RIDOS: A new system for online computation of the delivered dose distributions in scanning ion beam therapy

RIDOS: A NEW SYSTEM FOR ONLINE COMPUTATION OF THE DELIVERED DOSE DISTRIBUTIONS IN SCANNING ION BEAM THERAPY

S. Giordanengo[a], A. Vignati[a], A. Attili[a], M. Ciocca[b], M. Donetti[b], F. Fausti[a,c], L.Manganaro[a,e], F. M. Milian[d,e], S. Molinelli[b], V. Monaco[a,e], G. Russo[a,], R. Sacchi[a,e], M. Varasteh[a,e,†], R. Cirio[a,e]*

a. Istituto Nazionale di Fisica Nucleare, Via Giuria 1, 10125 Torino, Italy
b. Centro Nazionale di Adroterapia Oncologica, Strada Campeggi 53, 27100 Pavia, Italy
c. Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy
d. Universidade Estadual de Santa Cruz, Rod Jorge Amado, km16, 45652900 Ilheus, Brazil
e. Università di Torino, Via Giuria 1, 10125 Torino, Italy

* now at EuriX, Torino, Italy
† now at CERN, Geneva, Switzerland

Corresponding author. E-mail address: simona.giordanengo@to.infn.it

ABSTRACT

Purpose: To describe a new system for scanned ion beam therapy, named RIDOS (Real-time Ion DOse planning and delivery System), which performs real time delivered dose verification integrating the information from a clinical beam monitoring system with a Graphic Processing Unit (GPU) based dose calculation in patient Computed Tomography.

Methods: A benchmarked dose computation algorithm for scanned ion beams has been parallelized and adapted to run on a GPU architecture. A workstation equipped with a NVIDIA GPU has been interfaced through a National Instruments PXI-crate with the dose delivery system of the Italian National Center of Oncological Hadrontherapy (CNAO) to receive in real-time the measured beam parameters. Data from a patient monitoring system are also collected to associate the respiratory phases with each spot during the delivery of the dose. Using both measured and planned spot properties, RIDOS evaluates during the few seconds of inter-spill time the cumulative delivered and prescribed dose distributions and compares them through a fast γ-index algorithm.

Results: The accuracy of the GPU-based algorithms was assessed against the CPU-based ones and the differences were found below 1‰. The cumulative planned and delivered doses are computed at the end of each spill in about 300ms, while the dose comparison takes approximatively 400ms. The whole operation provides the results before the next spill starts.

Conclusions: RIDOS system is able to provide a fast computation of the delivered dose in the inter-spill time of the CNAO facility and allows to monitor online the dose deposition accuracy all along the treatment.

1. Introduction

The possibility of a precise dose localization makes scanning ion beam therapy favourable for highly conformal radiotherapy treatments, but also sensitive to uncertainties due to in-patient particle range variations, patient inter-fractional and intra-fractional motions, and interplay effects between moving targets and the time-dependent pencil beam delivery [1]. Therefore, treating targets like lung that are subject to considerable intra-fractional motion with scanned ion beams [2] is a major and very complex problem, which still requires progress in dose planning, delivery and verification methods.

As an example, many efforts have been made to minimize the severe interplay effects on lung cancer dose distributions [3], such as internal target volume covering range changes [4,5], voluntary breath hold [6], gating [7–10] , tumor tracking [11–13], re-scanning [14–17], and 4-dimensional (4D) treatment planning and delivery optimization [18–21].

As gantry-mounted Cone Beam Computed Tomography (CBCT) scanners are currently available on the market, offline updates of patient anatomy for daily or weekly adaptive purposes are now feasible [22,23]. Furthermore, more challenging adaptive radiotherapy methods and online approaches [24] are evolving quickly, largely as a result of breakthrough in computational [25] and delivery technologies [26]. To tackle most of the mentioned strategies new tools for fast segmentation, dose simulations, dose optimization, cumulative dose estimation are needed with the ultimate goal of automatic re-planning on the fly [27].

However, evident intermediate steps as software and hardware tools to perform online dose delivered verification are mandatory.

Currently, the effects of beam delivery uncertainties on the delivered dose can be evaluated only after treatment by using the computation of dose distributions based on the number of delivered particles and the beam positions measured during treatment. These data are usually saved in dedicated log-files [28,29]. The delivered dose is then compared with the one computed using the planned values by means of γ-index algorithm.

Due to the recent development of GPU-based pencil beam algorithms, a sub-second dose computation is today feasible [30]. Nevertheless, fast dose computation algorithms always have to face the trade-off between computing time and precision. Fast Monte Carlo (MC) based dose computations show a great potential to speed up and improve planning capabilities, as well as to perform independent pre- and post-treatment dose computations, and were recently proposed to replace measurements of dose verification [29,31]. However, the expected clinical enhancement from the new GPU-based MC dose computations does not yet include online applications, due to the fact that obtaining accurate results requires tens of seconds [32–34].

The aim of this paper is to describe a new system, named RIDOS (Real-time Ion DOse planning and delivery System) for online dose computation, which exploits software and hardware technologies aiming at real time dose monitoring system for pencil beam scanning therapy. The work presented here shows the RIDOS tools, which interface fast dose computation with a clinical dose delivery system to perform spill by spill dose computation

for scanned ion beam therapy performed with a synchrotron. These include the pre-treatment and online operations, the hardware and firmware to interface clinical devices and the dedicated Graphic User Interface (GUI). Additionally, the methods adopted for GPU-based codes validation and a sample of preliminary results are presented.

2. Materials and Methods

The RIDOS system has been designed to be installed at the Italian National Center of Oncological Hadrontherapy (CNAO) [35], a synchrotron-based facility with a beam delivery time structure characterized by about 1 second of spill time (beam on) followed by 2 to 5 seconds of inter-spill time (beam off).

The core of RIDOS consists of a GPU-based Fast Dose Computation (F-DC) tool, running on a dedicated high-performance workstation (WS), for the computation in a few seconds of the dose distribution delivered to the patient. Moreover, RIDOS includes a National Instruments (NI Austin, USA) PCI eXtensions (PXI) system to interface the WS with two clinical devices: the CNAO Dose Delivery System (DDS) [36] and the commercial Anzai AZ-733V respiratory gating system (Anzai Medical, Tokyo, Japan) [7,37]. The latter has been foreseen to implement in future real-time dose distribution check able to account for interplay effects and anatomic deformations.

The F-DC receives at the end of each CNAO spill the measurements of the pencil beam characteristics as measured by the beam monitors of the CNAO DDS [38], together with a list of the corresponding patient respiratory phases for each spot when the Anzai patient tracking system is available. The latter is added to build a system able to select the proper CT among a sequence of CT images. Thus, accounting for beam deviations (now) and for interplay effects and anatomic deformations (in future), RIDOS reconstructs the planned and delivered dose distributions for each spill and compares them by means of a GPU-based γ-index algorithm. In addition, a Fast Image Deformation (F-ID) algorithm has been developed (section 2.2.2 and 2.4.3) and tested independently (see section 3.4) to warp the dose of each spill onto points of the reference CT used by the Treatment Planning System (TPS) with the aim of computing a cumulative dose distribution to be compared with the planned one. A F-ID algorithm has been developed to foresee the use of RIDOS with CT images synchronized with the corresponding respiratory phases, referred to as breathing phase CTs. However, the use of synchronized CT images with F-ID has yet to be integrated into the whole system and tested.

The operations performed by RIDOS to provide the online dose computation and comparison can be divided in two main groups based on the execution time, as shown in Fig.1. The first group includes the tasks performed on the standard CPU before the beginning of the treatment (pre-treatment tasks). Considering that RIDOS cannot know in advance how many and which spots will be treated in each spill, only few tasks can be anticipated. These include setup and clinical data loading and initializations, the creation of the planned pencil beam properties (number of particles and direction), the image registration between each breathing CT-phase and the reference CT, the loading and interpolation of the corresponding

deformation vector fields. As reference CT we have considered the CT image at the end of exhalation among the CT images of a 4DCT scan; the remaining CT images will be synchronized with the corresponding respiratory phases and are here referred to as breathing phase CTs.

The second group includes all the other tasks, which are performed online during the spill and inter-spill time slots. Some of these operations take advantage of the GPU architecture to shorten the execution time, while some others rely on NI PXI platform to provide, in real time, readout and communication data.

The flowchart of Fig. 1 shows an overview of the RIDOS operations and when these occur with respect to the beam delivery. The hardware architecture is introduced in the next section followed by the description of each task, also summarized in Tab. 1.

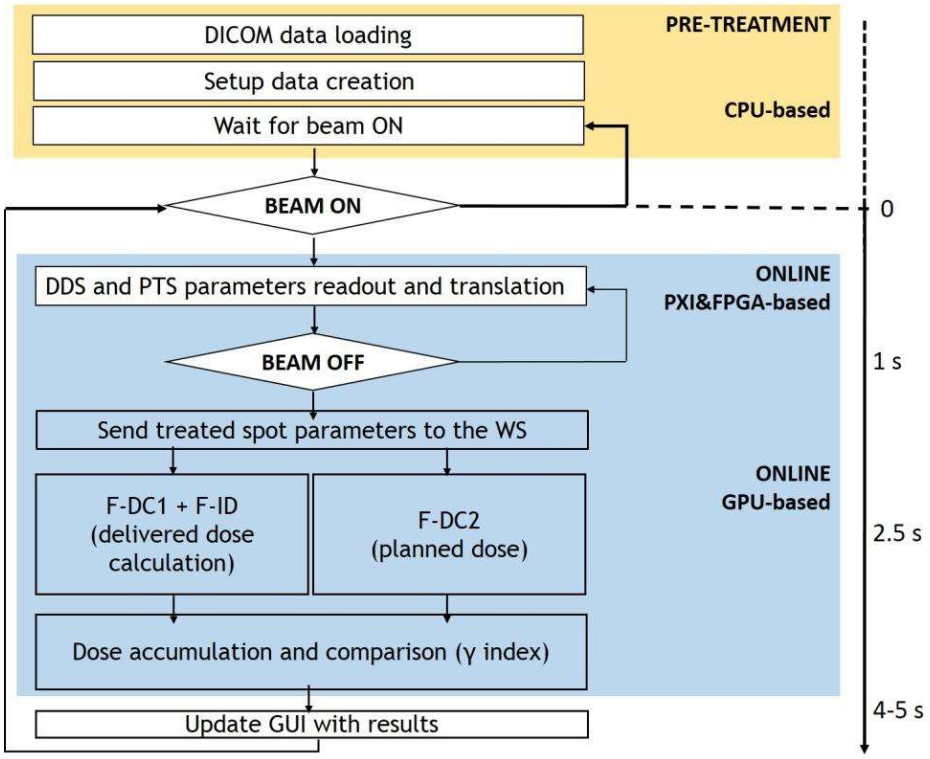

Fig 1. Flowchart of the RIDOS operations subdivided in two groups: pre-treatment tasks (low time constraints) and online tasks. Abbreviations: Patient Tracking System (PTS); Dose Delivery System (DDS); Fast dose calculation (F-DC); Fast Image Deformation (F-ID).

Table 1. List of the main RIDOS operations subdivided into three groups: *pre-treatment* (based on CPU), *online* based on GPU and *online* based on PXI-FPGA. Abbreviations used: Deformable Image Registration (DIR).

RIDOS operations	Pre-treatment CPU-based	Data loading	Patient CTs Treatment plan (list of planned spot parameters) LUTs for the dose computation
		Create setup data	Deformation Vector Fields from DIR between each CT-phase and the reference CT

			Build the planned rays (one for each spot)
	Online PXI-FPGA-based	Data loading	Measured beam fluences and positions from DDS; breathing data from the Anzai system
		Create and share online data	List with measured spot parameters, repeatedly sent to the workstation
	Online GPU-based	Dose Calculation	Delivered dose on the specific CT phases and planned dose on the reference CT Update the cumulative planned dose distributions
		Warp to reference CT	Warp the dose distribution of the specific CT phase according to the reference CT Update the cumulative delivered dose
		Dose Comparison	3D γ-index between the delivered and planned doses
	Online CPU-based	Result display	GUI update with the computed doses and comparisons (γ-index passing rate)

2.1 RIDOS hardware architecture

The RIDOS systems includes a National Instruments PXI chassis, which houses the controller NI PXI-8115, hereafter PXI-CPU, and a NI PXI-7831R module equipped with a Xilinx Virtex II (San Jose, California, US) Field-Programmable Gate Array (FPGA). The FPGA module accepts both digital and analog inputs, namely the measured beam parameters from the CNAO DDS and the respiratory waveform from the Anzai respiratory system, respectively. The WS used is a HP Z-820 equipped with 2 Xeon E5-2670V2 2.5 10C, 64GB RAM and a NVIDIA Tesla-K40c GPU. The main hardware and interfaces are also sketched in Fig. 2.

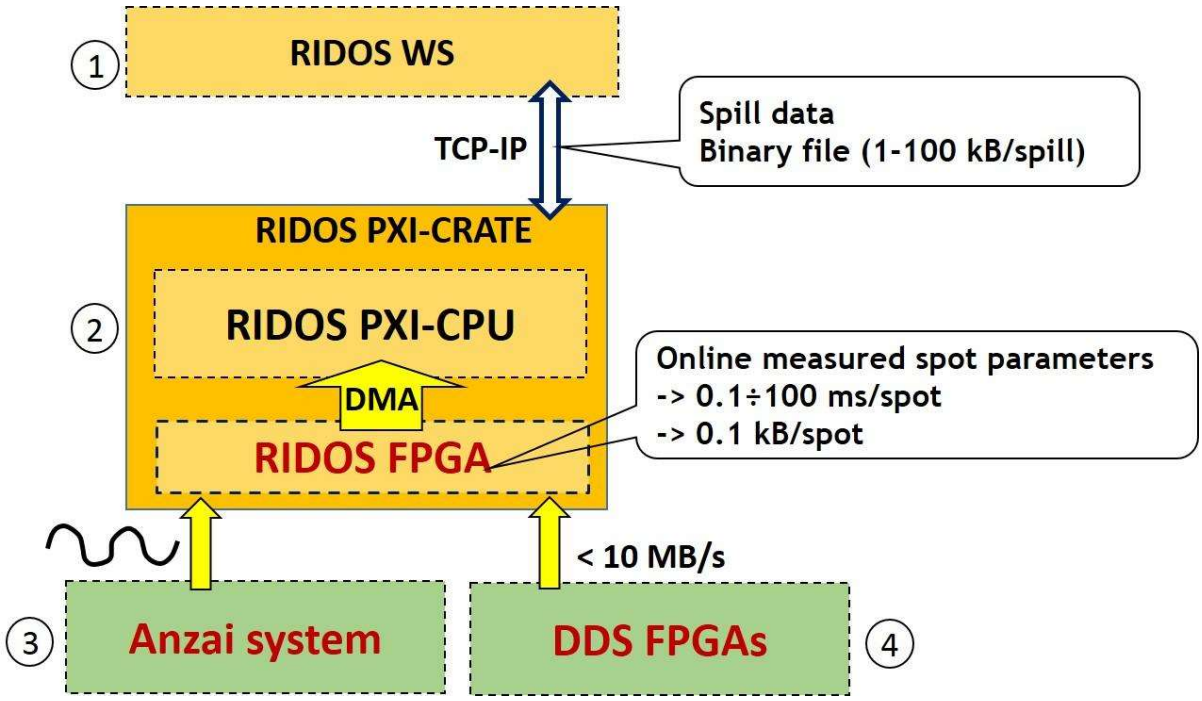

Fig. 2. RIDOS sub-systems and connections: 1) WS for F-DC computation, dose comparison and GUI execution; 2) National Instrument PXI-crate to share the spill data with the WS; 3) cabled connection with the Anzai system; 4) cabled connection with the CNAO-DDS FPGAs.

GPU features

RIDOS makes use of a NVIDIA Tesla-K40 GPU, equipped with 12 GB GDDR5 integrated memory, 288 GB/s memory bandwidth and 2880 CUDA core. The GPU code was developed using NVIDIA's Compute Unified Device Architecture (CUDA) version 6.5 of the software development kit. CUDA extends the C programming language by splitting intensive computations on a large number of small tasks (threads), running in parallel on different data [28,29]. Threads are organized in a hierarchy of blocks, each block corresponding to a maximum of 1024 threads with their own resource assignment.

2.2 Pre-treatment CPU-based operations

2.2.1 Data loading and initialization

Any software that performs forward dose computations makes use of clinical input data like the patient or phantom CT, the contours of CT structures and the characteristics of the beams; likewise, RIDOS needs the reference CT with its contours and the list of planned spot parameters. These data are saved in clinical DICOM files, which have been stored in the Oncological Information System database the day of treatment approval. The DICOM files are first translated into new data tailored to be used by the RIDOS algorithms. The RIDOS F-DC is based on the algorithms of the Dose Engine Kernel (DEK) treatment planning framework and beam model [39].

In a pre-treatment phase, RIDOS receives the main spot parameters as planned by the TPS, which are the number of particles, the transversal position at the isocenter and the energy. These data are used to allocate memories and initialize the system. In the future, also the 4DCT data and the Deformation Vector Field (DVF) maps pre-computed with the Deformable Image Registration (DIR) algorithms will be loaded in this phase.

2.2.2 Image registration

In order to evaluate online the cumulative delivered dose distribution, RIDOS includes an accurate and time-consuming pre-treatment DIR process on CPU between each breathing phase CT (moving images) and the CT used to plan the treatment (reference image). The DIR software tool creates a 3D array with transformations, i.e. a DVF, for each CT. Those DVFs are then exploited by a GPU-based image deformation algorithm (F-ID), with the aim of warping online the dose distributions calculated on each CT to the reference CT. The DIR algorithm implements a parametric non-rigid registration method, based on the ITK open-source framework [40]. It adopts a cubic-Bspline transformation, Mattes Mutual Information [41] as similarity measure, a linear interpolator and a L-BFGS-B optimization [42]. A detailed overview of image registration algorithms is given in [43].

2.3 Online PXI-FPGA-based operations

As shown in Fig. 2 the NI-PXI system collects and synchronizes data from the CNAO DDS and the Anzai device and sends the processed data to the RIDOS workstation at the end of each spill. In detail, the FPGA receives the measured spot parameters in real-time from the CNAO-DDS FPGAs [36]. The transfer speed ranges between a few kB/s and 10 MB/s because each spot needs 100 bytes and the CNAO spot duration changes between about 0.3 and 100 ms. In parallel, the FPGA receives the waveform measured by the Anzai pressure sensor. Thus, the delivered spot data are merged and synchronized with the respiratory phase in a new data sequence, continuously sent to the PXI-CPU through a Direct Memory Access (DMA) method. A customized LabVIEW program runs on the PXI-CPU to interface the FPGA with the WS where the fast dose computations and the RIDOS GUI are executed.

2.4 Online GPU-based operations

The DEK algorithms [39] have been ported on a GPU code optimized to provide a very fast forward dose computation, in order to update the cumulative dose distributions at the end of each spill. These algorithms compute both physical and RBE-weighted dose distributions for treatments with protons or carbon ions (see details in section 2.4.2). To perform online comparisons, the reference cumulative dose distribution is also computed, spill by spill for the irradiated spots using the planned parameters and the reference CT. When 4DCT data will be available, the dose distributions will be computed separately for each patient breathing phase, using the measured spot characteristics and respiratory phase. The partial dose distributions will be therefore mapped on the reference CT using the DVF from the pre-treatment DIR operations.

2.4.1 Initialization steps

The performance of algorithms running on GPU strongly depends on the proper use of the memory resources and data access [25], the major limitations coming from the data transfers between the host (CPU) and the device memories (GPU). All the GPU memories containing input and output data not depending on the irradiation time-structure are allocated and filled before the start of the treatment. In particular, the following data are loaded in GPU global memories:

- $N_{br\text{-}ph}$ matrices containing CT data for each breathing phase, where $N_{br\text{-}ph}$ is the number of respiratory phases in which the 4DCT is segmented. In particular, the values loaded in the matrix are the stopping powers relative to water, assigned to each voxel by a CT-scanner-specific calibration curve. Non regular CT grid spacing are allowed;
- a set of ($N_{br\text{-}ph}$ + 1) DVF maps, each one providing the displacements of the coordinates of each voxel of the reference computing grid to the corresponding coordinates in the computing grid of another breathing phase. These maps are built by interpolating the DVF maps provided by the DIR, usually calculated on a grid with a lower number of points than the CT;

- several sets of Look Up Tables (LUTs), one set for each possible beam energy, used to calculate the physical dose and the parameters needed for evaluation of the biological dose, as explained in the following;
- the fluence, trajectory and energy of the pencil beams prescribed by the TPS for the treatment.

In addition a set of ($N_{br\text{-}ph}$ + 2) 3D matrices (to store the calculated dose) are allocated. These matrices correspond to down-sampled computing grids compared to the CT grids and are filled with: a) the $N_{br\text{-}ph}$ cumulative partial dose distributions for each breathing phase, b) the total cumulative dose calculated on the reference CT coordinate system and c) the cumulative planned dose distribution.

2.4.2 Online GPU computations

At the end of each spill, a set of data is transferred to the GPU memories, containing for each irradiated spot the main measured properties (spot number of particles and positions) and the patient breathing phase.

The computation of the dose distributions is based on ray-tracing method through the patient CT and on the interpolation of LUTs. In order to provide results before the end of the inter-spill time, these tasks are accelerated with parallel execution of the kernel code running on multiple threads, each thread accessing different subsets of data (for example different voxels).

To accelerate memory transfers during the computation, the F-DC code makes extensive use of local and shared memories with coalescent memory accesses [25]. Moreover, the hardware built-in functions to fetch points and provide immediate interpolation in 3D texture memories are intensively used.

Ray-tracing

The lateral shape of the beam (both proton and carbon ion) is obtained as a superposition of elementary beamlets obtained from preliminary Monte Carlo simulations (implemented using the Fluka code[44]) carried out in water. The weights used in the superposition are obtained from the beam optics phase space derived experimentally for the specific beamline. For details of the method see also [45].

Each elementary beam is propagated through the patient CT along a straight line; the beam axis is defined as ray. The direction of each ray points to the corresponding spot and is identified by the X, Y and Z coordinates of the CT reference system.

The Siddon algorithm for ray-tracing [35] was used to compute the cumulative radiological path length along each ray, through the CT array of the corresponding breathing phase. The voxels through which the ray travels are determined from the intersections of the ray with all the orthogonal planes which define the CT grid. The computation is done in parallel for all the planes, with threads organized in CUDA blocks, one block for each ray. A stream compaction algorithm is used to remove intercepts not corresponding to voxels crossed by the ray. As a results of these computation, an unordered list of distances between the patient's entrance point of the ray in the CT and each intercept is provided. A CUDA-based bitonic

sort algorithm [36] has been implemented to sort the distances and evaluate of the total length travelled by the ray at the borders of each voxel.

For the dose computation, the path length of a beam in a heterogeneous material has to be converted into a Water Equivalent Path Length (WEPL). The cumulative WEPL values are determined by multiplying the length travelled inside each voxel by the stopping power ratio of the voxel loaded in the CT matrix with respect to the stopping power of water. For each ray, a mapping between the travelled lengths in water and in the CT medium is thus obtained and saved in texture memories to be used during the dose computation phase.

Dose computation

The algorithms used in this work feature a set of 3D beam-line specific LUT prepared during initialization, filled with several linearly-superposable quantities that are used to evaluate the mean physical and radiobiological effects in water as a function of the position along the ray and its transversal distances from the ray. The LUTs have been created for different beam energies using Monte Carlo simulations of the interactions in water of finite-size beamlets, including a realistic modeling of the beam optics extracted from experimental data collected at CNAO [39]. Other LUTs are created by processing the output of the Monte Carlo simulation with a radiobiological model, namely the Microdosimetric Kinetic Model (MKM [46]) or the Local Effect Model I (LEM-I) [47]. The output consists of different LUTs for each beam energy containing the spatial 3D distribution of the physical dose per unit fluence and other intensive parameters for the evaluation of radiobiological distributions (LET, RBE, RBE-weighted dose). For details on the DEK beam model and use of the linearly-superposable radiobiological quantities refer to [39]. As an example of LUT values, Fig. 3 shows the physical dose per unit fluence at different energies in the Y - $Z/Z_{BraggPeak}$ plane.

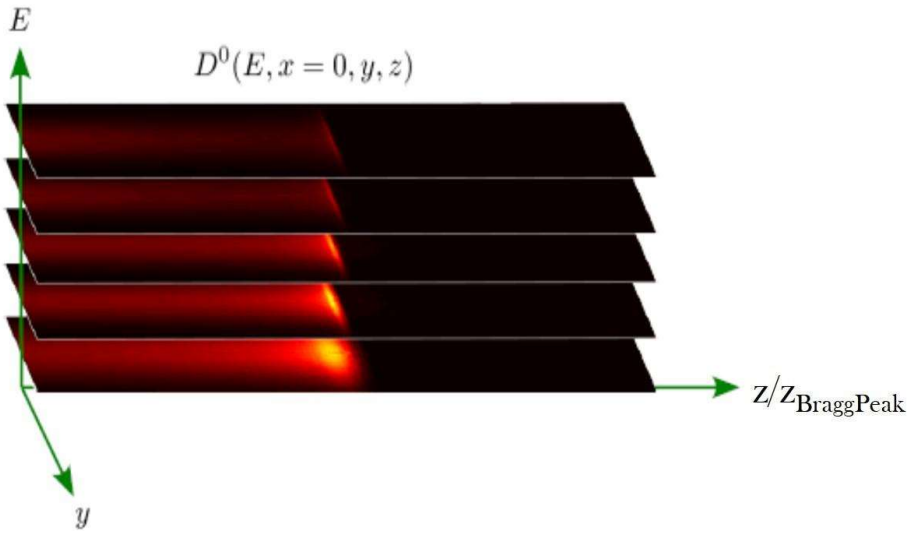

Fig. 3. 2D physical dose distributions per unit fluence at different energies stored in the DEK LUT.

In order to apply the LUTs created in water on an inhomogenous target, it is necessary to convert the coordinates of the considered medium in water equivalent coordinates. This is

done by computing, for the voxels of the computing grid around the ray direction, the longitudinal and transversal components of the voxel coordinates with respect to the ray axis, as shown in Fig. 4. In particular the longitudinal distance is defined as the distance of the projection of the voxel position on the beam axis from the entrance position of the ray in the body. It is worth noting that in general the beam might be asymmetric, and therefore two components of the transverse position are computed in the reference frame of the beam. The WEPLs of each voxel are obtained with a parallel interpolation of the WEPL maps extracted with the ray-tracing algorithms.

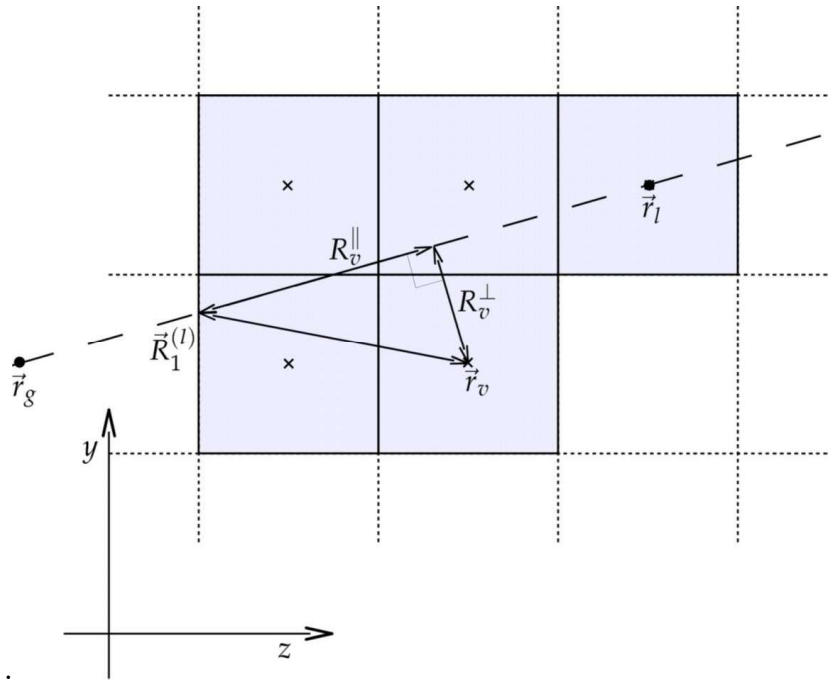

Fig. 4. Definition of the radial and transverse components of the voxel coordinates with respect to the ray trajectory.

Once the longitudinal WEPL and radial distances are known for the voxels of the computing grid, the intensive quantities needed for the physical and biological dose computation are determined from a texture interpolation of the proper LUTs for the given ray energy, weighted by the ray fluence. The output values are used to update the corresponding computing grids. We refer to Russo *et al.* [39] for the rules used to calculate radiobiological parameters and RBE-weighted dose distributions.

All these computations must be performed for a large number of voxels for each ray (10^4-10^6 voxels). Even if a parallel CUDA algorithm has been developed, the computing time of this part of code is dominant in all the GPU online calculations. In general only the voxels within a radial cut-off distance with respect to the ray direction are considered (the typical cut-off distance is 80 mm, assuming negligible dose contributions at larger distances from the beam axis). A further optimization in the computing time is obtained by using an adaptive algorithm for the preselections of a list of voxels contained inside a cylinder whose axis and radius are calculated on the base of the directions of all the previous rays, to be sure that all

the voxels within the radial cut-off of the current ray are included. Efficient GPU implementation of stream compaction algorithms are used for the voxel selections.

When all the spots of the previous spills have been processed, the computing grids containing the planned and measured cumulative dose distributions are updated and transmitted to the GUI program for fast comparison and visualization.

2.4.3 Application of Deformation Vector Fields.

In order to allow a comparison with the planned dose distribution, when different CT images will be available for different patient respiratory phases, synchronized with group of delivered spots, all the partial dose distributions will be mapped on the same reference system. Using the set of pre-computed DVF matrices, 3D texture interpolations are performed to extract the dose contributions in the displaced points of each breathing phase and assign them to the corresponding voxels of the reference computing grid. These interpolations are performed assigning a GPU thread for each voxel within the radial-cut off distance from the ray axis.

2.5 Online Dose Comparison

As soon as the GPU online computations are finished, at the end of each spill, the planned and the delivered cumulative dose distributions are used by dedicated software tool for analysis. The latter compares each pair of doses performing the 3D absolute difference and the 3D γ-index [48] computation.

Fast 3D γ-index

The γ-index algorithm was proposed by Low *et al*. [48] to quantify the differences in dose distributions by means of equation 1.

$$\gamma(\vec{r}_r) = min_{c=1\ldots M}\,\Gamma_r(\vec{r}_c, D_c) = \min_{c=1\ldots M}\sqrt{\frac{|\vec{r}_c-\vec{r}_r|^2}{\Delta d_M^2} + \frac{[D_c(\vec{r}_c)-D_r(\vec{r}_r)]^2}{\Delta D_M^2}} \qquad (1)$$

The γ-index results depend on the Distance To Agreement (DTA, Δd_M) and on the Dose-Difference (DD, ΔD_M) settings. Equation 1 defines the γ-index as the minimum distance between two corresponding points in the dose-distance space, where $D_r(\vec{r}_r)$ is the measured dose and $D_c(\vec{r}_c)$ is the computed dose. Voxels meet the acceptance criteria only where $\gamma(\vec{r}_r) \leq 1$. This evaluation method is very time consuming because all the points in both distribution are used during the evaluation.

For RIDOS a fast γ-index implementation, proposed by Wendling *et al.* [49] and implemented on GPU by Persoon *et al.* [50], was used. This method introduces a search sphere (or search box) for each measured dose point and centered on it. The γ-index values $\Gamma_r(\vec{r}_c, D_c)$ are computed only for points with $|\vec{r}_r - \vec{r}_c| \leqslant \Delta d_M$. The search is performed on the voxels of the calculated dose in a sphere centered on the reconstructed voxel, for increasing radii. An early stop is activated if $\sqrt{|\vec{r}_c - \vec{r}_r|^2/\Delta d_M^2}$ is bigger than the minimum $\Gamma_r(\vec{r}_c, D_c)$ found during evaluation in the search sphere. To increase the resolution, the search grid is

usually further segmented by linearly interpolating the dose between neighbouring voxels with a given interpolation distance, typically one tenth of the voxel distance.

The fast 3D γ-index algorithm has been developed for RIDOS using Matlab software environment and executed within the RIDOS GUI.

2.6 RIDOS Graphic User Interface

A Matlab-based GUI was developed to display, in the RIDOS monitor placed in the CNAO local control room, a summary of the computed doses and comparisons. Controls allow the user to set the computing grid dimensions, the dose-difference and distance-to-agreement criteria for the γ-index computation. Spill by spill the γ-index passing rate updates a histogram to provide a trend of the dose delivered accuracy.

3 Results

A sample of representative results are presented in the following to show the RIDOS capabilities. We start with the GPU-based algorithms developed and integrated in the RIDOS workflow that have been benchmarked against existing validated codes or new CPU-based ones developed for verification purpose. As an example, dedicated CPU-based ray-tracing was developed to verify the accuracy of the GPU code and to compare the time performance as a function of computing grid dimensions, i.e. number of voxels and threads. The dose accuracy and time performance of the whole F-DC computation were compared mainly with the DEK computations, which were implemented for execution on single core CPU and without time constraints.

3.1 GPU-based ray-tracing performance

As expected, using the CPU sequential computing, the ray-tracing execution time scales linearly with the number of intercepts and rays; the parallel computing features of the GPU are found to be very time effective above 500 rays, as shown in Tab. 2.

As an example, the ray-tracing algorithms for 5000 rays on a standard 512x512x127 CT grid, which leads to 512+512+127 intercepts, requires 400 ms when running on the CPU and 27 ms on the GPU. A summary of the CPU and GPU ray-tracing time performances is shown in Tab. 2 where the times for the intercepts search and sorting, for the segment length computation and for the overall ray-tracing computation are shown as a function of the number of rays.

Tab. 2 Example of CPU and GPU ray-tracing time performances: times required for the intercepts search, sorting, segment length computation and overall ray-tracing computation as a function of number of rays crossing a 512x512x127 CT grid.

	Time (ms)								
	Search		Sorting		Path length		**Total ray-tracing**		**Gain**
N. rays	CPU	GPU	CPU	GPU	CPU	GPU	**CPU**	**GPU**	**CPU/GPU**
1	0.0	2.5	0.1	0.1	0.2	3.1	**0.3**	**3.3**	**0.1**
10	0.2	2.8	0.5	0.1	1.3	4.0	**1.8**	**4.2**	**0.4**
50	1.2	2.9	1.7	0.2	5.2	4.8	**6.9**	**5.0**	**1.4**

100	1.5	3.1	3.5	0.4	10.3	5.1	**13.8**	**5.5**	**2.5**
500	7.4	3.1	17.5	1.4	48.2	6.5	**65.7**	**7.8**	**8.4**
1000	25.6	3.3	38.1	2.6	101.8	8.6	**139.9**	**11.2**	**12.5**
5000	72.8	3.4	170.9	12.6	443.7	12.6	**614.6**	**25.2**	**24.4**
10000	145.5	5.0	341.4	25.2	894.8	23.3	**1236.2**	**48.5**	**25.5**
50000	725.7	18.7	1650.4	127.6	3651.7	108.0	**5302.1**	**235.6**	**22.5**
100000	1487.3	40.7	2859.8	256.2	7396.7	218.5	**10256.5**	**474.7**	**21.6**

As we can see from Tab. 2, the ray-tracing computation for more than 50 rays is faster using GPU. The results also show that, for a typical spill comprising up to 10000 rays, ray-tracing can be accomplished within 50 ms. The times measured for the single algorithms include also the time spent to perform the time measurements; thus the time for the total ray-tracing is lower than the sum of the times of the single steps.

3.2 Fast dose calculation accuracy

The DEK physical and biological doses calculated in the patient CT by the DEK original code running on the CPU [39] were used as benchmark for the RIDOS F-DC.

Figure 5 shows the dose of a 2D central slice for a clinical skull-base tumor treatment calculated both using DEK DC (CPU) and RIDOS F-DC (GPU). The absolute difference, shown on the right, is negligible ($<10^{-4}$ Gy) and mainly due to the limited numeric precision of the interpolations performed using GPU textures. This treatment was selected among the ones used for DEK validation in patient CT, thus in heterogeneous tissues.

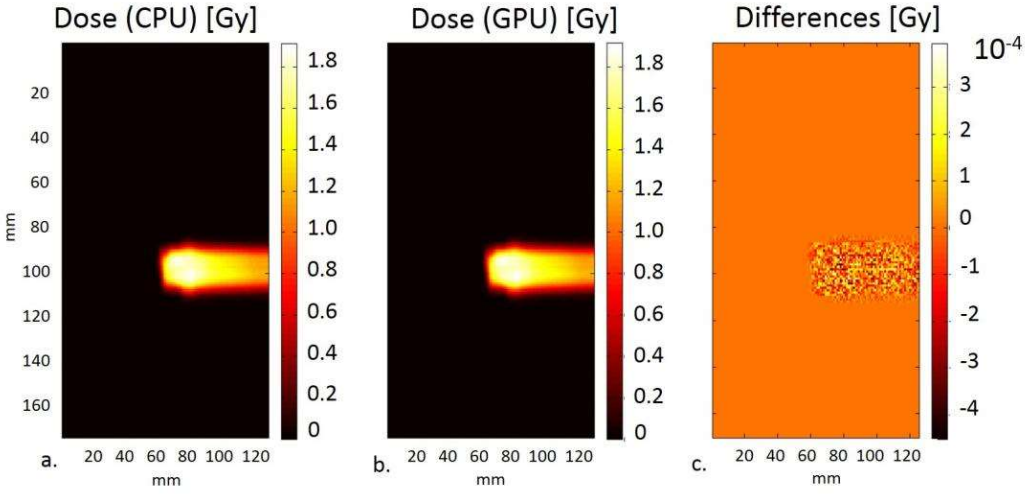

Fig. 5 2D samples of 3D proton physical dose distributions computed a) with DEK, i.e. CPU based FP, b) with RIDOS GPU-based DC and c) the absolute difference.

Similar results have been obtained for carbon ion physical and biological doses; Fig. 6 shows the results for carbon ion biological doses.

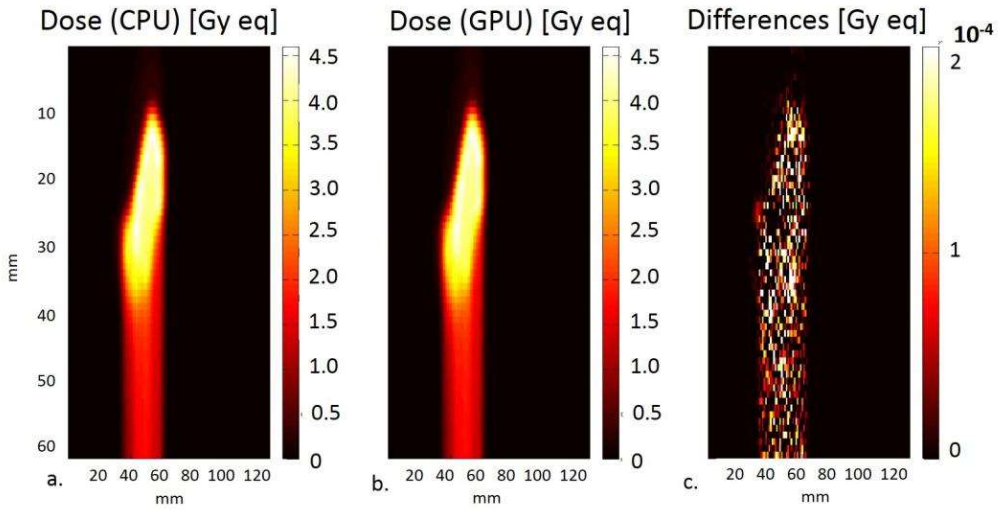

Fig. 6 2D samples of 3D carbon ion biological dose distributions computed a) with DEK, i.e. CPU based DC, b) with RIDOS GPU-based DC and c) absolute difference.

Additionally, dose comparisons with the clinical Syngo RT Planning (V13) (Siemens, Germany) [51] were also performed for irradiations with both protons and carbon ions of a clinical head and neck tumor in the patient CT and for a uniform dose distribution in a water cube. The lateral profiles are compared in Fig. 7 and Fig. 8 respectively.

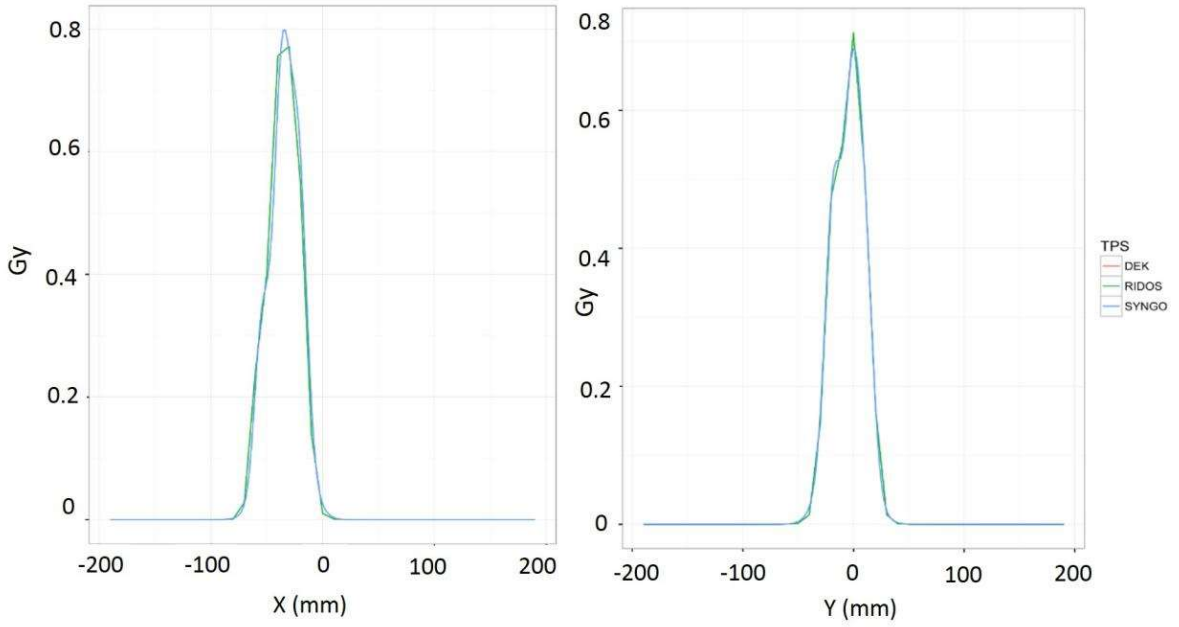

Fig. 7 Lateral profiles along X and Y for single field dose distributions of a clinical proton treatment computed with DEK, RIDOS and Syngo in the patient CT. DEK and RIDOS distributions are completely overlapped.

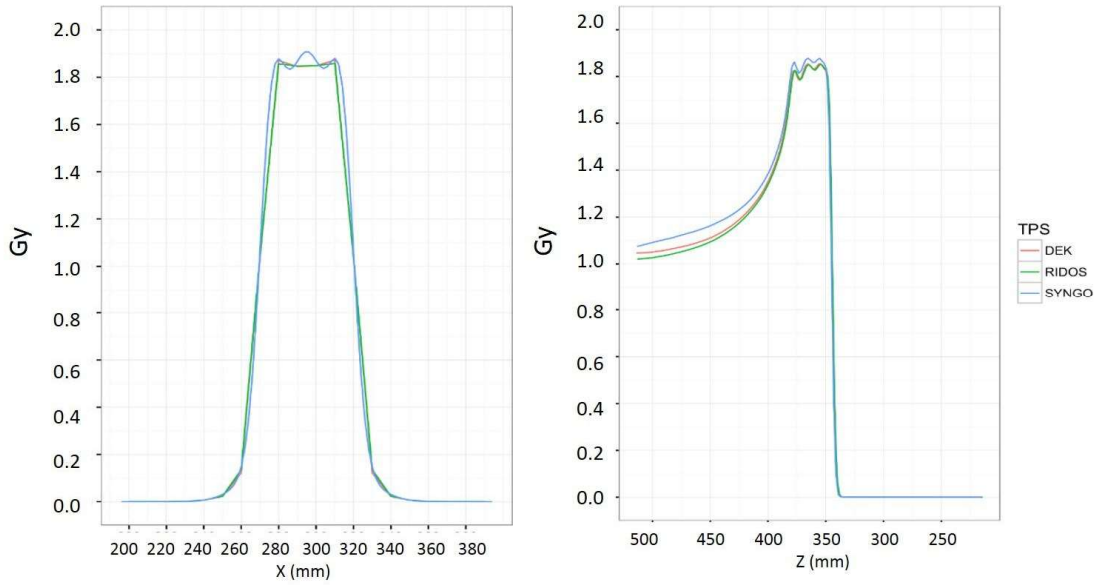

Fig. 8 Lateral and longitudinal profiles (along X and Z) for a uniform dose in a water cube, computed with DEK, RIDOS and Syngo.

3.3 Fast dose calculation time performance

The time gain achieved with the RIDOS F-DC with respect to the DEK DC is presented for the same clinical brain case, used to benchmark the dose computation. A single field is considered, characterized by 39 energy layers and 1248 spots. The dimension of the CT was a standard matrix of 512x512x125 voxels that cover 320x320x250 mm^3, while the selected computing grid was of 170x170x125 voxels. Within the computing grid, the number of voxels considered for the dose computation depends on the radial cut-off used. A larger cut-off leads to a larger number of voxels and to an increase of the computation time for the CPU-based code that performs the dose computation voxel by voxel. As an example, moving the cut-off from 10 to 80 mm leads to an increase of the voxels for the dose computation from 4×10^4 to 7×10^5.

Table 3 shows the gain in computing speed of the RIDOS F-DC compared to the DEK DC to compute the dose distribution of the total field and of a single energy, which represents the delivery of a single spill. Results show that the whole dose computation, using 80 mm as cut-off, is more than 600 times faster using GPU algorithms. It is worth underlying that DEK algorithms were optimised not to be fast, but versatile and robust. Concerning the times required for single spills, the RIDOS F-DC spends always less than 0.5 seconds for the dose computation and therefore its performance is well within requirements.

Table 3. F-DC computing times per full treatment and per spill as a function of the radial cut-off.

		F-DC GPU-based times (s)		CPU-based times (s)	Gain
Cut-off (mm)	N. of voxels	Total field	Single energy	Total field	Total field
10	11200	1.9	0.02	40	21
20	44678	1.9	0.03	157	83
40	178702	2.2	0.06	636	293
50	279243	2.6	0.07	997	383

80	714861	3.9	0.12	2534	655

Similar results are obtained considering a carbon-ion treatment.

3.4 Fast image deformation and dose accumulation performance

The F-ID accuracy depends on the quality of the DVFs provided off-line with existing DIR tools, which must be carefully tuned to achieve the best transformations. The issue of DIR accuracy and robustness is outside RIDOS purposes that aim at shortening as much as possible the time required to warp the dose distributions delivered on different CTs exploiting the DVFs. The GPU-based dose warping developed for RIDOS takes advantage of the use of trilinear texture interpolation and provides a remapped dose in 80 ms while the reference CPU algorithm needs about 5 seconds.

The preliminary results show the feasibility of computing the dose accumulation online even for spills delivered during different respiratory phases because in less than 100 ms RIDOS F-ID warps the current delivered dose with to the reference CT and evaluate the online cumulative dose.

3.5 Dose comparison performance and GUI update

To evaluate the fast γ-index (F-GI) accuracy, the software was tested with three different pairs of dose distributions. In all cases, the γ-index meeting the acceptance criteria were in perfect agreement with the computation performed using the classical method and the Persoon's code.

Four different dose distributions with 100^3 to 200^3 voxels of 1 mm^3, and DD=3% and DTA=3mm were used for the performance test, which results in the execution times summarized in Tab. 4.

Table 4. γ-index evaluation in RIDOS using DD = 3%, DTA = 3 mm, texture activated and no subdivision.

	Volume 1	**Volume 2**	**Volume 3**
Time to load data on GPU [ms]	218 to 226	304	201
Time for γ-index calculation [ms]	12 to 15	201	20
Time to free the GPU and CPU memories [ms]	1	2	1
Total time [ms]	**226 to 241**	**507**	**222**

Volume 1: 100^3 voxels of $1mm^3$. Volume 2: 200^3 voxels of 1 mm^3. Volume 3 with 170x170x125 voxels of 1.9x1.9x2.0 mm^3.

In RIDOS, the F-GI computation uses the planned and delivered dose distributions computed by the F-DC and the times for loading data are of the order of few ms.

The γ-index computation time is inversely proportional to the dose interpolation resolution. Using an interpolation of 0.1 mm as in Table 4 the γ-index values were obtained in approximately half a second.

A screenshot of the RIDOS GUI is shown in Fig. 9, where in the upper left corner the projections in a plane orthogonal to the beam direction of the planned and delivered doses during the last spill are displayed as 2D plots. The absolute difference between the planned and the delivered doses, and the integrated delivered dose are shown in the plots below. The γ-index results fill the upper right corner with two 2D plots that present the projections of the whole γ-index and of the dose difference contribution alone. Additionally, spill by spill a histogram is updated to show the γ-index passing rate trend.

The lower right corner is filled with strings and numerical indicators to display mainly the spill number progress, the path and name of the saved files. Moreover, three input parameters allow to choose the computing grid resolution, the DD and the DTA to be used for the γ-index evaluation.

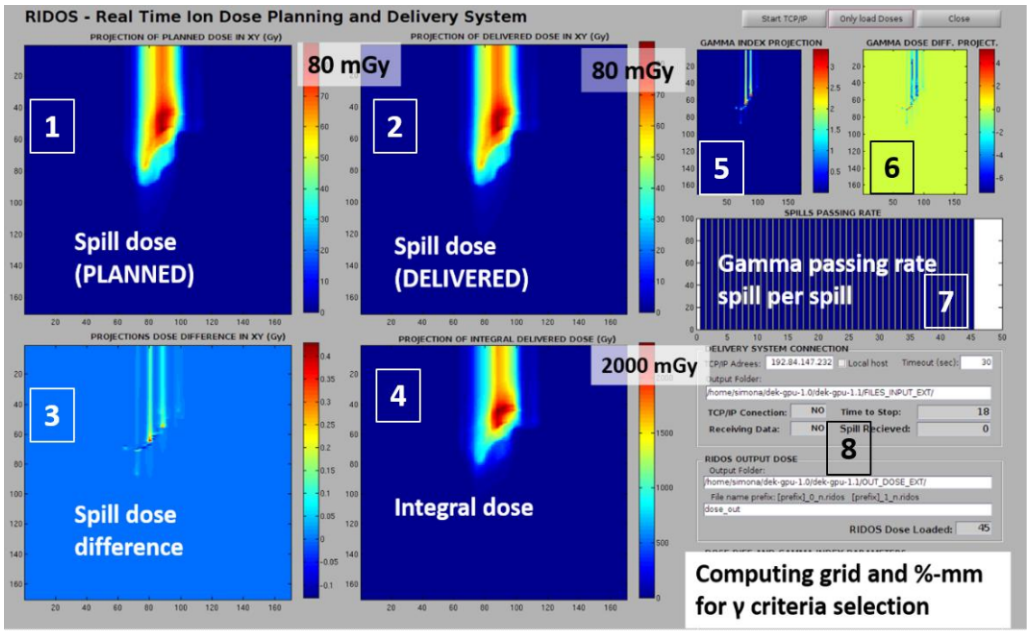

Fig. 9. RIDOS GUI screenshot: 1 and 2 are the planned and delivered dose projections in the XY plane; 3 and 4 show the absolute difference between the two doses and the integrated delivered dose respectively. 5 and 6 present the projections of the whole γ-index and of the dose difference contribution alone; 7 is a histogram to show the γ-index passing rate trend. 8 includes strings and numerical indicators to display the spill number progress, the path and name of the saved files and three input parameters to change the γ-index computing grid resolution, the DD and the DTA.

3.6 Online dose validation with PinPoint measurements

We performed a test to validate the RIDOS online capability in computing the dose by comparing with a reference. A 3x3x3 cm^3 volume of water was irradiated with protons delivering 1.8 Gy uniform dose. Spill by spill, the RIDOS cumulative doses were compared with the integral absolute dose measured by a calibrated PinPoint ionization chamber model T31015 (PTW, Germany) [36] placed in the middle of the 3D dose distribution. We remark that at CNAO the dose is delivered starting from the proximal slices (i.e. lower energies),

therefore most of the dose is delivered in the last spills, which have to reach the distal layers. The first six spills are at zero doses because the irradiated layers are at a shallow depth and particles are not reaching the point of measurement, which is in the middle of the cube.

Figure 10 shows the cumulative doses computed with RIDOS and measured with the PinPoint, whereas the absolute dose difference as a function of the delivered dose was evaluated and presented in Fig. 11.

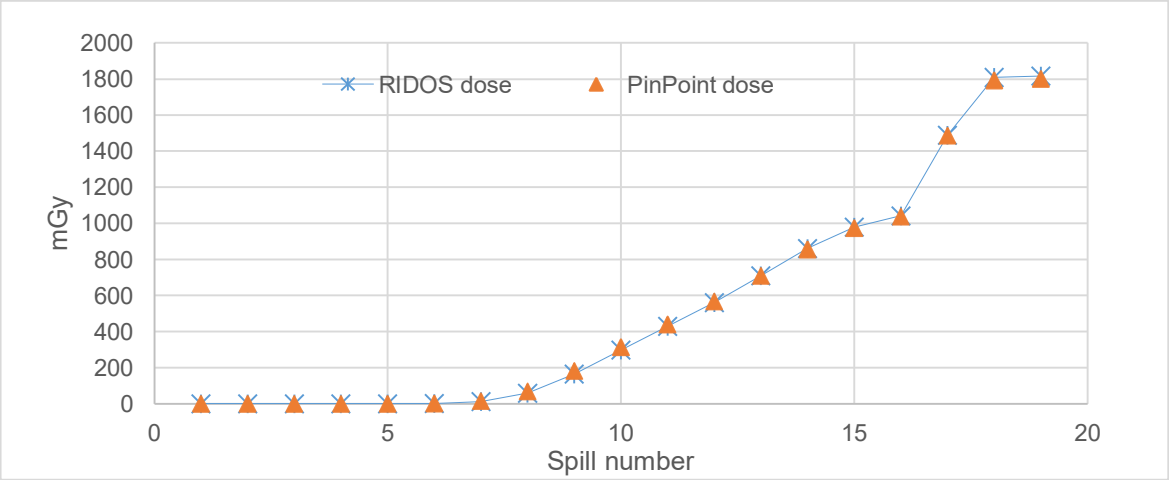

Fig. 10 The RIDOS and the PinPoint values of dose respectively computed and measured for each spill.

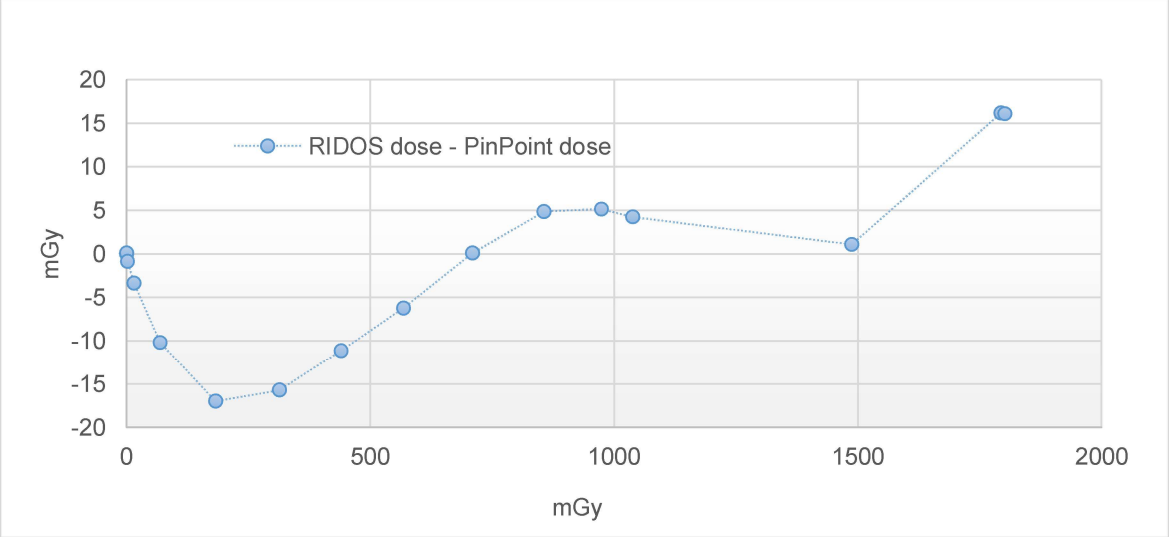

Fig. 11 Absolute dose difference between the RIDOS dose computation and the PinPoint measurement as a function of the integrated delivered dose.

It is found that the measurements of the PinPoint is well described by the computation made with RIDOS leading to a deviation less than ± 20 mGy in all the spills. The relative dose deviation is less than 1% for integral doses larger than 600 mGy.

4 Discussion

This paper demonstrates the feasibility of online dose delivered verification by means of dose calculation for scanned ion beam therapy and provides the description of all the operations and components required for the clinical implementation.
By porting the pencil beam algorithms to GPUs we reduced the time for dose computation from 20 to 600 times and achieved the goal of developing a sub-second dose computation. Such a high gain was feasible because the reference CPU-based code (DEK) was not optimized for speed but to be versatile in order to be easily adaptable to different beams, facilities and computational requirements.
Among the interesting results, we found that the GPU-based ray-tracing is faster than the same CPU computation for a large number of rays (more than 4000) which cross a standard 512x512x127 CT grid. Additionally, as reported in an earlier publication [25], an advantage of GPUs consists in a high performance for interpolating 1D, 2D and 3D tables by means of their characteristic texture memories. The RIDOS F-DC exploits the latter because, as for the reference DEK DC, it is based on pre-computed LUT and the dose values for thousands of voxels are obtained from trilinear LUT interpolations. We notice that a significant advantage of the DEK beam model [39] is the description of each pencil beam as composed of sub-units (hundreds of rays). The larger the number of sub-beams, the better the dose accuracy reconstruction. This feature is particularly useful to compute accurate dose in the patient CT with highly heterogeneous figures and favourable for GPU implementation where sub-beams can be processed in parallel.

Similarly to the study published by Richter at al. [52], RIDOS has been designed to use online patient motion monitoring data to correlate the breathing phase CTs with the temporal structure of the beam delivery in order to reconstruct the final delivered dose distribution, which incorporates both interplay effects and range changes.

To prepare RIDOS for a realistic evaluation of motion effects, we are developing tools to synchronize the dynamic patient deformations and the timeline of the delivery. These include the use of 4DCT images and online breathing measurements from the clinical Anzai respiratory tracking system. The measurements from the Anzai system were integrated in RIDOS and will be exploited in future integrating in the system the RIDOS fast dose accumulation procedure based on 4DCT imaging and standardized DIR algorithms, introduced in this paper.

Currently RIDOS computes in the inter spill time two dose distributions on the same reference CT: one based on measured beam parameters and the other one based on planned data. A fast γ-index implementation was also developed and tested, that compares two dose distributions in a few hundreds of milliseconds and shows results in the RIDOS GUI spill by spill.

Moving targets at CNAO are treated with gating [7] so the 4D RIDOS capabilities could not be tested online with real patients; the Anzai QA phantom was used in place of the patient to verify the online update of spot data with the current respiratory phase sent spill by spill to the fast dose calculation. The use of different CTs for spill-dose calculation has been also implemented in the F-DC exploiting GPU and the additional time required to provide the spill

dose comparison was mainly dependent on the time spent for dose accumulation procedure, which takes around 100 ms to warp and sum two different doses.

A key point of this work is the proof that intra-fraction dose feedback is feasible if one has access to the dose delivery hardware and firmware in order to integrate them with a sub-second dose computation. Although RIDOS was tailored for the CNAO synchrotron delivery time structure, we are confident that the tools developed can be easily adapted to a cyclotron-based facility.

5 Conclusions

This work describes a new system, RIDOS, which exploits the use of GPUs to build an extension of the dose delivery system for scanned ion beams able to compare the delivered and planned dose distributions spill by spill for a synchrotron-based facility.
Up to now, different algorithms and dose engines have been developed for GPU-based hardware but, to our knowledge, none of these has been integrated within a clinical dose delivery system to be used for the online dose computation.

ACKNOWLEDGMENTS

The authors would like to thank Joakim da Silva for the useful suggestions about GPU and CUDA programming and for the fruitful discussions on the RIDOS applications.

This work was funded in part by the INFN Grant for Young Researcher (Call INFN n. 15766).

REFERENCES

[1] Bert C, Durante M. Motion in radiotherapy: particle therapy. Phys Med Biol 2011;56:R113–44. doi:10.1088/0031-9155/56/16/R01.

[2] Lambert J, Suchowerska N, McKenzie DR, Jackson M. Intrafractional motion during proton beam scanning. Phys Med Biol 2005;50:4853–62. doi:10.1088/0031-9155/50/20/008.

[3] Seco J, Robertson D, Trofimov A, Paganetti H. Breathing interplay effects during proton beam scanning: simulation and statistical analysis. Phys Med Biol 2009;54:N283–94. doi:10.1088/0031-9155/54/14/N01.

[4] Graeff C, Durante M, Bert C. Motion mitigation in intensity modulated particle therapy by internal target volumes covering range changes. Med Phys 2012;39:6004–13. doi:10.1118/1.4749964.

[5] Knopf A-C, Boye D, Lomax A, Mori S. Adequate margin definition for scanned particle therapy in the incidence of intrafractional motion. Phys Med Biol 2013;58:6079–94. doi:10.1088/0031-9155/58/17/6079.

[6] Dueck J, Knopf A-C, Lomax A, Albertini F, Persson GF, Josipovic M, et al. Robustness of the Voluntary Breath-Hold Approach for the Treatment of Peripheral Lung Tumors Using Hypofractionated Pencil Beam Scanning Proton Therapy. Int J

Radiat Oncol 2016;95:534–41. doi:10.1016/J.IJROBP.2015.11.015.

[7] Ciocca M, Mirandola A, Molinelli S, Russo S, Mastella E, Vai A, et al. Commissioning of the 4-D treatment delivery system for organ motion management in synchrotron-based scanning ion beams. Phys Medica 2016;32:1667–71. doi:10.1016/J.EJMP.2016.11.107.

[8] Minohara S, Kanai T, Endo M, Noda K, Kanazawa M. Respiratory gated irradiation system for heavy-ion radiotherapy. Int J Radiat Oncol Biol Phys 2000;47:1097–103. doi:10.1016/S0360-3016(00)00524-1.

[9] Matsuura T, Miyamoto N, Shimizu S, Fujii Y, Umezawa M, Takao S, et al. Integration of a real-time tumor monitoring system into gated proton spot-scanning beam therapy: An initial phantom study using patient tumor trajectory data. Med Phys 2013;40:71729. doi:10.1118/1.4810966.

[10] Shimizu S, Matsuura T, Umezawa M, Hiramoto K, Miyamoto N, Umegaki K, et al. Preliminary analysis for integration of spot-scanning proton beam therapy and real-time imaging and gating. Phys Medica 2014;30:555–8. doi:10.1016/J.EJMP.2014.04.002.

[11] Grözinger SO, Bert C, Haberer T, Kraft G, Rietzel E. Motion compensation with a scanned ion beam: a technical feasibility study. Radiat Oncol 2008;3:34. doi:10.1186/1748-717X-3-34.

[12] Rietzel E, Bert C. Respiratory motion management in particle therapy. Med Phys 2010;37:449–60. doi:10.1118/1.3250856.

[13] Knopf A, Bert C, Heath E, Nill S, Kraus K, Richter D, et al. Special report: workshop on 4D-treatment planning in actively scanned particle therapy–Recommendations, technical challenges, and future research directions. Med Phys 2010;37:4608–14. doi:10.1118/1.3475944.

[14] Zenklusen SM, Pedroni E, Meer D. A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new Gantry 2 at PSI. Phys Med Biol 2010;55:5103–21. doi:10.1088/0031-9155/55/17/014.

[15] Knopf AC, Hong TS, Lomax A. Scanned proton radiotherapy for mobile targets - The effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. Phys Med Biol 2011;56:7257–71. doi:10.1088/0031-9155/56/22/016.

[16] Graeff C. Motion mitigation in scanned ion beam therapy through 4D-optimization. Phys Medica 2014;30:570–7. doi:10.1016/j.ejmp.2014.03.011.

[17] Schätti A, Zakova M, Meer D, Lomax AJ. Experimental verification of motion mitigation of discrete proton spot scanning by re-scanning. Phys Med Biol 2013;58:8555–72. doi:10.1088/0031-9155/58/23/8555.

[18] Knopf AC, St??tzer K, Richter C, Rucinski A, da Silva J, Phillips J, et al. Required transition from research to clinical application: Report on the 4D treatment planning workshops 2014 and 2015. Phys Medica 2016;32:874–82. doi:10.1016/j.ejmp.2016.05.064.

[19] Graeff C. Motion mitigation in scanned ion beam therapy through 4D-optimization. Phys Medica 2014;30:570–7. doi:10.1016/J.EJMP.2014.03.011.

[20] Bert C, Graeff C, Riboldi M, Nill S, Baroni G, Knopf A-C. Advances in 4D treatment

planning for scanned particle beam therapy - report of dedicated workshops. Technol Cancer Res Treat 2014;13:485–95. doi:10.7785/tcrtexpress.2013.600274.

[21] Graeff C, Lüchtenborg R, Eley JG, Durante M, Bert C. A 4D-optimization concept for scanned ion beam therapy. Radiother Oncol 2013;109:419–24. doi:10.1016/J.RADONC.2013.09.018.

[22] Landry G, Nijhuis R, Dedes G, Handrack J, Thieke C, Janssens G, et al. Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation. Med Phys 2015;42:1354–66. doi:10.1118/1.4908223.

[23] Kurz C, Kamp F, Park Y-K, Zöllner C, Rit S, Hansen D, et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. Med Phys 2016;43:5635–46. doi:10.1118/1.4962933.

[24] Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online Adaptive Radiation Therapy. Int J Radiat Oncol 2017;99:994–1003. doi:10.1016/j.ijrobp.2017.04.023.

[25] Després P, Jia X. A review of GPU-based medical image reconstruction. Phys Medica Eur J Med Phys 2017;42:76–92. doi:10.1016/J.EJMP.2017.07.024.

[26] Giordanengo S, Manganaro L, Vignati A. Review of technologies and procedures of clinical dosimetry for scanned ion beam radiotherapy. Phys Medica 2017;43:79–99. doi:https://doi.org/10.1016/j.ejmp.2017.10.013.

[27] Trnková P, Knäusl B, Actis O, Bert C, Biegun AK, Boehlen TT, et al. Clinical implementations of 4D pencil beam scanned particle therapy: Report on the 4D treatment planning workshop 2016 and 2017. Phys Medica 2018;54:121–30. doi:10.1016/J.EJMP.2018.10.002.

[28] Giordanengo S, Manganaro L, Vignati A. Review of technologies and procedures of clinical dosimetry for scanned ion beam radiotherapy. Phys Medica 2017;43. doi:10.1016/j.ejmp.2017.10.013.

[29] Trnková P, Bolsi A, Albertini F. Factors influencing the performance of patient specific quality assurance for pencil beam scanning IMPT fields. Med Phys 2016;43:5998–6008. doi:10.1118/1.4964449.

[30] da Silva J, Ansorge R, Jena R. Sub-second pencil beam dose calculation on GPU for adaptive proton therapy. Phys Med Biol 2015;60:4777–95. doi:10.1088/0031-9155/60/12/4777.

[31] Li H, Sahoo N, Poenisch F, Suzuki K, Li Y, Li X, et al. Use of treatment log files in spot scanning proton therapy as part of patient-specific quality assurance. Med Phys 2013;40:21703. doi:10.1118/1.4773312.

[32] Schiavi A, Senzacqua M, Pioli S, Mairani A, Magro G, Molinelli S, et al. Fred: a GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy. Phys Med Biol 2017;62:7482–504. doi:10.1088/1361-6560/aa8134.

[33] Wan Chan Tseung H, Ma J, Beltran C. A fast GPU-based Monte Carlo simulation of proton transport with detailed modeling of nonelastic interactions. Med Phys 2015;42:2967–78. doi:10.1118/1.4921046.

[34] Qin N, Botas P, Giantsoudi D, Schuemann J, Tian Z, Jiang SB, et al. Recent developments and comprehensive evaluations of a GPU-based Monte Carlo package for proton therapy. Phys Med Biol 2016;61:7347–62. doi:10.1088/0031-9155/61/20/7347.

[35] Rossi S. The National Centre for Oncological Hadrontherapy (CNAO): Status and perspectives. Phys Medica 2015;31:333–51. doi:10.1016/j.ejmp.2015.03.001.

[36] Giordanengo S, Garella MA, Marchetto F, Bourhaleb F, Ciocca M, Mirandola A, et al. The CNAO dose delivery system for modulated scanning ion beam radiotherapy. Med Phys 2015;42:263–75. doi:10.1118/1.4903276.

[37] Desplanques M, Tagaste B, Fontana G, Pella A, Riboldi M, Fattori G, et al. A comparative study between the imaging system and the optical tracking system in proton therapy at CNAO. J Radiat Res 2013;54:i129–35. doi:10.1093/jrr/rrt043.

[38] Giordanengo S, Donetti M, Garella MA, Marchetto F, Alampi G, Ansarinejad A, et al. Design and characterization of the beam monitor detectors of the Italian National Center of Oncological Hadron-therapy (CNAO). Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip 2013;698:202–7. doi:10.1016/j.nima.2012.10.004.

[39] Russo G, Attili A, Battistoni G, Bertrand D, Cappucci F, Ciocca M, et al. A novel algorithm for the calculation of the physical and biological irradiation effect in scanned ion beam therapy : the beamlet superposition approach. Phys Med Biol 2016;183:1–31. doi:10.1088/0031-9155/61/1/183.

[40] H.J. Johnson , M.M. McCormick LI. The ITK Software Guide Book 2: Design and Functionality. Fourth Edi. Inc., Kitware; 2015.

[41] Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging 2003;22:120–8. doi:10.1109/TMI.2003.809072.

[42] Byrd RH, Lu P, Nocedal J, Zhu C. A Limited Memory Algorithm for Bound Constrained Optimization. SIAM J Sci Comput 1995;16:1190–208. doi:10.1137/0916069.

[43] Oh S, Kim S. Deformable image registration in radiation therapy. Radiat Oncol J 2017;35:101–11. doi:10.3857/roj.2017.00325.

[44] Battistoni G, Bauer J, Boehlen TT, Cerutti F, Chin MPW, Dos Santos Augusto R, et al. The FLUKA Code: An Accurate Simulation Tool for Particle Therapy. Front Oncol 2016;6. doi:10.3389/fonc.2016.00116.

[45] Russo G, Attili A, Battistoni G, Bertrand D, Bourhaleb F, Cappucci F, et al. A novel algorithm for the calculation of physical and biological irradiation quantities in scanned ion beam therapy: the beamlet superposition approach. Phys Med Biol 2016;61:183–214. doi:10.1088/0031-9155/61/1/183.

[46] Kase Y, Kanai T, Matsufuji N, Furusawa Y, Elsässer T, Scholz M. Biophysical calculation of cell survival probabilities using amorphous track structure models for heavy-ion irradiation. Phys Med Biol 2008;53:37–59. doi:10.1088/0031-9155/53/1/003.

[47] Scholz M, Kraft G. Track structure and the calculation of biological effects of heavy charged particles. Adv Space Res 1996;18:5–14.

[48] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25:656–61. doi:10.1118/1.598248.

[49] Wendling M, Zijp LJ, McDermott LN, Smit EJ, Sonke J-J, Mijnheer BJ, et al. A fast algorithm for gamma evaluation in 3D. Med Phys 2007;34:1647–54.

doi:10.1118/1.2721657.

[50] Persoon LCGG, Podesta M, van Elmpt WJC, Nijsten SMJJG, Verhaegen F. A fast three-dimensional gamma evaluation using a GPU utilizing texture memory for on-the-fly interpolations. Med Phys 2011;38:4032–5. doi:10.1118/1.3595114.

[51] Molinelli S, Mairani A, Mirandola A, Vilches Freixas G, Tessonnier T, Giordanengo S, et al. Dosimetric accuracy assessment of a treatment plan verification system for scanned proton beam radiotherapy: one-year experimental results and Monte Carlo analysis of the involved uncertainties. Phys Med Biol 2013;58:3837–47. doi:10.1088/0031-9155/58/11/3837.

[52] Richter D, Schwarzkopf A, Trautmann J, Krämer M, Durante M, Jäkel O, et al. Upgrade and benchmarking of a 4D treatment planning system for scanned ion beam therapy. Med Phys 2013;40:51722. doi:10.1118/1.4800802.