IPEM
Institute of Physics and
Engineering in Medicine

PAPER • OPEN ACCESS

Fluence-modulated proton CT optimized with patient-specific dose and variance objectives for proton dose calculation

To cite this article: J Dickmann et al 2021 Phys. Med. Biol. **66** 064001

View the article online for updates and enhancements.

Physics in Medicine & Biology

PAPER

Fluence-modulated proton CT optimized with patient-specific dose and variance objectives for proton dose calculation

J Dickmann[1] , F Kamp[2,*], M Hillbrand[3], S Corradini[2], C Belka[2,4], R W Schulte[5] , K Parodi[1] , G Dedes[1] and G Landry[1,2]

[1] Department of Medical Physics, Fakultät für Physik, Ludwig-Maximilians-Universität München (LMU Munich), D-85748 Garching bei München, Germany
[2] Department of Radiation Oncology, University Hospital, LMU Munich, D-81377 Munich, Germany
[3] Institut für Radio–Onkologie, Kantonsspital Graubünden, CH-7000 Chur, Switzerland
[4] German Cancer Consortium (DKTK), D-81377 Munich, Germany
[5] Division of Biomedical Engineering Sciences, Loma Linda University, Loma Linda, CA 92354, United States of America
[*] Currently at Department of Radiation Oncology and Cyberknife Center, University Hospital of Cologne, D-50937 Cologne, Germany.

E-mail: guillaume.landry@med.uni-muenchen.de

Abstract

Particle therapy treatment planning requires accurate volumetric maps of the relative stopping power, which can directly be acquired using proton computed tomography (pCT). With fluence-modulated pCT (FMpCT) imaging fluence is concentrated in a region-of-interest (ROI), which can be the vicinity of the treatment beam path, and imaging dose is reduced elsewhere. In this work we present a novel optimization algorithm for FMpCT which, for the first time, calculates modulated imaging fluences for joint imaging dose and image variance objectives. Thereby, image quality is maintained in the ROI to ensure accurate calculations of the treatment dose, and imaging dose is minimized outside the ROI with stronger minimization penalties given to imaging organs-at-risk. The optimization requires an initial scan at uniform fluence or a previous x-ray CT scan. We simulated and optimized FMpCT images for three pediatric patients with tumors in the head region. We verified that the target image variance inside the ROI was achieved and demonstrated imaging dose reductions outside of the ROI of 74% on average, reducing the imaging dose from 1.2 to 0.3 mGy. Such dose savings are expected to be relevant compared to the therapeutic dose outside of the treatment field. Treatment doses were re-calculated on the FMpCT images and compared to treatment doses re-recalculated on uniform fluence pCT scans using a 1% criterion. Passing rates were above 98.3% for all patients. Passing rates comparing FMpCT treatment doses to the ground truth treatment dose were above 88.5% for all patients. Evaluation of the proton range with a 1 mm criterion resulted in passing rates above 97.5% (FMpCT/pCT) and 95.3% (FMpCT/ground truth). Jointly optimized fluence-modulated pCT images can be used for proton dose calculation maintaining the full dosimetric accuracy of pCT but reducing the required imaging dose considerably by three quarters. This may allow for daily imaging during particle therapy ensuring a safe and accurate delivery of the therapeutic dose and avoiding excess dose from imaging.

1. Introduction

Particle therapy treatment planning for irradiation of tumors requires a precise knowledge of a patient's anatomy (Engelsman *et al* 2013, Landry and Hua 2018), in particular of the stopping power relative to water relative stopping power (RSP). With volumetric maps of the RSP, the dose of protons or heavier ions can be calculated and optimized such that the prescription is deposited inside the tumor and surrounding healthy tissue is spared as much as possible (Weber *et al* 2012, Park *et al* 2015, Nakajima *et al* 2017). Inaccuracies in RSP maps

result in inaccurate calculations of the particles' range (Paganetti 2012) and can cause over-or under-dosages both in the tumor and the surrounding tissue and are worth minimizing. Remaining uncertainties are considered either directly in a robust optimization (Cubillos-Mesías *et al* 2017) or as additional margins around the tumor and necessarily increase the dose to healthy tissue. In current clinical practice, RSP maps are calculated from x-ray computed tomography (CT) images by conversion of the x-ray attenuation coefficient to RSP, introducing uncertainties of up to 3% (Yang *et al* 2012). Recently, dual-energy x-ray CT has become available in the clinic, reducing the RSP uncertainty to around 1% (Hünemohr *et al* 2014, Hudobivnik *et al* 2016, Wohlfahrt *et al* 2017, Bär *et al* 2018, Taasti *et al* 2018, Niepel *et al* 2020). To limit the patient's exposure to imaging dose and due to time constraints during treatment, CTs are typically not acquired prior to every fraction of the treatment and also not in treatment position. More frequent and ideally daily imaging could detect anatomical changes that inevitably occur during treatment. Since this is currently not possible, additional margins have to be considered to cover such changes (Wedenberg *et al* 2018). In the future, an improvement of RSP accuracy and daily imaging in treatment position with low imaging dose may be possible using proton CT (pCT), which was proposed by Cormack (1963) and later realized by Hanson *et al* (1977). It directly uses the energy loss of protons for tomographic imaging of the RSP. This achieves a better dose efficiency than x-ray CT assuming ideal detectors (Schulte *et al* 2005) and a pre-clinical prototype scanner produced a comparable accuracy to state-of-the-art clinical dual-energy x-ray CT (Dedes *et al* 2019).

Unlike traditional radiotherapy, the clinically relevant dose in particle therapy typically covers a small fraction of the image used for treatment planning since only a few treatment fields are used and particles stop inside the tumor. This enables the application of fluence field modulation (Graham *et al* 2007, Bartolac *et al* 2011) to pCT as proposed by Dedes *et al* (2017). Fluence-modulated pCT (FMpCT) further reduces the imaging dose by maintaining image quality in a region-of-interest (ROI), i.e. the vicinity of the treatment beam path, but reducing imaging fluence elsewhere. Dickmann *et al* (2020) proposed a three-step optimization algorithm for FMpCT which calculates imaging fluences achieving a target image variance map. It used an iterative variance forward-projection approach allowing to solve for the fluence modulation for each projection independently. With this, dose savings of up to 40% outside of the ROI could be achieved, outperforming the simple intersection-based approach used in Dedes *et al* (2017). Optimized fluence modulations were employed experimentally by modulating pencil beams and using a prototype pCT scanner resulting in good agreements between simulated and experimental scans in terms of image variance and RSP accuracy (Dickmann *et al* 2020a). The studies of Dickmann *et al* (2020, 2020a) were limited to phantoms with tissue-equivalent materials and the optimization algorithm only took into account image variance and not imaging dose. Dose savings were achieved implicitly by prescribing higher image variance outside of the ROI, and it remained unclear if the chosen variance level achieved the optimal dose saving, and whether this level was the lowest achievable.

In this study we propose a novel FMpCT optimization algorithm which jointly optimizes imaging fluences for dose and image variance targets. The optimization requires knowledge of the patient's RSP and therefore an initial scan at uniform fluence. Alternatively, this information could be obtained from a previous x-ray CT scan. It allows to achieve a desired variance level inside the ROI while minimizing imaging dose outside, avoiding the need to prescribe an arbitrary high variance outside of the ROI. At the same time imaging dose inside the ROI and variance outside can be disregarded in the cost function. Moreover, we can define imaging organs-at-risk (OARs) where the imaging dose saving is reinforced. Such an optimization can be computationally expensive since the cost function is defined in image domain while fluences are modulated in projection domain, thus requiring one dose calculation and one variance reconstruction in each iteration of the algorithm. Adapting concepts of treatment plan optimization (Bortfeld 1999, Scholz *et al* 2003), we formulated the algorithm using sparse matrix multiplications, which can be executed efficiently using specialized libraries. In a simulation study, we applied the FMpCT algorithm to three pediatric cases of brain tumors to assess the accuracy of the resulting ROI variance and the potential imaging dose savings. Therapeutic doses were re-calculated on the FMpCT images as well as un-modulated pCT images to evaluate the dosimetric accuracy of FMpCT scans compared to a ground truth RSP image and to standard pCT.

2. Materials and methods

2.1. Simulation framework

Data for this study was simulated using the GEANT4 Monte Carlo simulation framework (Agostinelli *et al* 2003) and a detailed implementation (Giacometti *et al* 2017) of the phase-II prototype pCT scanner (Bashkirov *et al* 2016, Johnson *et al* 2016). The simulation framework produces equivalent output to the one of the physical scanner, which can be processed using the same reconstruction chain. It was validated for its fidelity in terms of RSP (Giacometti *et al* 2017, Dedes *et al* 2019). In terms of prediction of the image variance, its accuracy was estimated by Dickmann *et al* (2019) to be better than 7% (root-mean-square error).

The prototype pCT scanner consists of two tracking detectors, one prior and one after the object, as well as a longitudinally segmented energy detector, which was fully modeled in the Monte Carlo code. The two silicon strip tracking detectors measure both position and direction of travel of each incident proton. The plastic scintillator energy detector produces five independent energy measurements, one in each of its longitudinal segments (called stages). An empiric calibration (Bashkirov *et al* 2016) is used to infer the protons' water-equivalent path length (WEPL) from those energy measurements. Eventually, the input to the reconstruction algorithm are two positions, two directions and the WEPL.

To simulate modulated pencil beams, we used the pencil beam model of Dickmann *et al* (2020) with an elliptical Gaussian profile and a small divergence. The beam shape was elliptical because standard deviations were different for the horizontal coordinate u and the vertical coordinate v due to operation in research mode. In accordance with recent experiments (Dickmann *et al* 2020a) at the Northwestern Medicine Chicago proton center using the phase-II pCT scanner the standard deviations were chosen as $\sigma_u = 3.7$ mm and $\sigma_v = 2.9$ mm and in agreement with Dickmann *et al* (2020) the divergence was $\delta_u = 5.2 \times 10^{-4}$ mm^{-1} and $\delta_v = 5.8 \times 10^{-4}$ mm^{-1}. The smaller spot size compared to Dickmann *et al* (2020) was necessary due to an upgrade of the treatment facility. Horizontally, $N_{\mathrm{PB},u} = 51$ columns of pencil beams were interspaced by 5.6 mm and vertically $N_{\mathrm{PB},v} = 21$ rows were interspaced by 4.5 mm resulting in a total of 1071 pencil beams covering 285.6 mm $\times$ 94.5 mm.

2.2. Image and variance reconstruction

Data were simulated in step-and-shoot mode for 360 degrees in 1 degree steps. Before image reconstruction, protons of one projection angle are grouped together based on their coordinates u and v at the front tracker in bins of 2 mm $\times$ 2 mm. In each of those bins, distributions of direction angles and WEPL are estimated and protons outside of a three standard deviation interval are rejected (Schulte *et al* 2005).

For each of the remaining protons, a most likely path (Schulte *et al* 2008) is calculated based on the tracking information. This allows to estimate the track coordinates $(u(d), v(d))$ at any distance d between the front and the rear tracker. To exploit this path information, three-dimensional discretized projections are calculated on a grid of 1 mm $\times$ 3 mm $\times$ 1 mm in u, v and d. Each voxel of such a projection contains the mean WEPL value of all protons intersecting it as described in Rit *et al* (2013). Additionally, variance projections can be calculated as the WEPL variance of all intersecting protons divided by the number of intersecting protons (Dickmann *et al* 2019). Such projections are labeled with v^α for a given rotation angle α. A third type of three-dimensional projections containing only the number of protons intersecting a given voxel will be called counts projection and labeled with f^α.

To calculate volumetric RSP maps from the WEPL projections using a cone-beam filtered backprojection algorithm for pCT from Rit *et al* (2013), each projection at each depth d is convolved individually with a ramp filter before backprojection. When backprojecting, the algorithm chooses the optimal binning depth d which corresponds to a voxel's position along the beam for a given rotation angle α. This way the path information is fully exploited and spatial resolution improved (Rit *et al* 2013).

Analogously, image variance maps can be reconstructed from the variance projections as suggested for pCT by Rädler *et al* (2018) and validated in Dickmann *et al* (2019). Instead of applying a ramp filter, the projections are convolved with the squared elements of the ramp filter, and then backprojected as for the WEPL projections (see equations (16) and (20) in Rädler *et al* 2018). This produces an image variance map from a single dataset. Calculating the variance from a set of protons in each pixel covers all possible sources of noise without the need to model single contributions such as energy straggling or multiple Coulomb scattering (Rädler *et al* 2018, Collins-Fekete *et al* 2020).

2.3. Dose and variance optimization algorithm

In this work, we propose an optimization algorithm for FMpCT which calculates pencil beam weights achieving given objectives in terms of spatial distributions of image variance and imaging dose. The workflow of the algorithm is illustrated in figure 1. We first describe forward models predicting the image variance and imaging dose for a given incident fluence. The forward models are based on Monte Carlo simulations which requires knowledge of the patient's RSP, which could be obtained through an initial scan at uniform fluence. Alternatively, an x-ray CT scan can be imported to the simulation as described in section 2.7. The algorithm then uses these forward models in a bixel-wise fluence optimization. This step is illustrated in figure 1(a) and concepts are similar to early optimization of intensity-modulated radiotherapy (Bortfeld 1999, Markman *et al* 2002) which were also used in proton therapy in more recent studies (Wilkens and Oelfke 2006, Kamp *et al* 2017). Secondly, we optimize pencil beam weights approaching this bixel-wise fluence. This step is illustrated in figure 1(b). All equations in this section are formulated in parallel beam geometry, which is a fair assumption, given that the virtual source of the pencil beams is at 1.8 m from the isocenter (Dickmann *et al* 2020). The
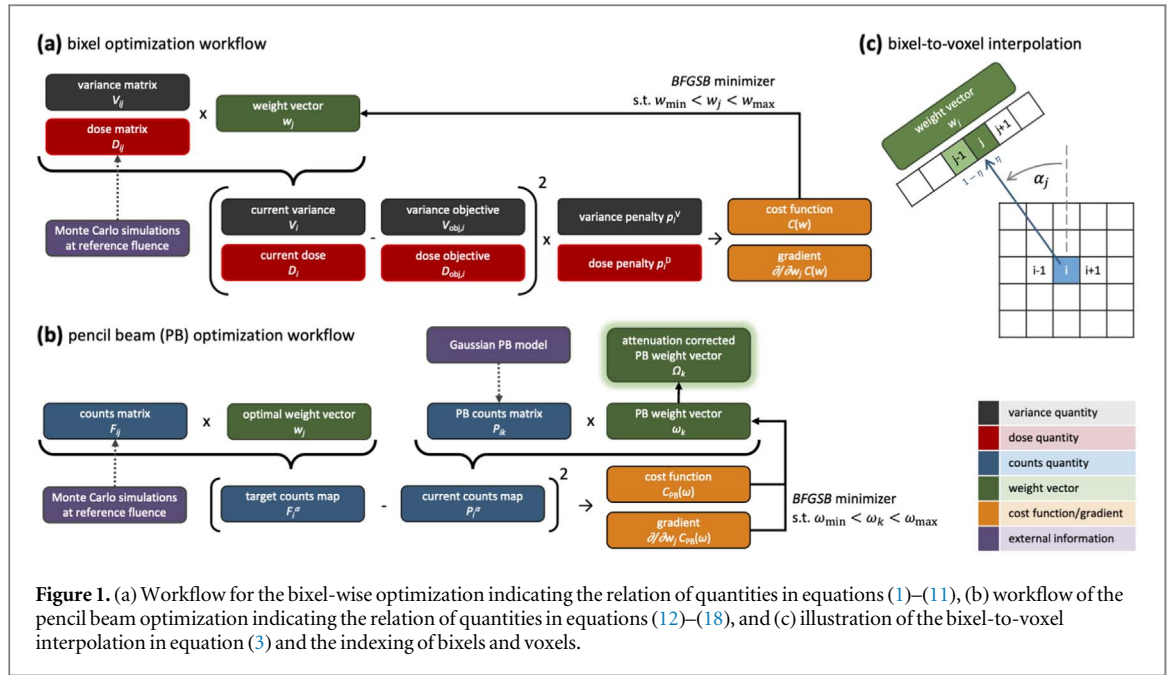
Figure 1. (a) Workflow for the bixel-wise optimization indicating the relation of quantities in equations (1)–(11), (b) workflow of the pencil beam optimization indicating the relation of quantities in equations (12)–(18), and (c) illustration of the bixel-to-voxel interpolation in equation (3) and the indexing of bixels and voxels.

divergence of the pencil beams, however, is considered implicitly in the final pencil beam optimization in section 2.3.5.

To establish a forward model for image variance and imaging dose, we discretize the incident fluence into bixels as illustrated in figure 1(c). Each bixel refers to a virtual detector element of size 4 mm × 4 mm at a given rotation angle. Bixels are numbered for all rotation angles consecutively with the bixel index $j \in \{1, \ldots, M\}$. In this work $M = N_{\mathrm{P}} \cdot N_{\mathrm{b},u} \cdot N_{\mathrm{b},v} = 518\,400$ with $N_{\mathrm{P}} = 360$ projections, $N_{\mathrm{b},u} = 60$ bixels along the u coordinate of the virtual detector and $N_{\mathrm{b},v} = 24$ bixels along the v coordinates of the virtual detector. Each bixel is associated to a rotation angle α_j. Additionally, the image volume is discretized using $N_x = N_y = 60$ virtual voxels in the left–right and anterior–posterior directions and $N_z = 24$ virtual voxels in the superior–inferior direction, with virtual voxel size of 4 mm × 4 mm × 4 mm. Voxels are numbered consecutively with the voxel index $i \in \{1, \ldots, N\}$ with $N = N_x \cdot N_y \cdot N_z = 86\,400$. Each bixel's center on the detector is denoted as (u_j, v_j) and each voxel's center is denoted as (x_i, y_i, z_i).

2.3.1. Imaging dose forward model

The forward model for imaging dose D_i at voxel i is formulated as a matrix multiplication of the dose matrix D_{ij} with a weight vector w_j as

$$D_i = \sum_{j=1}^{M} D_{ij} \cdot w_j. \tag{1}$$

The weight vector w_j describes the relative fluence modulation at bixel j with respect to a given reference fluence. This is similar to the approach in Scholz *et al* (2003). The dose matrix D_{ij} can be calculated from the dose $d_i^{\alpha_j}$ scored in a Monte Carlo simulation at the homogeneous reference fluence for a single projection angle α_j and voxel i. Instead of scoring the dose for each bixel individually, the full projection dose is sliced into bixel contributions and the dose matrix will then be

$$D_{ij} = d_i^{\alpha_j} \cdot \delta_{ij}, \tag{2}$$

where in the simplest case δ_{ij} would be 1, if bixel j intersects with voxel i and 0 otherwise. To be more accurate this was implemented as a linear interpolation between two voxels, and therefore

$$\delta_{ij} = \begin{cases} \eta_{ij} & \text{if } \lfloor (x_i \cos \alpha_j - y_i \sin \alpha_j - u_j)/u_j \rfloor = 0 \text{ and } z_i = v_j \\ 1 - \eta_{ij} & \text{if } \lceil (x_i \cos \alpha_j - y_i \sin \alpha_j - u_j)/u_j rceil = 0 \text{ and } z_i = v_j, \\ 0 & \text{else} \end{cases} \tag{3}$$

where $\lfloor \cdot \rfloor$ is the floor operator, $\lceil \cdot \rceil$ is the ceil operator and the interpolation fraction η_{ij} calculates as

$$\eta_{ij} = (x_i \cos \alpha_j - y_i \sin \alpha_j - u_j)/u_j - \lfloor (x_i \cos \alpha_j - y_i \sin \alpha_j - u_j)/u_j \rfloor. \tag{4}$$

The bixel-to-voxel interpolation is illustrated in figure 1(c) where two bixels correspond to one voxel. Consequently, the matrix D_{ij} is sparse and only has two non-zero elements in every row or column.

2.3.2. Image variance forward model

The forward model for the image variance V_i in voxel i is formulated analogously to equation (1) as

$$V_i = c \cdot \sum_{j=1}^{M} V_{ij} \cdot \tilde{w}_j, \tag{5}$$

where c is a constant which will be defined later, V_{ij} is the variance matrix and $\tilde{w}_j = 1/w_j$ are the inverse fluence weights, since variance is inversely proportional to the fluence. The variance matrix V_{ij} can be calculated from the variance projections of a simulation performed at the reference fluence as described in section 2.2. Those variance projections, which are defined in the (u, v, d) coordinate system are rotated by their corresponding rotation angle and interpolated at (x_i, y_i, z_i). The rotated variance projection at rotation angle α_j will then be called $v_i^{\alpha_j}$ and the variance matrix is consequently defined as

$$V_{ij} = v_i^{\alpha_j} \cdot \delta_{ij}, \tag{6}$$

where δ_{ij} is defined by equation (3). The constant c is defined as

$$c = f_{\text{interp}} \cdot f_{\text{filter}} \cdot \left(\frac{\pi \Delta u}{N_{\text{P}}} \right)^2, \tag{7}$$

with $\Delta u = 4$ mm the bixel size, $f_{\text{interp}} = 2/3 - 2/\pi^2 \approx 0.46$ an interpolation term (Rädler *et al* 2018) and $f_{\text{filter}} \approx 1.33/(2\Delta u)^4$ a filter term. The filter term is a consequence of ignoring the filtration with the squared ramp filter and simply using the sum of the filter elements, assuming variance projection values to be locally constant within the extent of non-negligible filter elements, as done in equation (4) in Hsieh and Pelc (2014). It calculates as

$$f_{\text{filter}} = \frac{1}{(2\Delta u)^4} + \frac{2}{(\pi \Delta u)^4} + \frac{2}{(3\pi \Delta u)^4} + \frac{2}{(5\pi \Delta u)^4} + \ldots \approx \frac{1.33}{(2\Delta u)^4}. \tag{8}$$

2.3.3. Bixel-wise optimization

Using equations (1) and (5) one can predict imaging dose and image variance maps for any arbitrary bixel-wise fluence modulation with weights w_j (from which $\tilde{w}_j$ can be calculated). Using optimization, we tried to achieve a certain imaging dose objective $D_{\text{obj},i}$ and a variance objective $V_{\text{obj},i}$ in voxel i. We also imposed a dose penalty p_i^{D} and a variance penalty p_i^{V}, which describe the relative strength with which a certain violation of the objective in voxel i is to be avoided. The bixel-wise optimization is illustrated in figure 1(a) and objectives and penalties will be defined in section 2.9.

To optimize bixel weights w_j the cost function

$$C(w) = \sum_{i=1}^{N} p_i^{\text{D}} (D_i(w) - D_{\text{obj},i})^2 + \sum_{i=1}^{N} p_i^{\text{V}} (V_i(w) - V_{\text{obj},i})^2 \tag{9}$$

was used, where the dependence of D_i and V_i on the weights w_j is stated explicitly. The gradient with respect to one weight w_j can then be expressed as

$$\frac{\partial}{\partial w_j} C(w) = 2 \sum_{i=1}^{N} p_i^{\text{D}} (D_i(w) - D_{\text{obj},i}) D_{ij} - 2 \sum_{i=1}^{N} p_i^{\text{V}} (V_i(w) - V_{\text{obj},i}) \frac{V_{ij}}{w_j^2}. \tag{10}$$

With this, the optimized weights $\hat{w}_j$ can be found as

$$\hat{w} = \arg\min_w C(w) \text{ s.t. } w_{\min} \leqslant w_j \leqslant w_{\max} \ \forall j, \tag{11}$$

where the limits $w_{\min} \geqslant 0$ and $w_{\max} > w_{\min}$ ensure that only physical (non-negative and finite) weights are allowed. The optimization is performed using the limited-memory BFGSB algorithm (Zhu *et al* 1997), which is a quasi-Newton method that minimizes the cost function equation (9) along the known gradient equation (10) subject to simple bounds equation (11).

2.3.4. Target counts maps

Since experimentally fluence can only be modulated using pencil beams (and not bixels), the problem is first transformed to maps of a target proton number and then, as described in the next section, to relative pencil beam weights. Similar to the matrix multiplications before, we can define a projection counts map in image space for a given rotation angle α as

$$F_i^{\alpha} = \sum_{j \in \{j_\alpha\}} F_{ij} \cdot w_j, \tag{12}$$

where the set $\{j_\alpha\}$ contains all j for which $\alpha_j = \alpha$. The matrix F_{ij} is defined as

$$F_{ij} = f_i^{\alpha_j} \cdot \delta_{ij}, \tag{13}$$

where $f_i^{\alpha_j}$ is the three-dimensional counts projection described in section 2.2 and δ_{ij} is as defined in equation (3). Those counts maps contain only protons used for reconstruction and are not equivalent to the initial proton fluence, from which protons are lost due to nuclear interactions causing attenuation. In presence of an object $f_i^{\alpha_j}$ will be reduced compared to the incident fluence, which will be accounted for in the following step. Using the optimized weights $\hat{w}_j$ from equation (11) target projection counts maps $\hat{F}_i^\alpha$ are generated using equation (12).

2.3.5. Pencil beam optimization

To calculate relative pencil beam weights, which are needed to employ the fluence modulation in a simulation or experiment, a second optimization needs to be performed that tries to achieve the bixel fluence using pencil beams. This step is illustrated in figure 1(b). It uses the same pencil beam model used in the simulation (Dickmann *et al* 2020). For each of the $K = N_P \cdot N_{PB,u} \cdot N_{PB,v} = 385\,560$ pencil beams, we used the pencil beam model to calculate the counts map P_{ik} of pencil beam k in voxel i. The counts map was normalized to have a maximum value of 1. Each pencil beam is associated with one rotation angle α_k. For given pencil beam weights ω_k and a selected rotation angle α, the summed pencil beam counts map P_i^α can be calculated as

$$P_i^\alpha = \sum_{k \in \{k_\alpha\}} P_{ik} \cdot \omega_k, \tag{14}$$

where the set $\{k_\alpha\}$ contains k for which $\alpha_k = \alpha$. Please note the Greek notation of pencil beam weights ω_k which is different to the Latin notation of bixel weights w_j. P_{ik} can directly be calculated from the analytical pencil beam model by calling the Gaussian model on a fine grid with voxels of 1 mm × 1 mm × 1 mm, rotating the map by its rotation angle α_k and resampling it to the coarse optimization grid with voxels i. Due to these operations, no interpolation as in equations (2), (6) and (13) is needed. To avoid unnecessary calls to the Gaussian function, P_{ik} was set to zero if the distance between voxel i and pencil beam k was more than three standard deviations. Eventually, we can formulate the cost function

$$C_{PB}(\omega) = \sum_{i=1}^{N} (P_i^\alpha(\omega) - F_i^\alpha)^2 \tag{15}$$

and the corresponding gradient

$$\frac{\partial}{\partial \omega_k} C_{PB}(\omega) = 2 \sum_{i=1}^{N} (P_i^\alpha(\omega) - F_i^\alpha) P_{ik}, \tag{16}$$

with which pencil beam weights ω_k can be optimized as

$$\hat{\omega} = \arg\min_\omega C_{PB}(\omega) \text{ s.t. } \omega_{min} \leqslant \omega_k \leqslant \omega_{max} \ \forall k \in \{k_\alpha\}, \tag{17}$$

using the bounds ω_{min} and ω_{max} which in section 2.9 were chosen to be equal to the bounds w_{min} and w_{max}. Note, that only weights for one rotation angle α can be optimized and the remaining weights are found by changing α. This also allows for an efficient parallelization of the optimization.

As in Dickmann *et al* (2020), to account for attenuation affecting the F_i^α in equations (12) and (15), the pencil beam optimization in equation (17) is performed twice: once for the target projection counts map $\hat{F}_i^\alpha$ and once for a reference counts map $\dot{F}_i^\alpha$. This reference counts map can be generated using equation (12) with all weights $w_j = 1 \ \forall j$. The two optimizations result in weights $\hat{\omega}$ and $\dot{\omega}$, respectively, which are both affected by attenuation. Finally, the desired pencil beam weights Ω_k are found as the ratio of these two weights cancelling out the effect of attenuation, and thus

$$\Omega_k = \frac{\hat{\omega}_k}{\dot{\omega}_k}. \tag{18}$$

Ω_k must be in the the interval ω_{min} to ω_{max} and otherwise is thresholded. A second Monte Carlo simulation can now be run using the relative pencil beam weights Ω_k by simulating $\Omega_k \cdot N$ particles for pencil beam k instead of N particles in the initial simulation with uniform fluence, that served as a reference in equations (2), (6) and (13).

2.4. Optimization hard-and software

Fluence optimizations were performed on a computer with two Intel Xeon E5-2667 v4 processors at 3.2 GHz and with in total 16 physical cores. The machine was equipped with 252 GB memory, which was required to store the dose and variance matrices. No GPU was used. Such high-performance workstations are likely to be available in proton therapy centers.

We used a C++ implementation of the optimization algorithm from the Insight Toolkit (ITK) (McCormick *et al* 2014), which was also used for handling and manipulation of image data. The sparse matrices and vector operations were implemented in C++ using the Eigen3 library (Guennebaud and Jacob *et al* 2010).

Table 1. Clinical characteristics of the delivered photon plans for the three pediatric patients under investigation. Patient 1 had an sequential boost prescription. The last two rows list the proton gantry angles for treatment plan generation and the gantry angle used for range evaluation, both are in the international technical commission scale. Abbreviations: RT—radiotherapy, GA—gantry angle.

	Patient 1	Patient 2	Patient 3
Age at first RT	4.0 years	5.8 years	4.4 years
Tumor site	Left orbit	Pons	Parotid gland
Tumor type	Embryonal rhabdomyosarcoma	Glioma	Acute lymphatic leukemia
Prescription/Gy	50.4[a]/36.0	54.0	20.0
fractions	28	30	8
Proton GA/deg	30, 90	90, 180, 270	270, 315
Proton GA (range)/deg	90	180	270

[a] boost to primary tumor.

2.5. Binary fluence modulation

To compare the performance of our proposed algorithm, a more simple intersection-based fluence modulation (Dedes *et al* 2018) was implemented. A pencil beam was assigned a weight of $\Omega_k = 1$ if it intersected a target volume T_i and $\Omega_k = \omega_{\min}$ elsewhere. This target volume was 0 if the variance penalty p_i^{V} was 0 and 1 elsewhere, and thus $T_i = p_i^{\mathrm{V}} > 0$. Pencil beams weights can be found by simple scalar product

$$\Omega_k = \begin{cases} 1 & \text{for } \sum_{i=1}^{N} P_{ik} T_i > 0 \\ \omega_{\min} & \text{else} \end{cases}. \tag{19}$$

Essentially this will result in an image with the unit fluence variance inside the target volume and a dose reduction outside. Images acquired with this fluence modulation will be labeled *binary*.

2.6. Patient data and original photon plans

In this work, we simulated pCT scans and optimized proton plans for three pediatric patients that underwent photon radiotherapy of the head. Due to the limited field-of-view and longitudinal coverage of the pCT scanner, we limited this investigation to pediatric tumors of the head region. Pediatric patients in particular are more susceptible to damage induced by imaging radiation due to their age. They can, therefore, benefit more from a radiation dose reduction. Tumor sites and indications of the original photon treatments are listed in table 1 together with the prescription dose and the number of fractions of the delivered photon plans. The three patients were chosen to be representative for typical cases in clinical practice with different tumor size and location within the brain. All plans were delivered using a 6 MV photon beam. Patient 1 had a photon plan with a sequential boost to the primary tumor, which is why two dose levels are reported. This plan was realized for protons as an integrated boost. Dose-volume histograms of the photon plans can be found in the supplementary material (available online at stacks.iop.org/PMB/66/064001/mmedia).

2.7. Study design and proton treatment plans

To assess the accuracy of pCT and FMpCT scans, we first created ground-truth RSP maps from the treatment planning CTs using the GEANT4 simulation code. We then imported the RSP maps to the treatment planning system (TPS) RayStation (*RaySearch Laboratories AB*, Stockholm, Sweden) and generated proton treatment plans while trying to match or improve dose-volume statistics of the delivered photon plans. Using the simulation code we then created uniform fluence pCT scans as well as FMpCT scans with modulation weights calculated by the optimization algorithm. Both pCT and FMpCT scans were then imported to RayStation as additional image sets where doses could be re-calculated to evaluate the accuracy of pCT and FMpCT scans with respect to the ground truth RSP map.

Ground-truth RSP maps were generated from a voxelized geometry in the GEANT4 simulation code using the x-ray CT images. Material and densities of the voxelized geometry were found from a piecewise-linear calibration curve used in Schmid *et al* (2015) and Resch *et al* (2017). In the GEANT4 code, each voxel of the geometry was then queried and its stopping power value relative to the one of water was output for a proton energy of 150 MeV. These three-dimensional maps served as ground truth to compare pCT and FMpCT scans to.

To import RSP maps back to the TPS, we generated a voxelized geometry with CT values ranging from −1000 to 4000 HU. The resulting RSP values were matched to the original CT values resulting in a RSP-to-CT value mapping. Since the same RSP value can be produced by two different CT values (due to a change of the elemental composition), the mapping was forced to be bijective by fitting continuous splines to local subsets of

Table 2. Fluence optimization objectives and imaging OARs for the three patients investigated in this study. The variance objective is also reported as RSP standard deviation (σ_{obj}). Abbreviations: OAR—organ at risk, r—right, l—left.

	Patient 1	Patient 2	Patient 3
Dose objective D_{obj}/mGy	0.0	0.0	0.0
Variance objective V_{obj}	5.41×10^{-4}	6.09×10^{-4}	6.72×10^{-4}
σ_{obj}	0.023	0.025	0.026
Imaging OARs	r. eye, brainstem	r./l. eye, r./l. optical nerve	r./l. eye, spinal cord

the data. The patients' RSP maps were subsequently run through the mapping to convert them back to CT values before importing them to the TPS. While this resulted in small differences between the original CT and the imported ground-truth CT, it allowed to import pCT and FMpCT scans using the same mapping and enabled us to compare them to the ground truth scan using the TPS.

Using the initial delineations and dose prescriptions of the photon plans, we then optimized proton treatment doses in the TPS using the clinical target volume (CTV) as prescription volume and the ground truth RSP maps. Gantry angles used for the proton plans are listed in table 1. For all patients, a generic IBA machine (`RSL_IBA_DED`) was used with a 40 mm range shifter and an air gap of 100 mm. The dose grid was uniformly 3 mm. Beam spots were interspaced laterally by 3 mm and in depth the adaptive spacing with energy of the TPS was used. Treatment plans were obtained using a robust optimization accounting for a positional uncertainty of 3 mm and a range uncertainty of 3%. For all OARs and the CTV we achieved a comparable or better dose distribution using protons with respect to photon plans. A comparison of dose-volume histograms of the photon and proton plans can be found in the supplementary material.

2.8. Unit fluence pCT scans
Besides the ground truth RSP maps, we simulated pCT scans with all pencil beam weights set to $\Omega_k = 1$ and 933 protons per pencil beam, which resulted in about 360 million primary protons per tomography, an incident proton fluence of 37 mm^{-2} (approximately 26 mm^{-2} after data cuts) and an imaging dose of 1.2 mGy. These are typical values for experimental scans with this prototype scanner (Johnson *et al* 2016, Dickmann *et al* 2020a). Throughout this paper these reconstructions are referred to as pCT and serve as a reference for a standard pCT scan performed with a prototype pCT scanner today.

### 2.9. Objectives for fluence optimization
To optimize FMpCT scans we used the optimization algorithm described in section 2.3. The dose matrix D_{ij} and the variance matrix V_{ij} were generated from the unit fluence scans described in section 2.8. To define the dose and the variance penalties p_i^{D} and p_i^{V}, we calculated a ROI volume by thresholding the ground truth dose map. The ROI was defined as all voxels with at least 10% of the prescription dose, which is in accordance to the recommendation of the AAPM task group 119 for gamma analysis (Ezzell *et al* 2009, Song *et al* 2015). The ROI volumes were 223 cm^3 for patient 1785 cm^3 for patient 2, and 321 cm^3 for patient 3.

The imaging dose objective was set to $D_{\mathrm{obj},i} = D_{\mathrm{obj}} = 0$ mGy throughout the volume, thus minimizing dose. The variance objective was also set to a constant value $V_{\mathrm{obj},i} = V_{\mathrm{obj}}$ throughout the volume, which was chosen as the 95th-percentile of variance values in the unit fluence scan inside the ROI and was therefore different between the patients. Prescription values are listed in table 2.

Both p_i^{D} and p_i^{V} were set to 0 outside of the patients' skin. p_i^{V} was also set to 0 outside of the ROI and to 100 inside the ROI. p_i^{D} was set to 0.01 inside the ROI and to 1 outside of the ROI. Furthermore, we defined imaging OARs where p_i^{D} was increased to 20. For each patient, the imaging OARs are listed in table 2. The choice of p_i^{V} and p_i^{D} was made empirically. With the relatively high p_i^{V} we compensated for the ROI being smaller than the complete volume, and made sure that $V_{\mathrm{obj},i} = V_{\mathrm{obj}}$ was achieved and not impaired by the dose minimization. The low p_i^{D} inside the ROI avoided too high doses for pencil beams that only intersected the ROI. By increasing p_i^{D} inside imaging OARs an additional dose saving should be achieved there. The minimum weight was set to $w_{\mathrm{min}} = \omega_{\mathrm{min}} = 0.05$ to ensure that at least 4 protons could be used for image reconstruction in each pixel on average. The maximum weight was $w_{\mathrm{max}} = \omega_{\mathrm{max}} = 10$, again to avoid too high doses inside the ROI.

2.10. Evaluation of image variance
To assess the performance of the proposed optimization algorithm to achieve V_{obj} inside the ROI, we performed variance reconstructions of the simulated pCT and FMpCT data as described in section 2.2 and compared them to V_{obj} inside the ROI, which was defined as the volume where $p_i^{\mathrm{V}} > 0$. To directly estimate the effect of fluence-modulation on the RSP, we also calculated the mean RSP error inside the ROI with respect to the ground truth RSP.
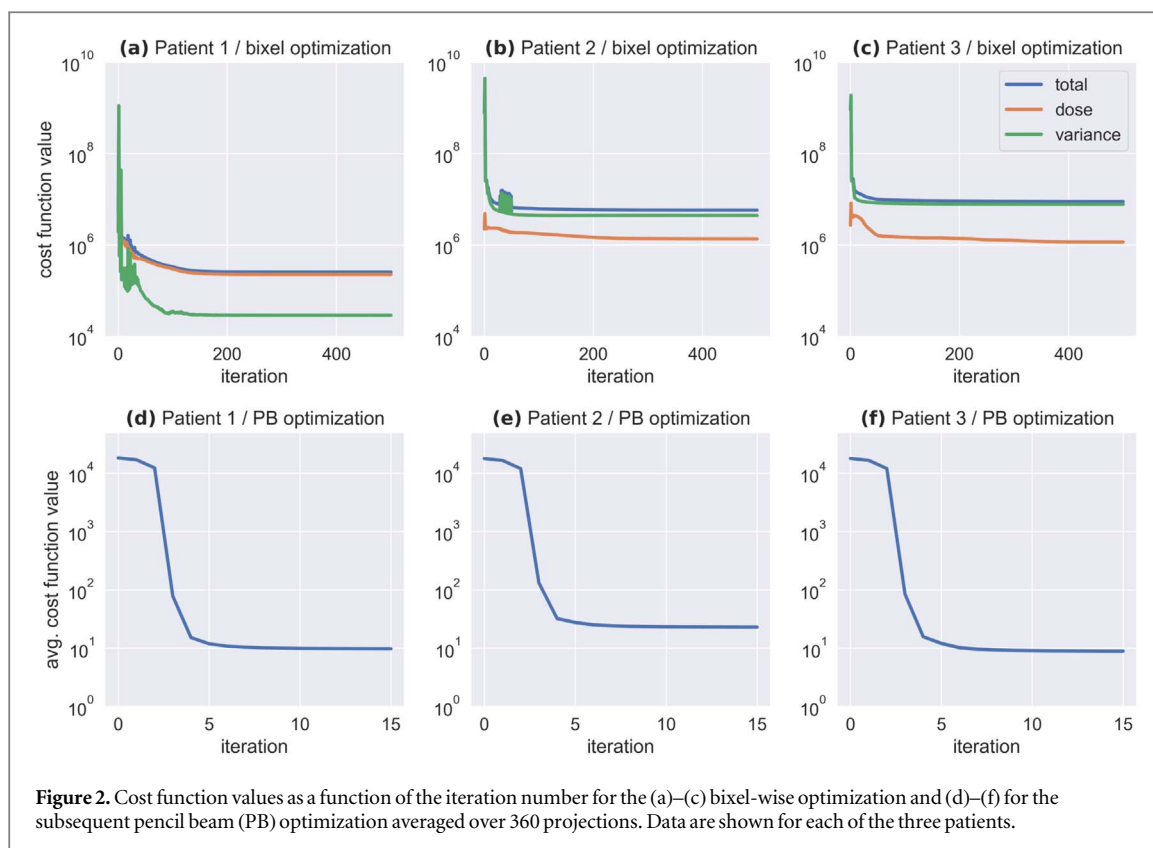
Figure 2. Cost function values as a function of the iteration number for the (a)–(c) bixel-wise optimization and (d)–(f) for the subsequent pencil beam (PB) optimization averaged over 360 projections. Data are shown for each of the three patients.

2.11. Evaluation of treatment dose accuracy

To evaluate the dosimetric accuracy of pCT and FMpCT scans, we imported them to the TPS as additional image sets. The therapeutic dose, which was originally optimized on the ground truth RSP map, was re-calculated on pCT and FMpCT. We visually assessed resulting dose distributions for a slice at the center of the CTV as well as with dose-volume-histograms computed by the TPS. Moreover, we calculated a passing rate for a relative dose difference criterion of 1% applied to the difference of the re-calculated dose to the ground truth dose and divided by the prescription dose. For the passing rate, the entire dose grid with voxels receiving more than 10% of the prescription was considered.

2.12. Evaluation of range accuracy

To evaluate the dosimetric accuracy also in terms of proton range, we optimized a second set of treatment plans on the ground truth RSP map with just a single field and prescribing a uniform dose to the CTV. The gantry angle chosen for each patient is listed in table 1. The single field dose was re-calculated on pCT and FMpCT. The dose calculation was performed with a dose grid of 1 mm $\times$ 1 mm $\times$ 3 mm. The range was calculated in beam-eye-view for each voxel by determining the 80% dose-falloff. Thereby, we used linear interpolation to achieve sub-millimeter precision. The beam-eye-view map of ranges of the pCT and FMpCT scan was then compared to the ground truth range map for all voxels intersecting the CTV (for patient 1 the low dose CTV with 36 Gy).

2.13. Evaluation of imaging dose reduction

For all pCT and FMpCT scans, imaging doses were scored using GEANT4 as absorbed doses, summing contributions from all projections. We calculated dose-volume-histograms as well as median doses for all OARs as well as the ROI. Moreover, we also scored imaging dose for patient 2 using the simple binary fluence modulation described in section 2.5.

3. Results

3.1. Dose and variance optimization

In figure 2 cost function values as a function of the iteration number both for the bixel-wise optimization and the subsequent pencil beam optimization are shown for each of the three patients. In figures 2(a)–(c) the cost function of the bixel-wise optimization is shown for 500 iterations. For all patients it reduces quickly within 100–200 iterations. For patient 1 the cost function is dominated by the contribution of the dose term (the first

Figure 3. Image variance maps for the three patients for (a), (d), (g) pCT and (b), (e), (h) FMpCT scans overlaid on the corresponding RSP maps. Each ROI is indicated by a solid white line delineation. (c), (f), (i) Profiles along the white dashed lines of the variance maps with an indication of the ROI and the variance objective.
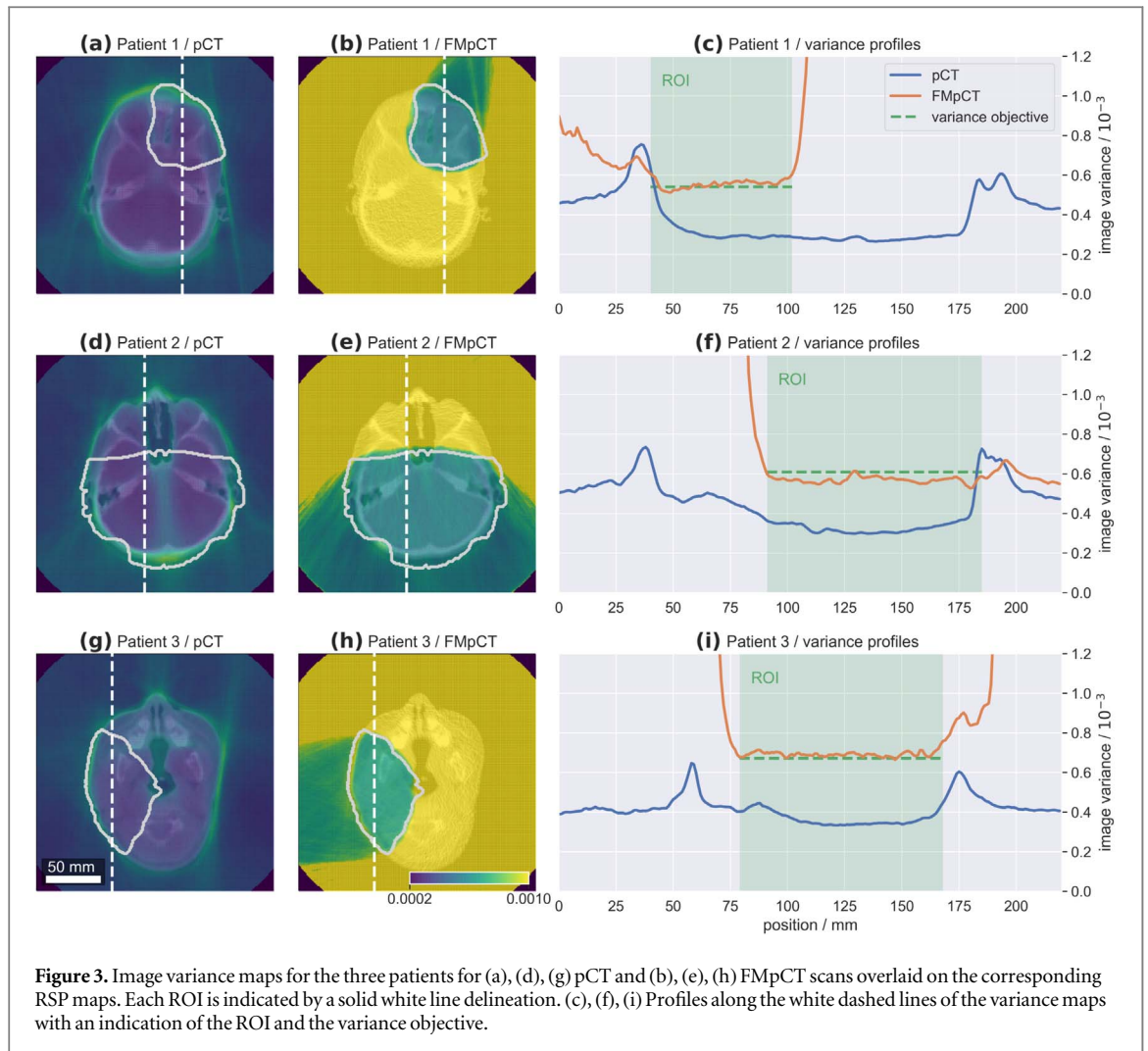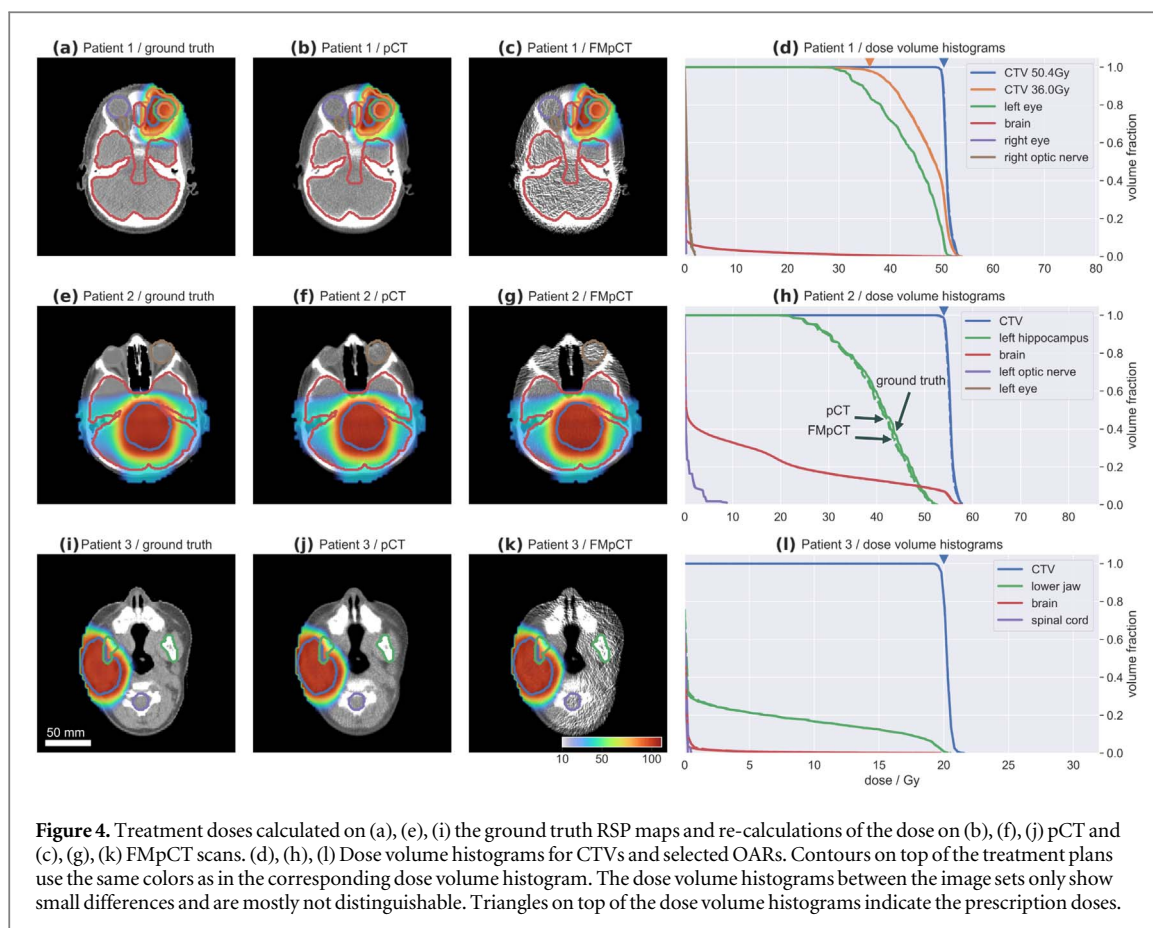
summand in equation (9)) while for patients 2 and 3 the variance term prevails. Even though for patients 1 and 3 one of the two contributions is an order of magnitude larger than the other, both reduce strongly within the first 100–200 iterations. In figures 2(d)–(f) the cost function of the pencil beam optimization averaged over all projections is shown. The cost function converges quickly within less than 10 iterations.

All optimizations were performed on the hardware described in section 2.4. One matrix multiplication for the evaluation of the bixel-wise optimization's cost function took 54 ms and one iteration of the optimization algorithm 302 ms. The total optimization time over 500 iterations was 151 s. The pencil beam optimization took 86 s in total or 238 ms per projection. The total optimization time, including creation of the optimization matrices was 19 min.

3.2. Evaluation of image variance

Figure 3 shows image variance maps for all three patients for pCT scans (a), (d), (g) as well as for FMpCT scans (b), (e), (h). While pCT scans have a low noise in homogeneous regions of the scan, noise is elevated close to the hull, but also close to heterogeneities such as in the nasal cavity or the pharynx. The FMpCT scan shows a homogeneous variance map inside the ROI and a sharp increase of variance outside. These observations are confirmed in the profile plots (c), (f), (i) where pCT and FMpCT variances are compared. Inside the ROI, FMpCT agrees well with its variance objective $V_{\mathrm{obj},i}$, which was the 95th percentile of variance values inside the entire ROI volume of the pCT scan. Profiles for pCT are therefore generally below those of FMpCT, even though they agree in particular at the edges. For patient 3, the 95th percentile value of the pCT scan (and thus the FMpCT variance target) was outside of the displayed slice and the two curves do not intersect. Outside of the ROI, variances for FMpCT increase sharply within a few centimeters.

The absolute mean RSP error relative to the ground truth RSP inside the ROI was below 0.6% for all patients when comparing the pCT images to the ground truth, and below 0.8% for all patients when comparing the FMpCT images to the ground truth.

Figure 4. Treatment doses calculated on (a), (e), (i) the ground truth RSP maps and re-calculations of the dose on (b), (f), (j) pCT and (c), (g), (k) FMpCT scans. (d), (h), (l) Dose volume histograms for CTVs and selected OARs. Contours on top of the treatment plans use the same colors as in the corresponding dose volume histogram. The dose volume histograms between the image sets only show small differences and are mostly not distinguishable. Triangles on top of the dose volume histograms indicate the prescription doses.

Table 3. Dose difference passing rates in percent for a 1% criterion comparing ground truth (GT) dose maps to those evaluated on the pCT and FMpCT scans.

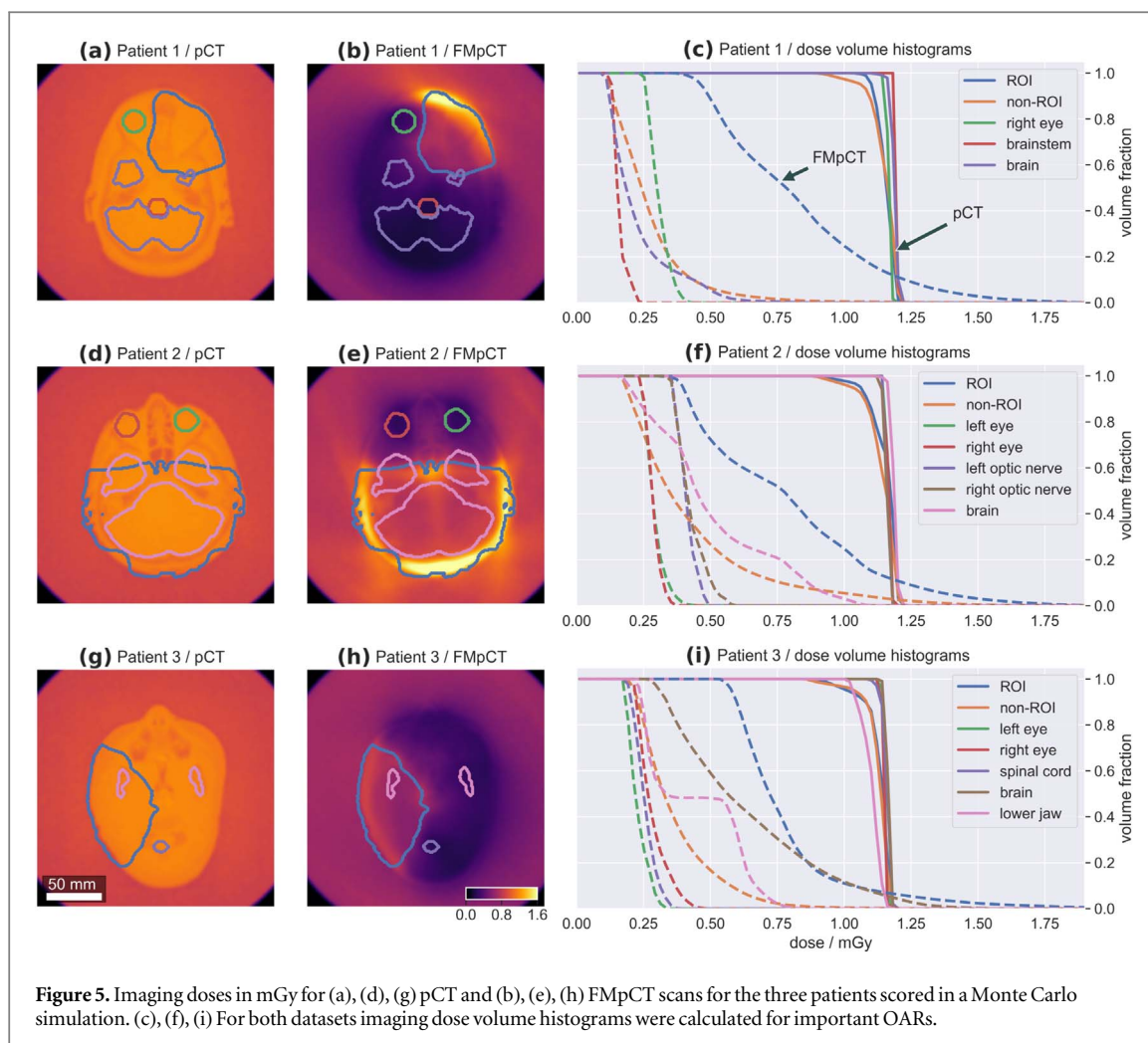| Comparison	Patient 1	Patient 2	Patient 3
pCT versus GT	90.6	91.8	90.6
FMpCT versus GT	91.3	93.1	88.5
FMpCT versus pCT	98.3	99.7	98.9

3.3. Evaluation of dose accuracy

In figure 4 treatment doses are displayed for all three patients together with dose volume histograms. The ground truth treatment dose (a), (e), (f) agrees well with the doses re-calculated on the pCT (b), (f), (j) and FMpCT (c), (g), (k) scans. This is confirmed in the dose volume histograms (d), (h), (j) where most lines are not distinguishable from each other. The largest difference can be seen for the left hippocampus of patient 2, which is close to the end of the range of one treatment field. For the FMpCT scans (c), (g), (k) an increased noise level can be observed outside of the area covered by the treatment dose. However, even organs like the brain of patient 1 show a good agreement in terms of the dose volume statistics.

Table 3 shows passing rates for a 1% criterion on the absolute difference between doses calculated on the three imaging sets. The lowest passing rate, thus the highest discrepancy of doses is observed for the comparison of FMpCT versus the ground truth dose for patient 3. Passing rates between pCT and ground truth as well as between FMpCT and ground truth are all around 90%, while passing rates between FMpCT and pCT are close to 100%.

3.4. Evaluation of range accuracy

Table 4 shows passing rates for a 1 mm criterion on range differences between single-field uniform doses calculated on the three imaging sets. All passing rates are well above 90% with some reaching 100%. The lowest passing rate is observed for the comparison of FMpCT and the ground truth dose and best agreement is between FMpCT and pCT. The table also shows mean absolute and mean range differences, which are all well below 0.5 mm. In general, a larger mean absolute difference correlates with a lower passing rate.

Figure 5. Imaging doses in mGy for (a), (d), (g) pCT and (b), (e), (h) FMpCT scans for the three patients scored in a Monte Carlo simulation. (c), (f), (i) For both datasets imaging dose volume histograms were calculated for important OARs.
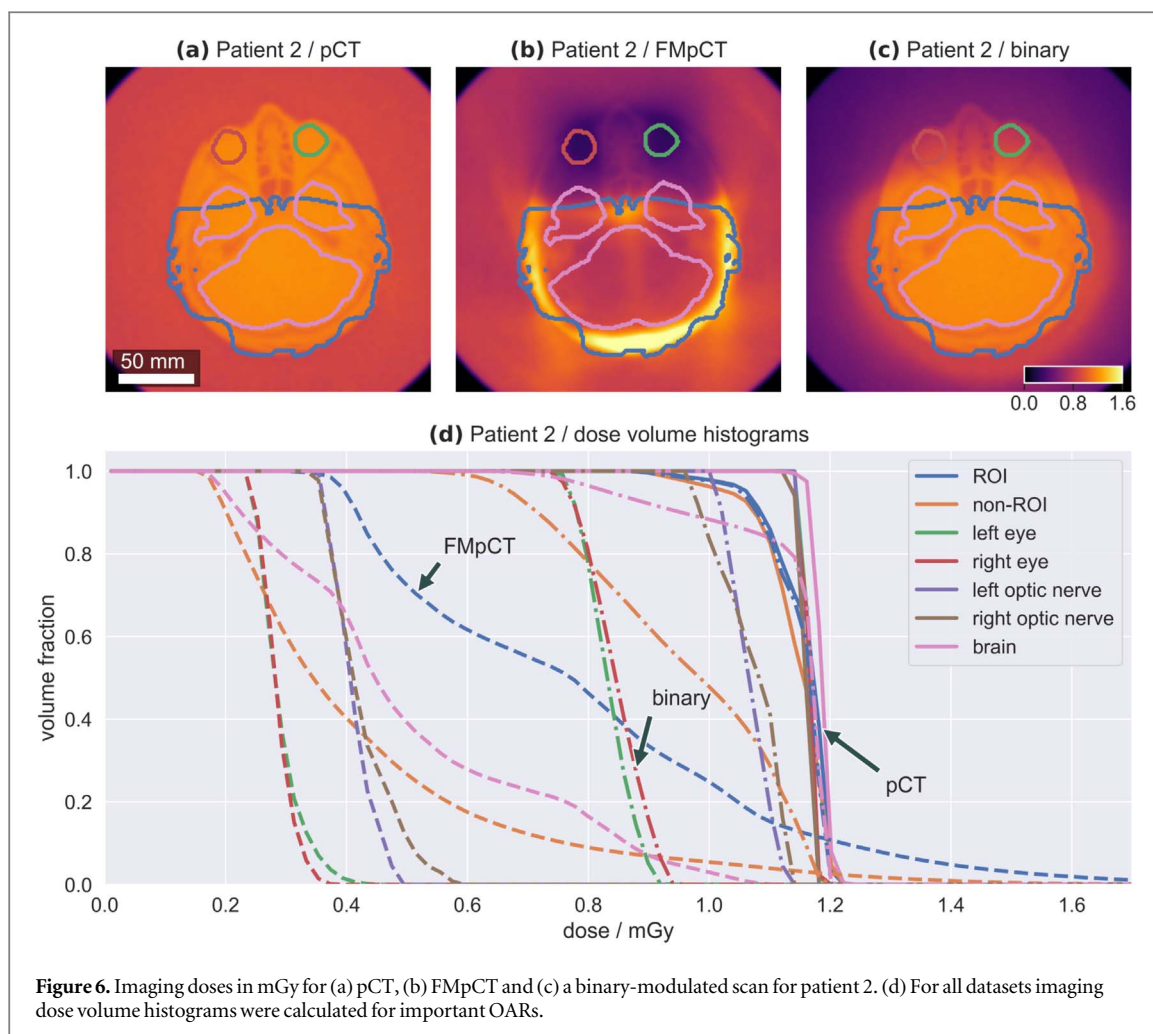
Table 4. Range difference passing rates for a 1 mm criterion comparing ground truth (GT) range maps to those re-calculated on the pCT and FMpCT scans together with corresponding mean absolute range differences and mean range differences.

Quantity	Comparison	Patient 1	Patient 2	Patient 3
Passing rate / %	pCT versus GT	97.4	100.0	95.3
	FMpCT versus GT	96.5	96.8	95.0
	FMpCT versus pCT	100.0	97.5	99.9
Mean absolute difference / mm	pCT versus GT	0.29	0.25	0.28
	FMpCT versus GT	0.31	0.32	0.32
	FMpCT versus pCT	0.15	0.17	0.05
Mean difference / mm	pCT versus GT	0.00	0.04	0.27
	FMpCT versus GT	0.15	0.20	0.31
	FMpCT versus pCT	−0.15	−0.16	−0.05

3.5. Evaluation of imaging dose reduction

In figure 5 imaging doses for pCT and FMpCT scans are shown for all three patients. The pCT scans (a), (d), (g) show a homogeneous dose at around 1.2 mGy per scan with reduced doses in bones and the nasal cavity. In contrast, FMpCT scans (b), (e), (h) show heterogeneous dose distributions with a clear dose reduction outside of the ROI. The dose generally reduces further away from the ROI. Imaging OARs listed in table 2, such as the right eye of patient 1 or the spinal cord of patient 3, show a reduced dose compared to their direct vicinity. Moreover, imaging dose inside the ROI is not homogeneous: it is elevated at the hull of the object and in general where variance was increased in the pCT scans as shown in figure 3. Dose in these regions can be higher than in the uniform fluence scans.

Figure 6. Imaging doses in mGy for (a) pCT, (b) FMpCT and (c) a binary-modulated scan for patient 2. (d) For all datasets imaging dose volume histograms were calculated for important OARs.

These observations are confirmed in the imaging dose volume histograms of figures 5(c), (f), (i) comparing pCT and FMpCT scans. The pCT scans all show doses of around 1.2 mGy for all OARs and the ROI. Lower doses occur where the elemental composition of the OAR is different (i.e. the lower jaw of patient 3) or where parts of air are included (some of the ROIs). As could be already seen visually, imaging doses for the FMpCT scan show increased peak doses for the ROI and for organs overlapping with the ROI of up to 1.8 mGy. These increased doses are limited to a small fraction of the ROI (less than 10%) and most of the ROI receives a reduced dose compared to the pCT scan. All OARs (except for those overlapping with the ROI) have a strongly reduced imaging dose in the FMpCT scan. The non-ROI volume, which contains all voxels inside the patient, but outside the ROI, receives a clearly reduced dose in the FMpCT scan and no dose above the dose of the pCT scan, except for patient 2, where 2.5% of the non-ROI volume receives a dose above 1.2 mGy.

Sinograms of the imaging fluence employed in all scans are included in the supplementary material.

In figures 6(a) and (b) imaging doses of the pCT and FMpCT scan for patient 2 are shown again. In addition, figure 6(c) shows the imaging dose for a binary fluence modulation according to section 2.5. Imaging dose volume histograms for the three scans are shown in (d). Inside the ROI, imaging dose of the binary modulation is equal to the pCT scan. The non-ROI region receives an imaging dose starting at the value inside the ROI and slowly reducing towards lower values. For the eyes, which were furthest apart from the ROI, imaging dose is considerably reduced, but still clearly above the imaging dose for FMpCT.

Table 5 summarizes median imaging doses for pCT, FMpCT and the binary-modulated scan. For the pCT scans imaging doses are constant across all patients and equal inside and outside of the ROI. Imaging doses and dose savings for the FMpCT scans are of comparable magnitude across the three patients. Outside of the ROI about three quarters of the imaging dose could be saved. For certain OARs, the imaging dose saving reaches values of -80% or more. The binary fluence had the same imaging dose inside the ROI and a moderate dose saving outside. For the eyes, which were furthest apart from the ROI, the dose saving reached up to -29%, but was still considerably less compared to the FMpCT scan.

Table 5. Median imaging doses in pCT, FMpCT and binary FMpCT scans for the ROI and relevant OARs. In parentheses the imaging dose saving compared to pCT is given. Abbreviations: ROI—region-of-interest, l—left, r—right.

	Region	pCT Dose/mGy	FMpCT Dose/mGy (saving/%)	Binary Dose/mGy (saving/%)
Patient 1	ROI	1.16	0.78 (−33)	—
	Non-ROI	1.16	0.24 (−80)	—
	r. eye[a]	1.17	0.30 (−74)	—
	Brainstem[a]	1.19	0.15 (−87)	—
	Brain	1.19	0.19 (−84)	—
Patient 2	ROI	1.17	0.77 (−35)	1.17 (0)
	Non-ROI	1.16	0.34 (−71)	0.98 (−16)
	l. eye[a]	1.17	0.28 (−76)	0.83 (−29)
	r. eye[a]	1.17	0.28 (−76)	0.85 (−27)
	l. optic nerve[a]	1.16	0.41 (−65)	1.07 (−8)
	r. optic nerve[a]	1.16	0.41 (−65)	1.08 (−7)
	Brain	1.19	0.45 (−62)	1.17 (− 2)
Patient 3	ROI	1.14	0.72 (−37)	—
	Non-ROI	1.14	0.33 (−71)	—
	l. eye[a]	1.15	0.22 (−81)	—
	r. eye[a]	1.15	0.27 (−77)	—
	Spinal cord[a]	1.16	0.25 (−78)	—
	Brain	1.17	0.56 (−52)	—
	Lower jaw	1.11	0.34 (−69)	—
Average	ROI	1.16	0.76 (−35)	1.17 (0)
	Non-ROI	1.15	0.30 (−74)	0.98 (−16)

[a] imaging OAR.

4. Discussion

4.1. Dose and variance optimization

The cost functions of both optimizations reached their minimum quickly and consistently across patients. While the bixel-wise optimization, which considered fluences from all angles at once, needed less than 200 iterations, the pencil beam optimization, which only considered fluences from one angle, reached convergence within less than 10 iterations. For the bixel-wise optimization the final value of the cost function depended on the patient and thus on the dose and variance objectives. While in this study we kept the penalties for dose and variance the same for all patients, these are typically adjusted by the user for treatment plan optimization (which is a similar task to fluence field optimization). The proposed method would allow to adapt these values to each patient, but we opted to not modify the penalties in order to demonstrate this effect and the robustness of the method. For patient 1 with the smallest ROI the cost function was dominated by the dose term while for the two larger ROIs of patients 2 and 3 it was dominated by the variance term. This trend is in agreement with the fact that for patient 1 less voxels contributed to the variance term and more to the dose term.

A difference to treatment plan optimization was that in this work it would not have been possible to avoid the bixel-wise optimization and directly optimize pencil beam weights. While it would be possible to formulate the dose in a given voxel as a linear combination of doses from single pencil beams, this is not possible for the variance term of the cost function due to the inverse relationship between fluence and variance. The non-overlapping nature of bixels was therefore required.

Due to the sparse implementation of the matrix multiplications, the evaluation of the cost function was possible in a few hundred milliseconds even though the matrix had $N \cdot M \approx 5 \times 10^{10}$ entries (most of them zero). The time needed for the fluence field optimization of one patient was mainly dominated by reading data from disk and the subsequent calculation of the variance, dose and fluence matrices. In particular the matrix P_{ik} in equation (14) required to query the pencil beam model for each pencil beam and each voxel of the fine grid. Since these matrices are independent of the dose and variance objectives, they could be calculated offline as soon as the uniform fluence pCT scan and the corresponding Monte Carlo simulation are available. The matrices would then be available prior to the next imaging session reducing the optimization time needed to a few minutes with the current configuration. Optimization time could further be decreased by reducing the number of iterations from 500–200 or even 100 given that the cost function only slightly reduced beyond iteration 100. This may reduce the time requirement to few tens of seconds. Since the code was not fully optimized for speed,

further improvements (e.g. not calculating variance values where the variance weight is zero), may reduce the optimization time to be comparable to the one needed for treatment planning. With this, fluence field optimization for FMpCT may be feasible with respect to time within a clinical workflow.

The forward model from the last iteration of the bixel-wise optimization agreed well with the quantities calculated from the subsequent Monte Carlo scans both for variance and imaging dose. A comparison is included in the supplementary material.

4.2. Evaluation of image variance

Analysis of the uniform fluence pCT scans showed an elevated image variance close to heterogeneities as shown previously in Dickmann *et al* (2019) using phantoms. Therefore, the subsequent FMpCT optimization used the 95th percentile uniform pCT variance as the variance objective inside the ROI. This way, pCT and FMpCT images were comparable in terms of peak image variance. Using the proposed algorithm, image variance objectives were met inside the ROI, where resulting variance maps were much more homogeneous compared to pCT. Outside of the ROI variance increased sharply, as intended to achieve the desired dose saving.

4.3. Evaluation of dose and range accuracy

The dosimetric accuracy of pCT and FMpCT was satisfactory both in terms of re-calculation of the treatment dose as well as the protons' range for a single uniform dose treatment field. Passing rates with a 1% (for treatment dose) or a 1 mm (for range) criterion were all above 88% with respect to the ground truth RSP image. In particular, passing rates comparing dose and range between pCT and FMpCT were all above 97%. Therefore, the principal contribution to dosimetric errors was caused by using pCT/FMpCT images instead of the ground truth RSP map. Errors of the magnitude observed in this study (1% or below) are expected for pCT and comparable to those introduced by the current state-of-the-art imaging using dual-energy x-ray CT (Dedes *et al* 2019). They also agree with the mean RSP error inside the ROI observed in the pCT and FMpCT images. Range differences observed for FMpCT compared to the ground truth map were slightly larger compared to those of pCT for some patients, but small compared to variations between patients. Therefore, a clear dosimetric difference between pCT and FMpCT compared to the ground truth could not be shown. In fact, differences between pCT and FMpCT compared directly were much smaller compared to the differences with the ground truth RSP map. We therefore conclude that FMpCT scans maintain the full dosimetric value that pCT scans have and confirm expected errors introduced by using protons for imaging (Meyer *et al* 2019).

4.4. Evaluation of imaging dose reduction

Using uniform fluence pCT, imaging doses were mostly constant across the patient while, as reported before, variance distributions were not. Instead, using FMpCT optimization constant variance inside the ROI could be achieved and imaging dose could be reduced outside of this ROI. However, because pCT and FMpCT image quality were matched in terms of peak noise, homogeneous regions of the ROI were also imaged with reduced imaging dose while regions with heterogeneities and increased noise in the uniform scan (like the nasal cavity) received an increased imaging dose with FMpCT. The median imaging dose inside the ROI was reduced by 35% on average over the three patients. However, this dose reduction is not very relevant in practice, since those volumes will also receive a considerably higher treatment dose in each fraction.

Outside of the ROI, median imaging dose could be reduced by 74% on average compared to the uniform fluence pCT scan. This dose saving was similar across the three patients (ranging from 71% to 80%), even though treatment fields and consequently optimization targets were considerably different. The largest dose saving was achieved for patient 1 who had the smallest ROI. Using the novel optimization allowed to further reduce imaging dose to imaging OARs. Dose savings in imaging OARs reached up to 87%, however the dose saving was less, the closer an OAR was located to the ROI. Estimates of the imaging dose volume statistics are available as part of the optimization. A selection of imaging OARs and a trade-off between OAR dose and dose to the remaining non-ROI tissue will need to be decided based on clinical criteria. In this study we selected the eyes, the optical nerve, the brainstem and the spinal cord as critical organs. The eye is more sensitive to radiation-induced damage which can cause cataract (Shore 2016). The other organs were selected as showcases and eventually a physician would need to define imaging OARs, if any.

For patient 1, the median variance in the ROI was 3.45×10^{-4}. Assuming that this were the variance objective instead of the peak variance 5.41×10^{-4} used in this study, we can estimate the resulting imaging doses by simply scaling up the doses with the inverse ratio of the two variance values. This would result in an imaging dose of 1.22 mGy instead of 0.78 mGy inside the ROI, which is a dose increase of 5%. Outside the ROI the imaging dose would be 0.38 mGy instead of 0.24 mGy with a dose saving of -67% instead of -80%. This simple linear scaling is potentially overestimating the dose since the optimizer may be able to further reduce it.

Nevertheless, the dose saving is still relevant outside of the ROI and a slight dose increase in the ROI is irrelevant as argued below.

Dose savings reported in Dickmann *et al* (2020) were up to 40% outside of the ROI and also matched the peak variance inside the ROI with the uniform fluence scans. In comparison, dose savings achieved in this work are considerably improved, for which we see two reasons: (1) the ROI based on the treatment doses covered a smaller volume compared to the ROI in Dickmann *et al* (2020) and (2) the optimization algorithm in this work can jointly minimize dose and optimize variance, while in Dickmann *et al* (2020) dose was only implicitly minimized by prescribing a certain variance level outside of the ROI, potentially leading to non-optimal results.

In general, imaging dose reductions (and increases) due to FMpCT need to be compared to the local treatment dose: if a change in imaging dose is an order of magnitude lower than the treatment dose, it is probably irrelevant. This holds true for the imaging dose increases of up to 0.6 mGy inside the ROI. Assuming that imaging were to be performed prior to every fraction, this dose increase is negligible compared to the treatment dose per fraction. On the contrary, dose savings outside of the ROI were on average 0.9 mGy. Those, however, need to be compared to the treatment dose outside of the ROI, which was between 180 and 250 mGy per fraction at the border of the ROI (10% by definition) and dropped rapidly further away from the ROI. For most OARs the treatment dose was so low that it was not determined by the TPS. While it is known that treatment planning underestimates neutron dose and thus doses further away from the treatment field, it is likely that those doses will be in the order of magnitude of few milli-Sievert (Schneider and Hälg 2015). Therefore, the imaging dose saving achieved with FMpCT appears to be relevant. Nevertheless, a definitive answer to this question requires a careful study including a precise calculation of the neutron dose during treatment.

This study did not investigate the potential impact of anatomical changes on image variance and dose, and consequently on the FMpCT patterns. Translations and rotations are typically avoided for head and neck patients using face masks. Additionally, internal anatomical changes may occur, as well as weight loss. Future studies should investigate the impact of anatomical changes on the FMpCT scans. Should these be important, mitigation strategies, such as repeated uniform fluence scans or updates of the uniform fluence scan using the FMpCT data, need to be developed.

5. Conclusions

In this work we propose an algorithm with dose and variance objectives for fluence field optimization of proton CT for particle therapy treatment planning. The optimization algorithm aims at minimizing imaging dose while maintaining a certain variance within the treated area. Increased optimization penalties were given to imaging OAR to further reduce imaging dose to susceptible structures. In a Monte Carlo study simulating an existing pCT prototype scanner and realistic pencil beams, we showed that image variance objectives were met and demonstrated that the resulting FMpCT images maintained the dosimetric accuracy compared to uniform fluence pCT images based on the analysis of three pediatric head cases. Passing rates comparing pCT to FMpCT for a 1% criterion on dose and a 1 mm criterion on range were well above 90%. Dose volume statistics showed only minor differences even for organs outside of the FMpCT ROI. We therefore conclude that FMpCT using optimization does not deteriorate image quality for treatment planning dose calculations. At the same time we could demonstrate considerable dose reductions of 74% outside of the ROI on average outperforming a previously used intersection-based modulation. These dose reductions are also relevant compared to the dose per fraction outside of the beam path, which is generally low for particle therapy. Thus, fluence-modulated proton CT with dose and variance objectives can be used to carefully tailor and generally reduce imaging dose with a minor impact on therapeutic dose calculation accuracy of 0.3 mm or less in terms of average range.

Acknowledgments

Declaration of interests

Reinhard W Schulte is member of the advisory board of ProtonVDA LLC.

Ethical statement

The use of patient imaging data in this retrospective study was exempt from requiring ethics approval. Bavarian state law (Bayerisches Krankenhausgesetz/Bavarian Hospital Law §27 Absatz 4 Datenschutz (data protection)) allows the use of patient data for research, provided that any person's related data are kept anonymous. German radiation protection laws request a regular analysis of outcomes in the sense of quality control and assurance, thus in the case of purely retrospective studies no additional ethical approval is needed under German law.

ORCID iDs

J Dickmann ⓘ https://orcid.org/0000-0001-8410-3995
R W Schulte ⓘ https://orcid.org/0000-0002-7892-2756
K Parodi ⓘ https://orcid.org/0000-0001-7779-6690
G Landry ⓘ https://orcid.org/0000-0003-1707-4068

References

Agostinelli S *et al* 2003 Geant4—a simulation toolkit *Nucl. Instrum. Methods Phys. Res.* A **506** 250–303

Bär E, Lalonde A, Zhang R, Jee K W, Yang K, Sharp G, Liu B, Royle G, Bouchard H and Lu H M 2018 Experimental validation of two dual-energy ct methods for proton therapy using heterogeneous tissue samples *Med. Phys.* **45** 48–59

Bartolac S, Graham S, Siewerdsen J and Jaffray D 2011 Fluence field optimization for noise and dose objectives in CT *Med. Phys.* **38** S2–17

Bashkirov V A, Schulte R W, Hurley R F, Johnson R P, Sadrozinski H F-W, Zatserklyaniy A, Plautz T and Giacometti V 2016 Novel scintillation detector design and performance for proton radiography and computed tomography *Med. Phys.* **43** 664–74

Bortfeld T 1999 Optimized planning using physical objectives and constraints *Semin. Radiat. Oncol.* **9** 20–34

Collins-Fekete C-A, Dikaios N, Royle G J and Evans P M 2020 Statistical limitations in proton imaging *Phys. Med. Biol.* **65** 085011

Cormack A M 1963 Representation of a function by its line integrals, with some radiological applications *J. Appl. Phys.* **34** 2722–7

Cubillos-Mesías M, Baumann M, Troost E G C, Lohaus F, Löck S, Richter C and Stützer K 2017 Impact of robust treatment planning on single- and multi-field optimized plans for proton beam therapy of unilateral head and neck target volumes *Radiat. Oncol.* **12** 190

Dedes G, Johnson R P, Pankuch M, Detrich N, Pols W M A, Rit S, Schulte R W, Parodi K and Landry G 2018 Experimental fluence-modulated proton computed tomography by pencil beam scanning *Med. Phys.* **45** 3287–96

Dedes G *et al* 2017 Application of fluence field modulation to proton computed tomography for proton therapy imaging *Phys. Med. Biol.* **62** 6026–43

Dedes G *et al* 2019 Experimental comparison of proton ct and dual energy x-ray ct for relative stopping power estimation in proton therapy *Phys. Med. Biol.* **64** 165002

Dickmann J, Rit S, Pankuch M, Johnson R P, Schulte R W, Parodi K, Dedes G and Landry G 2020 An optimization algorithm for dose reduction with fluence-modulated proton CT *Med. Phys.* **47** 1895–906

Dickmann J *et al* 2019 Prediction of image noise contributions in proton computed tomography and comparison to measurements *Phys. Med. Biol.* **64** 145016

Dickmann J *et al* 2020a Experimental realization of dynamic fluence field optimization for proton computed tomography *Phys. Med. Biol.* **65** 195001

Engelsman M, Schwarz M and Dong L 2013 Physics controversies in proton therapy *Semin. Radiat. Oncol.* **23** 88–96

Ezzell G A *et al* 2009 Imrt commissioning: Multiple institution planning and dosimetry comparisons, a report from aapm task group 119 *Med. Phys.* **36** 5359–73

Giacometti V *et al* 2017 Software platform for simulation of a prototype proton CT scanner *Med. Phys.* **44** 1002–16

Graham S A, Moseley D J, Siewerdsen J H and Jaffray D A 2007 Compensators for dose and scatter management in cone-beam computed tomography *Med. Phys.* **34** 2691–703

Guennebaud G, Jacob B *et al* 2010 *Eigen v3* http://eigen.tuxfamily.org

Hanson K M, Bradbury J N, Cannon T M, Hutson R L, Laubacher D B, Macek R, Paciotti M A and Taylor C A 1977 The Application of Protons to Computed Tomography *IEEE Trans. Nucl. Sci.* **25** 657–60

Hsieh S S and Pelc N J 2014 Control algorithms for dynamic attenuators *Med. Phys.* **41** 061907

Hudobivnik N *et al* 2016 Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy CT images *Med. Phys.* **43** 495–504

Hünemohr N, Krauss B, Tremmel C, Ackermann B, Jäkel O and Greilich S 2014 Experimental verification of ion stopping power prediction from dual energy ct data in tissue surrogates *Phys. Med. Biol.* **59** 83–96

Johnson R P *et al* 2016 A fast experimental scanner for proton CT: technical performance and first experience with phantom scans *IEEE Trans. Nucl. Sci.* **63** 52–60

Kamp F, Carlson D J and Wilkens J J 2017 Rapid implementation of the repair-misrepair-fixation (RMF) model facilitating online adaption of radiosensitivity parameters in ion therapy *Phys. Med. Biol.* **62** N285–96

Landry G and Hua C-H 2018 Current state and future applications of radiological image guidance for particle therapy *Med. Phys.* **45** e1086–95

Markman J, Low D A, Beavis A W and Deasy J O 2002 Beyond bixels: Generalizing the optimization parameters for intensity modulated radiation therapy *Med. Phys.* **29** 2298–304

McCormick M, Liu X, Jomier J, Marion C and Ibanez L 2014 ITK: enabling reproducible research and open science *Frontiers Neuroinform.* **8** 13

Meyer S, Kamp F, Tessonnier T, Mairani A, Belka C, Carlson D J, Gianoli C and Parodi K 2019 Dosimetric accuracy and radiobiological implications of ion computed tomography for proton therapy treatment planning *Phys. Med. Biol.* **64** 125008

Nakajima K *et al* 2017 Clinical outcomes of image-guided proton therapy for stage I non-small cell lung cancer *Int. J. Radiat. Oncol. *Biol. *Phys.* **99** E483–4

Niepel K B *et al* 2020 Animal tissue-based quantitative comparison of dual-energy CT to SPR conversion methods using high-resolution gel dosimetry *Phys. Med. Biol.* (https://doi.org/10.1088/1361-6560/abbd14)

Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99–117

Park J, Park Y, Lee S U, Kim T, Choi Y-K and Kim J-Y 2015 Differential dosimetric benefit of proton beam therapy over intensity modulated radiotherapy for a variety of targets in patients with intracranial germ cell tumors *Radiat. Oncol.* **10** 135

Rädler M, Landry G, Rit S, Schulte R W, Parodi K and Dedes G 2018 Two-dimensional noise reconstruction in proton computed tomography using distance-driven filtered back-projection of simulated projections *Phys. Med. Biol.* **63** 215009

Resch A, Landry G, Kamp F, Cabal G, Belka C, Wilkens J, Parodi K and Dedes G 2017 Quantification of the uncertainties of a biological model and their impact on variable rbe proton treatment plan optimization *Phys. Med.* **36** 91–102

Rit S, Dedes G, Freud N, Sarrut D and Létang J M 2013 Filtered backprojection proton CT reconstruction along most likely paths *Med. Phys.* **40** 031103

Schmid S, Landry G, Thieke C, Verhaegen F, Ganswindt U, Belka C, Parodi K and Dedes G 2015 Monte carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional changes in prostate cancer patients *Phys. Med. Biol.* **60** 9329–47

Schneider U and Hälg R 2015 The impact of neutrons in clinical proton therapy *Frontiers Oncol.* **5** 235

Scholz C, Nill S and Oelfke U 2003 Comparison of imrt optimization based on a pencil beam and a superposition algorithm *Med. Phys.* **30** 1909–13

Schulte R W, Bashkirov V, Klock M C, Li T, Wroe A J, Evseev I, Williams D C and Satogata T 2005 Density resolution of proton computed tomography *Med. Phys.* **32** 1035–46

Schulte R W, Penfold S N, Tafas J T and Schubert K E 2008 A maximum likelihood proton path formalism for application in proton computed tomography *Med. Phys.* **35** 4849–56

Shore R E 2016 Radiation and cataract risk: Impact of recent epidemiologic studies on icrp judgments *Mutation Res./Rev. Mutation Res.* **770** 231–7

Song J H, Kim M J, Park S H, Lee S R, Lee M Y, Lee D S and Suh T S 2015 Gamma analysis dependence on specified low-dose thresholds for vmat qa *J. Appl. Clin. Med. Phys.* **16** 263–72

Taasti V T, Michalak G J, Hansen D C, Deisher A J, Kruse J J, Krauss B, Muren L P, Petersen J B and McCollough C H 2018 Validation of proton stopping power ratio estimation based on dual energy ct using fresh tissue samples *Phys. Med. Biol.* **63** 015012

Weber D C, Schneider R, Goitein G, Koch T, Ares C, Geismar J H, Schertler A, Bolsi A and Hug E B 2012 Spot scanning-based proton therapy for intracranial meningioma: long-term results from the Paul Scherrer institute *Int. J. Radiat. Oncol. ∗Biol. ∗Phys.* **83** 865–71

Wedenberg M, Beltran C, Mairani A and Alber M 2018 Advanced treatment planning *Med. Phys.* **45** e1001–23

Wilkens J J and Oelfke U 2006 Fast multifield optimization of the biological effect in ion therapy *Phys. Med. Biol.* **51** 3127–40

Wohlfahrt P, Möhler C, Stützer K, Greilich S and Richter C 2017 Dual-energy CT based proton range prediction in head and pelvic tumor patients *Radiother. Oncol.* **125** 526–33

Yang M, Zhu X R, Park P C, Titt U, Mohan R, Virshup G, Clayton J E and Dong L 2012 Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration *Phys. Med. Biol.* **57** 4095–115

Zhu C, Byrd R H, Lu P and Nocedal J 1997 Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization *ACM Trans. Math. Softw.* **23** 550–60