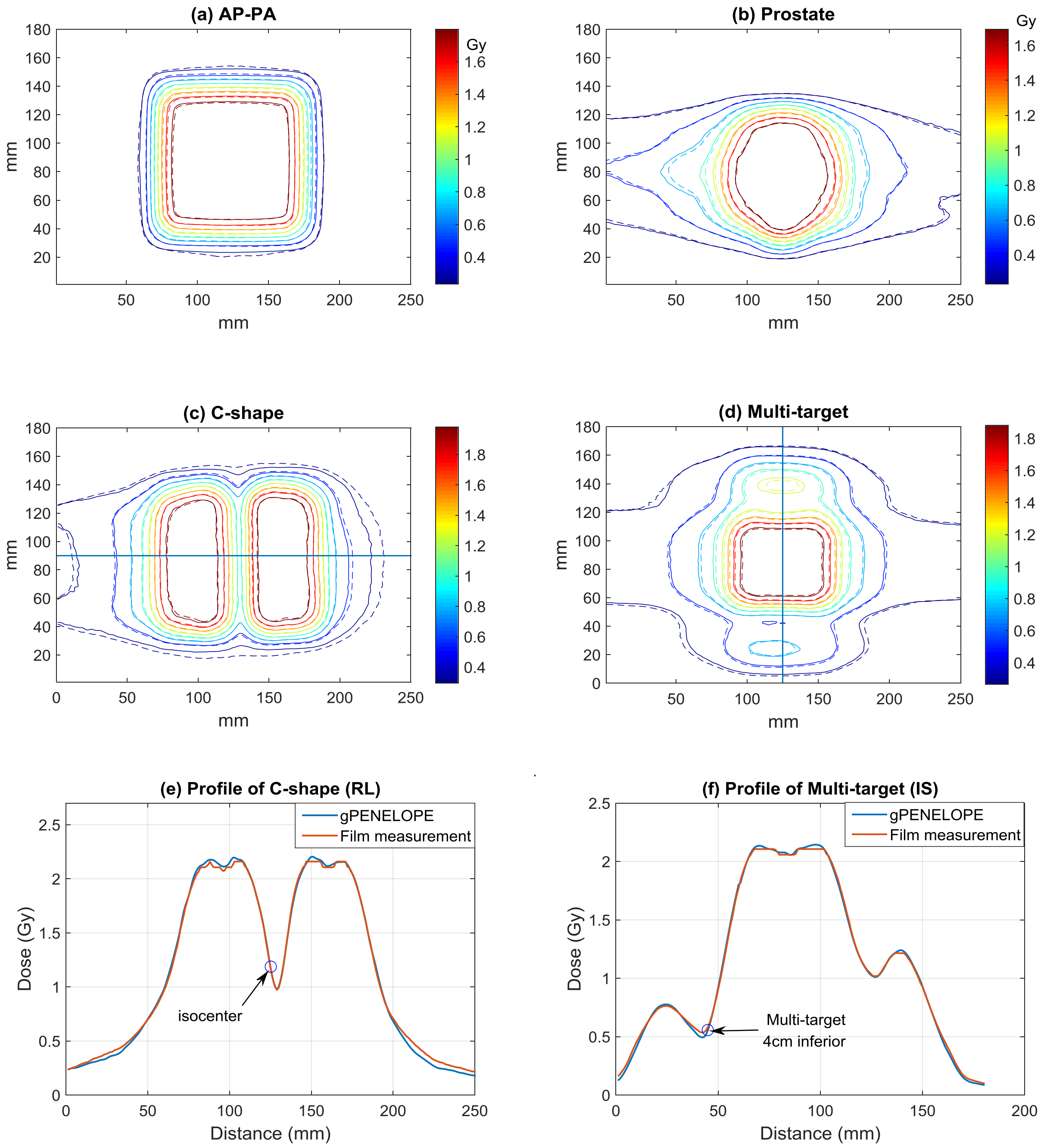
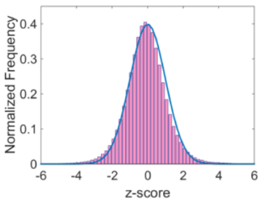
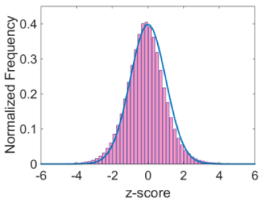
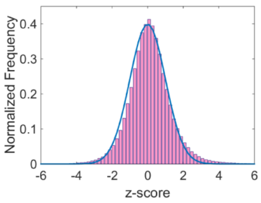
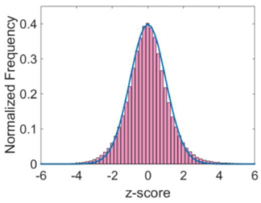
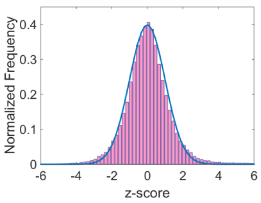
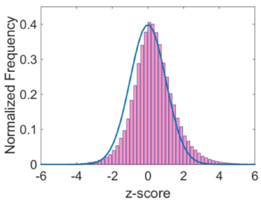
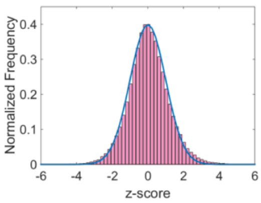
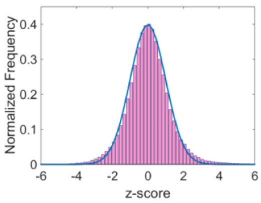
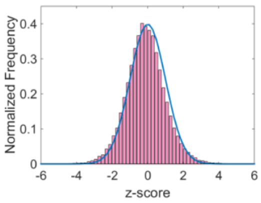
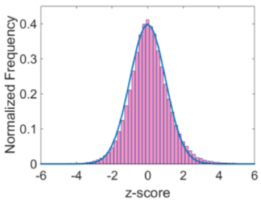
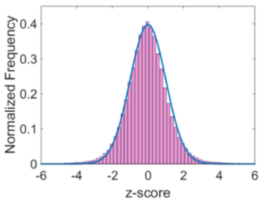
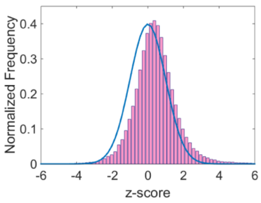
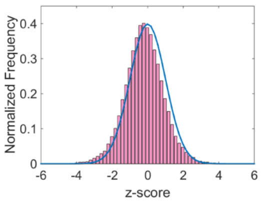
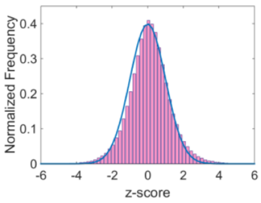
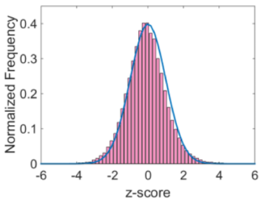
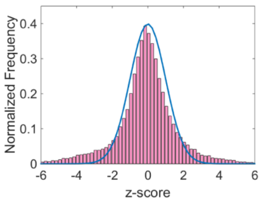


FIG. 5. Isodose and profile comparison between gPENELOPE and radiographic film measurements: (a) AP-PA, (b) Prostate, (c) C-shape, and (d) Multitarget, where solid lines represent gPENELOPE and dashed lines represent film measurement. (e) Profile of C-shape along the right-to-left central axis. (f) Profile of Multitarget along the inferior-to-superior central axis. Note that the circled points are located in the high gradient region.

precise dosimetric commissioning remains a challenge.[39] In the era of MRI-guided IMRT, the permanent magnetic field is augmenting another dimension of error and uncertainty to the already error-prone IMRT process.

As a result of many limitations to experimental approaches, largely due to the dearth of appropriate multidimensional water-equivalent dosimeters, a hybrid approach that includes a computational component is needed for MRI-IMRT commissioning and validation. For example, Ding *et al.*[40] studied the feasibility of using a Monte Carlo method to commission stereotactic radiosurgery beams shaped by micro multileaf collimators. This hybrid approach is especially valuable for MRI-IMRT where the Monte Carlo method may be the only method that is capable of dealing with complex dose deposition in a heterogeneous medium subject to a magnetic field.[18,41] The Monte Carlo methods like KMC, on the other hand, may require many approximations in order to be practical in the clinic, and these approximations may not be thoroughly communicated to an end-user for proprietary reasons. We therefore developed a fast, GPU-accelerated Monte Carlo dose calculation system based on PENELOPE. Unlike some other GPU implementations, the accuracy of our adaptation is at the same level as the original code. Our implementation achieved 152 times faster

Table IV. Gamma passing rates (2%/2 mm and 10% threshold) and z-score distributions comparing gPENELOPE and MRIdian's KMC.

			
Stomach γ = 97.7%	Stomach γ = 98.6%	Stomach γ = 99.0%	Stomach γ = 99.4%
			
Lung γ = 99.6%	Lung γ = 98.8%	Liver γ = 99.7%	Liver γ = 99.3%
			
Liver γ = 99.9%	Adrenal γ = 99.1%	Adrenal γ = 98.5%	Pancreas γ = 98.7%
			
Pancreas γ = 98.5%	Spleen γ = 99.4%	Mediastinum γ = 99.2%	Breast γ = 99.4%

speed than that of the original PENELOPE implementation. Furthermore, we integrated the ^{60}Co head model of the MRIdian system into our system and performed a series of experimental benchmarks to examine the accuracy of the entire system. Finally, when comparing to MRIdian's KMC for a number of patients that span multiple disease sites, an average of 99.1% ± 0.6% gamma passing rates at 2%/2 mm provides another layer of confidence in treating patients that may benefit from IMRT with simultaneous MRI guidance.

In the clinic, gPENELOPE should be applicable to nearly any application requiring high dose accuracy, such as beam modeling,[21] IMRT optimization,[42] dosimeter response modeling,[28,43] dose validation,[19] dose accumulation,[44] among others. As an example, due to the three-source nature of the MRIdian system, quasi-3D dosimeters, such as ArcCHECK (Sun Nuclear Corp., Melbourne, FL), are quite useful for dosimetry measurements. However, the combined field size dependence and angular dependence of an ArcCHECK have been reported to be on the order of 10%–15% for a LINAC delivery. This can be corrected by using look-up tables as a function of beam angle and field size, for which the beam angle must first be determined using a virtual inclinometer in the ArcCHECK software. However, this cannot be corrected for the MRIdian system due to the simultaneous delivery of all three sources. One possibility to solve this problem is to model the dosimeter response using gPENELOPE so that the radiation transport in the diodes and surrounding buildup/backscatter material can be explicitly simulated. As a result, dose to individual diodes instead of to water can be calculated and subsequently compared to diode's raw response during measurements. By doing this, we cannot only convert the ArcCHECK from a relative, 3D dosimeter to an absolute one; more importantly, tighter criteria can be used for the gamma analysis, for example, 2%/2 mm. Nelms *et al.*[45] have recently made a convincing case that adoption of more sensitive metrics/tighter tolerances enables continual improvement of the accuracy of radiation therapy dose delivery not only at the end-user level but also at the level of product design by the manufacturer. This is especially important for MRI-guided IMRT which is at the early stage of its clinical implementation.

In conclusion, a GPU version of PENELOPE has been developed with its accuracy completely faithful to the original code. The comparisons with MRIdian dose calculation results suggest that MRIdian's fast dose calculation for the ^{60}Co source subject to a 0.35 T magnetic field is accurate using 2%/2 mm criteria. gPENELOPE will be useful for many MRI–IMRT applications including dose validation and accumulation, IMRT optimization, and dosimetry system modeling.

ACKNOWLEDGMENTS

The authors graciously acknowledge Dr. Iwan Kawrakow and Dr. James Dempsey of ViewRay, Inc. for many discussions on this subject. This work was supported in part by NIH Grant No. R01CA148853.

CONFLICT OF INTEREST DISCLOSURE

The authors have no COI to report.

[a] Author to whom correspondence should be addressed. Electronic mail: hli@radonc.wustl.edu

[1] I. J. Chetty, B. Curran, J. E. Cygler, J. J. DeMarco, G. Ezzell, B. A. Faddegon, I. Kawrakow, P. J. Keall, H. Liu, C. M. Ma, D. W. Rogers, J. Seuntjens, D. Sheikh-Bagheri, and J. V. Siebers, "Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning," Med. Phys. **34**, 4818–4853 (2007).

[2] D. W. Rogers, "Fifty years of Monte Carlo simulations for medical physics," Phys. Med. Biol. **51**, R287–R301 (2006).

[3] J. F. Briesmeister, "MCNP—A general Monte Carlo N-particle transport code," Los Alamos National Laboratory Report LA-12625-M, 1993.

[4] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. J. G. Cadenas, I. Gonzalez, G. G. Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampen, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. M. de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Seil, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. S. Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J. P. Wellisch, T. Wenaus, D. C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche, "GEANT4—A simulation toolkit," Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250–303 (2003).

[5] I. Kawrakow and D. W. O. Rogers, *"The EGSnrc Code System: Monte Carlo Simulation of Electron and Photon Transport. Ionizing Radiation Standards," Report No. PIRS-701* (NRC, Ottawa, Ontario, 2003).

[6] I. Kawrakow, "Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version," Med. Phys. **27**, 485–498 (2000).

[7] F. Salvat, "The PENELOPE code system. Specific features and recent improvements," Ann. Nucl. Energy **82**, 98–109 (2015).

[8] J. Baro, J. Sempau, J. M. Fernandezvarea, and F. Salvat, "PENELOPE—An algorithm for Monte-Carlo simulation of the penetration and energy-loss of electrons and positrons in matter," Nucl. Instrum. Methods Phys. Res., Sect. B **100**, 31–46 (1995).

[9] G. Pratx and L. Xing, "GPU computing in medical physics: A review," Med. Phys. **38**, 2685–2697 (2011).

[10] I. Kawrakow, M. Fippel, and K. Friedrich, "3D electron dose calculation using a voxel based Monte Carlo algorithm (VMC)," Med. Phys. **23**, 445–457 (1996).

[11] M. Fippel, "Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm," Med. Phys. **26**, 1466–1475 (1999).

[12] J. Gardner, J. Siebers, and I. Kawrakow, "Dose calculation validation of VMC++ for photon beams," Med. Phys. **34**, 1809–1818 (2007).

[13] J. Sempau, S. J. Wilderman, and A. F. Bielajew, "DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations," Phys. Med. Biol. **45**, 2263–2291 (2000).

[14] X. Jia, X. Gu, J. Sempau, D. Choi, A. Majumdar, and S. B. Jiang, "Development of a GPU-based Monte Carlo dose calculation code for coupled electron-photon transport," Phys. Med. Biol. **55**, 3077–3086 (2010).

[15] X. Jia, X. Gu, Y. J. Graves, M. Folkerts, and S. B. Jiang, "GPU-based fast Monte Carlo simulation for radiotherapy dose calculation," Phys. Med. Biol. **56**, 7017–7031 (2011).

[16] S. Hissoiny, B. Ozell, H. Bouchard, and P. Despres, "GPUMCD: A new GPU-oriented Monte Carlo dose calculation platform," Med. Phys. **38**, 754–764 (2011).

[17] L. Jahnke, J. Fleckenstein, F. Wenz, and J. Hesser, "GMC: A GPU implementation of a Monte Carlo dose calculation based on GEANT4," Phys. Med. Biol. **57**, 1217–1229 (2012).

[18] G. H. Bol, S. Hissoiny, J. J. Lagendijk, and B. W. Raaymakers, "Fast online Monte Carlo-based IMRT planning for the MRI linear accelerator," Phys. Med. Biol. **57**, 1375–1385 (2012).

[19] H. Li, V. L. Rodriguez, O. L. Green, Y. Hu, R. Kashani, H. O. Wooten, D. Yang, and S. Mutic, "Patient-specific quality assurance for the delivery of Co intensity modulated radiation therapy subject to a 0.35-T lateral magnetic field," Int. J. Radiat. Oncol., Biol., Phys. **1**, 65–72 (2015).

[20] H. O. Wooten, O. Green, H. Li, V. Rodriguez, and S. Mutic, "Measurements of the electron-return-effect in a commercial magnetic resonance image guided radiation therapy system. WE-G-17A-4," in *AAPM Annual Meeting* (Austin, TX, 2014).

[21] E. Gete, C. Duzenli, M. P. Milette, A. Mestrovic, D. Hyde, A. M. Bergman, and T. Teke, "A Monte Carlo approach to validation of FFF VMAT treatment plans for the TrueBeam Linac," Med. Phys. **40**, 021707 (13pp.) (2013).

[22] T. Teke, A. M. Bergman, W. Kwa, B. Gill, C. Duzenli, and I. A. Popescu, "Monte Carlo based, patient-specific RapidArc QA using Linac log files," Med. Phys. **37**, 116–123 (2010).

[23] J. Sempau, A. Sanchez-Reyes, F. Salvat, H. O. ben Tahar, S. B. Jiang, and J. M. Fernandez-Varea, "Monte Carlo simulation of electron beams from an accelerator head using PENELOPE," Phys. Med. Biol. **46**, 1163–1186 (2001).

[24] J. Sempau, J. M. Fernández-Varea, E. Acosta, and F. Salvat, "Experimental benchmarks of the Monte Carlo code PENELOPE," Nucl. Instrum. Methods Phys. Res., Sect. B **207**(2), 107–123 (2003).

[25] J. Sempau, P. Andreo, J. Aldana, J. Mazurier, and F. Salvat, "Electron beam quality correction factors for plane-parallel ionization chambers: Monte Carlo calculations using the PENELOPE system," Phys. Med. Biol. **49**, 4427–4444 (2004).

[26] S. J. Ye, I. A. Brezovich, P. Pareek, and S. A. Naqvi, "Benchmark of PENELOPE code for low-energy photon transport: Dose comparisons with MCNP4 and EGS4," Phys. Med. Biol. **49**, 387–397 (2004).

[27] B. A. Faddegon, I. Kawrakow, Y. Kubyshin, J. Perl, J. Sempau, and L. Urban, "The accuracy of EGSnrc, GEANT4 and PENELOPE Monte Carlo systems for the simulation of electron scatter in external beam radiotherapy," Phys. Med. Biol. **54**, 6151–6163 (2009).

[28] M. Reynolds, B. G. Fallone, and S. Rathee, "Dose response of selected solid state detectors in applied homogeneous transverse and longitudinal magnetic fields," Med. Phys. **41**, 092103 (12pp.) (2014).

[29] PGI CUDA FORTRAN Compiler (2010), available at http://www.pgroup.com/resources/cudafortran.htm.

[30] W. H. Payne, J. R. Rabung, and T. P. Bogyo, "Coding Lehmer pseudorandom number generator," Commun. ACM **12**, 85–86 (1969).

[31] Y. Collet, "LZ4-Extremely fast compression" (2015), available at https://github.com/Cyan4973/lz4.

[32] Nvidia, NVIDIA CUDA C Programming Guide, 2011.

[33] D. Van Antwerpen, "Improving SIMD efficiency for parallel Monte Carlo light transport on the GPU," *Proceedings of High Performance Graphics*, *2011*.

[34] E. Woodcock, T. Murphy, P. Hemmings, and S. Longworth, "Techniques used in the GEM code for Monte Carlo neutronics calculation," *Proceedings of the Conference on the Applications of Computing Methods to Reactors ANL-7050*, *1965*.

[35] K. Kowari, "Validity of the continuous-slowing-down approximation in electron degradation, with numerical results for argon," Phys. Rev. A **41**, 2500–2505 (1990).

[36] A. J. Raaijmakers, B. W. Raaymakers, and J. J. Lagendijk, "Magnetic-field-induced dose effects in MR-guided radiotherapy systems: Dependence on the magnetic field strength," Phys. Med. Biol. **53**, 909–923 (2008).

[37] G. A. Ezzell, J. W. Burmeister, N. Dogan, T. J. LoSasso, J. G. Mechalakos, D. Mihailidis, A. Molineu, J. R. Palta, C. R. Ramsey, B. J. Salter, J. Shi, P. Xia, N. J. Yue, and Y. Xiao, "IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119," Med. Phys. **36**, 5359–5373 (2009).

[38] I. Kawrakow and M. Fippel, "Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC," Phys. Med. Biol. **45**, 2163–2183 (2000).

[39] A. Molineu, N. Hernandez, T. Nguyen, G. Ibbott, and D. Followill, "Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom," Med. Phys. **40**, 022101 (8pp.) (2013).

[40] G. X. Ding, D. M. Duggan, and C. W. Coffey, "Commissioning stereotactic radiosurgery beams using both experimental and theoretical methods," Phys. Med. Biol. **51**, 2549–2566 (2006).

[41] J. J. Lagendijk, B. W. Raaymakers, C. A. Van den Berg, M. A. Moerland, M. E. Philippens, and M. van Vulpen, "MR guidance in radiotherapy," Phys. Med. Biol. **59**, R349–R369 (2014).

[42] N. Dogan, J. V. Siebers, P. J. Keall, F. Lerma, Y. Wu, M. Fatyga, J. F. Williamson, and R. K. Schmidt-Ullrich, "Improving IMRT dose accuracy via deliverable Monte Carlo optimization for the treatment of head and neck cancer patients," Med. Phys. **33**, 4033–4043 (2006).

[43] A. Palm, A. S. Kirov, and T. LoSasso, "Predicting energy response of radiographic film in a 6 MV x-ray beam using Monte Carlo calculated fluence spectra and absorbed dose," Med. Phys. **31**, 3168–3178 (2004).

[44] D. A. Jaffray, P. E. Lindsay, K. K. Brock, J. O. Deasy, and W. A. Tome, "Accurate accumulation of dose for improved understanding of radiation effects in normal tissue," Int. J. Radiat. Oncol., Biol., Phys. **76**, S135–S139 (2010).

[45] B. E. Nelms, M. F. Chan, G. Jarry, M. Lemire, J. Lowden, C. Hampton, and V. Feygelman, "Evaluating IMRT and VMAT dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels," Med. Phys. **40**, 111722 (15pp.) (2013).