IPEM Institute of Physics and Engineering in Medicine

PAPER • OPEN ACCESS

Prediction of image noise contributions in proton computed tomography and comparison to measurements

View the article online for updates and enhancements.

Recent citations

- Experimental comparison of proton CT and dual energy x-ray CT for relative stopping power estimation in proton therapy
George Dedes *et al*

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Prediction of image noise contributions in proton computed tomography and comparison to measurements

J Dickmann[1] , P Wesp[1], M Rädler[1], S Rit[2], M Pankuch[3], R P Johnson[4], V Bashkirov[5] , R W Schulte[5] ,
K Parodi[1], G Landry[1,6,7,8] and G Dedes[1,8,9]

[1] Fakultät für Physik, Department of Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching bei München, Germany
[2] Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Étienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69373, Lyon, France
[3] Northwestern Medicine Chicago Proton Center, Warrenville, IL 60555, United States of America
[4] Department of Physics, U.C. Santa Cruz, Santa Cruz, CA 95064, United States of America
[5] Division of Biomedical Engineering Sciences, Loma Linda University, Loma Linda, CA 92354, United States of America
[6] Department of Radiation Oncology, University Hospital, LMU Munich, 81377 Munich, Germany
[7] German Cancer Consortium (DKTK), 81377 Munich, Germany
[8] Senior authorship is shared equally.
[9] Author to whom any correspondence should be addressed.

E-mail: g.dedes@physik.uni-muenchen.de

Abstract

We present a method to accurately predict image noise in proton computed tomography (pCT) using data generated from a Monte Carlo simulation and a patient or object model that may be generated from a prior x-ray CT image. This enables noise prediction for arbitrary beam fluence settings and, therefore, the application of fluence-modulated pCT (FMpCT), which can achieve prescribed noise targets and may significantly reduce the integral patient dose.

We extended an existing Monte Carlo simulation of a prototype pCT scanner to include effects of quenching in the energy detector scintillators and constructed a beam model from experimental tracking data. Simulated noise predictions were compared to experimental data both in the projection domain and in the reconstructed image. Noise prediction agreement between simulated and experimental data in terms of the root-mean-square (RMS) error was better than 7% for a homogeneous water phantom and a sensitometry phantom with tubular inserts. For an anthropomorphic head phantom, modeling the anatomy of a five-year-old child, the RMS error was better than 9% in three evaluated slices. We were able to reproduce subtle noise features near heterogeneities. To demonstrate the feasibility of Monte Carlo simulated noise maps for fluence modulation, we calculated a fluence profile that yields a homogeneous noise level in the image. Unlike for bow-tie filters in x-ray CT this does not require constant fluence at the detector and the shape of the fluence profile is fundamentally different.

Using an improved Monte Carlo simulation, we demonstrated the feasibility of using simulated data for accurate image noise prediction for pCT. We believe that the agreement with experimental data is sufficient to enable the future optimization of FMpCT fluence plans to achieve prescribed noise targets in a fluence-modulated acquisition.

1. Introduction

The use of proton computed tomography (pCT), initially proposed by Cormack (1963), promises superior accuracy of relative stopping power (RSP) images required for proton therapy dose calculations employed in treatment planning or for various adaptive proton therapy strategies. Instead of using a stoichiometric calibration to convert x-ray linear attenuation coefficients to proton stopping power (Paganetti 2012, Yang *et al* 2012), pCT scanners directly measure RSP line integrals, which can be used to reconstruct RSP images using

dedicated reconstruction algorithms (Li *et al* 2006, Penfold *et al* 2009, 2010, Rit *et al* 2013, Poludniowski *et al* 2014, Hansen *et al* 2016). First prototype proton (and heavier ion) CT scanners (Coutrakon *et al* 2013, Rinaldi *et al* 2013, Johnson *et al* 2016, Sadrozinski *et al* 2016, Meyer *et al* 2017, Pettersen *et al* 2017, Esposito *et al* 2018) obtain an RSP line integral—the water-equivalent path length (WEPL)—by measuring the proton's residual energy behind the patient and converting it to WEPL using a calibration obtained prior to the measurement. Reports on RSP accuracy (Giacometti *et al* 2017a, Esposito *et al* 2018, Volz *et al* 2018) suggest that pCT (and heavier ion CT) could equal or outperform accuracy currently achievable with dual–energy CT (Hudobivnik *et al* 2016, Wohlfahrt *et al* 2017). Dosimetric accuracy was studied in simulations (Arbor *et al* 2015, Meyer *et al* 2019) and suggests good performance of using ion CT images for treatment planning.

Apart from RSP accuracy, pCT benefits from its dose-efficiency (Schulte *et al* 2005). Reported doses for experimental operation of certain designs and for central pixel noise levels comparable to those of x-ray CT are at only 1 mGy (Sadrozinski *et al* 2016). This is comparable or lower to an in-room cone beam CT (Alaei and Spezi 2015) and can pave the way for daily in-room imaging prior to each treatment session to prevent inaccurate dose delivery during treatment due to positioning errors or anatomical changes. An emerging modality to further reduce imaging dose is fluence modulation, which was originally proposed for x-ray CT (Graham *et al* 2007, Bartolac *et al* 2011). Fluence-modulated imaging aims at a dose reduction in parts of the patient by delivering an inhomogeneous imaging dose, and, therefore achieving different noise levels within the image. Fluence modulation has recently gained a strong research interest, particularly due to technical improvements allowing implementation of such systems in x-ray CT (Bartolac and Jaffray 2013, Szczykutowicz and Mistretta 2013a, 2013b, Szczykutowicz *et al* 2015, Stayman *et al* 2016, Gang *et al* 2017, Mao *et al* 2018, Huck *et al* 2019, Shunhavanich *et al* 2019).

Fluence modulation is particularly meaningful for dose recalculation during particle therapy, where good image quality is only required in the proximity of the beam path and imaging dose can be reduced in healthy tissue where an increased noise level is acceptable (Dedes *et al* 2017). This could allow for frequent imaging prior to treatment while maintaining the low integral dose to normal tissue achievable with particle therapy. For imaging with pCT, imaging dose can be accurately delivered in prescribed patterns using the pencil beam scanning functionality of current treatment systems, and the feasibility of fluence-modulated pCT (FMpCT) has been demonstrated both in simulations and proof-of-concept experiments (Dedes *et al* 2017, 2018). Investigations so far focused on static non-optimized fluence maps. To optimize FMpCT plans such that they yield a prescribed image noise map, a prior treatment planning CT may be used as a guide from which pCT noise levels at any fluence setting can be deduced. This requires a method to reconstruct image noise maps from raw pCT data as well as a Monte Carlo simulation that can accurately predict such raw data. A noise reconstruction has been developed in Rädler *et al* (2018) for use with filtered backprojection along curved proton paths (Rit *et al* 2013). It has so far only been investigated for idealized pCT data of a homogeneous water cylinder. Monte Carlo simulations based on GEANT4 have been used to study a pCT scanner in Giacometti *et al* (2017a).
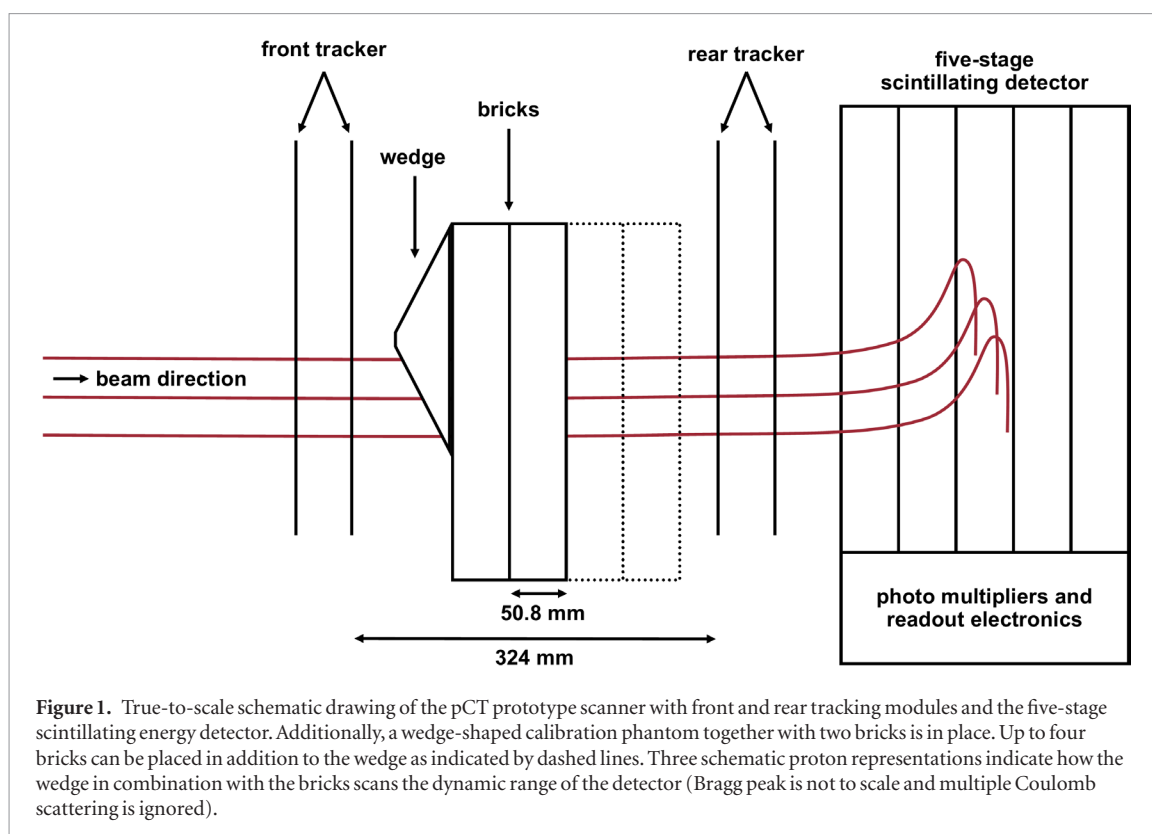
The objective of this study was to demonstrate the feasibility of using a realistic Monte Carlo simulation to accurately predict three-dimensional image noise maps for a given fluence setting. This extends previous studies on pCT noise that either relied on an idealized detector geometry (Rädler *et al* 2018) or on a central pixel noise model (Schulte *et al* 2005). Data for this study were acquired at a specific pCT prototype scanner described in Sadrozinski *et al* (2016). We compared simulated noise predictions to experimental data for three different phantoms, including an anthropomorphic head phantom. We aimed at explaining the shape of resulting image noise maps by disentangling noise contributions in the simulation. Therefore, we hypothesized noise contributions due to multiple Coulomb scattering, energy straggling within the object, uncertainty of the tracking information, energy detection process, and beam energy spread. Finally, we exploited simulated noise maps to apply a fluence profile that yields homogeneous image noise and thus has the same effect as a bow-tie filter for x-ray CT.

2. Materials and methods

2.1. Experimental setup

Experimental data for this study were acquired using the phase II preclinical pCT prototype scanner developed at Loma Linda University and the University of California, Santa Cruz with details published in Johnson *et al* (2016). The single proton tracking scanner consists of two tracking modules, one prior to and one after the imaging object, and a five-stage scintillating detector, which is described in Bashkirov *et al* (2016). The scintillators are built from the polystyrene-based material UPS-923A (Artikov *et al* 2005) with an RSP of about 1.038, according to Bashkirov *et al* (2016). For each proton, two positions before and two after the object, as well as five energy deposits are recorded. For each pair of position informations a direction vector can be calculated. A schematic drawing of the scanner geometry is shown in figure 1.

Experimental data were acquired at the Northwestern Medicine Chicago Proton Center by wobbling a narrow beam (FWHM $\approx$ 40 mm) of 200 MeV protons over a FWHM area of 80 mm $\times$ 200 mm for phantom runs and 80 mm $\times$ 300 mm for calibration runs, respectively. During phantom runs, the object was rotated con-

Figure 1. True-to-scale schematic drawing of the pCT prototype scanner with front and rear tracking modules and the five-stage scintillating energy detector. Additionally, a wedge-shaped calibration phantom together with two bricks is in place. Up to four bricks can be placed in addition to the wedge as indicated by dashed lines. Three schematic proton representations indicate how the wedge in combination with the bricks scans the dynamic range of the detector (Bragg peak is not to scale and multiple Coulomb scattering is ignored).

tinuously at a speed of 1 rpm to acquire tomographic data for 6 min at a rate of about 10^6 registered protons per second.

2.2. Calibration and reconstruction

For each proton, the energy deposit above 1 MeV to the furthest stage (referred to as stopping stage) was mapped to a WEPL using a two-step calibration procedure. In a first step, each stage's channel numbers of the analog-to-digital converter (ADC numbers) obtained in a degrader-free acquisition were mapped to pre-calculated energy values to compensate spatial non-uniformities of the detector response and to achieve an absolute energy measurement. These five energy values E_n^{G4} were obtained in a previous study by Bashkirov *et al* (2016). In a second step, data of a wedge-shaped calibration phantom (described in the next section) were used to map the stopping stage's energy deposit to a known WEPL. During subsequent acquisitions, this lookup-table was used to calculate unknown WEPLs from measured energy deposits to the stopping stage.

With the spatial information of the tracking modules, a most likely path as described in Schulte *et al* (2008) was estimated that then was used to reconstruct tomographic images by binning the WEPL data to virtual projections at every depth of interaction. This process of distance-driven binning is detailed in Rit *et al* (2013) and aims at improving spatial resolution in pCT images reconstructed using filtered backprojection. Prior to reconstruction, we performed a rejection of proton histories, referred to in the field as cuts. Protons were rejected if their WEPL or direction information was outside of 3σ boundaries around the median value for bins defined by a 2D grid based on the front tracker position information. This is a standard procedure for pCT (Schulte *et al* 2008, Rit *et al* 2013).

For reconstruction distance-driven binning was performed at a grid of 180×50 bins laterally and at 180 depths longitudinally with a uniform voxel size of 1 mm in all dimensions. Bins for the calculation of cuts were calculated on a uniform grid of 2 mm by 2 mm. Both RSP and noise maps were reconstructed to a volume of $180 \times 180 \times 50$ voxels with a uniform voxel size of 1 mm in all dimensions. Both the projection grid and the reconstruction grid were centered at the isocenter.

2.3. Phantoms

The following phantoms were used in this study. Wherever RSP values for physical phantoms are stated with uncertainties, these were experimentally determined using measurements with a variable water column.

To calibrate the setup, a wedge-shaped calibration phantom made from polystyrene (RSP $= 1.030 \pm 0.003$, Piersimoni *et al* (2017)) was used. The flattened peak of the wedge faces the front tracker as shown in figure 1. It has a maximum longitudinal thickness of 50.8 mm while the lateral width is 209 mm. Between zero and four

polystyrene bricks with a thickness of 50.8 mm each were placed behind the wedge to cover the whole dynamic range of the detector up to 254 mm.

For tomographic acquisitions, a water phantom, the sensitometric CTP404 phantom (*Phantom Laboratory*, New York, USA) as well as a pediatric head dosimetry phantom (ATOM®, Model 715 HN, *CIRS Inc.*, Norfolk, USA) were used. The water phantom consists of a water-filled PMMA cylinder (RSP ≈ 1.17) with an outer diameter of 150.5 mm and a wall thickness of 6.35 mm. The CTP404 phantom's cylindrical body made from epoxy (RSP = 1.144 ± 0.001, Giacometti *et al* (2017a)) has a diameter of 150 mm and multiple cylindrical inserts with RSPs ranging from air (RSP < 0.01) to Teflon (RSP = 1.790 ± 0.002, Giacometti *et al* (2017a)). The pediatric head phantom is a realistic anatomical model of a 5-year-old child built from tissue-equivalent materials and was used in previous pCT studies (Giacometti *et al* 2017b). Since the height of the head phantom was larger than the height of the detector aperture, it needed to be scanned in two consecutive acquisitions with an overlap of several millimeters.

2.4. Simulation platform

To simulate acquisitions, a dedicated Monte Carlo simulation platform was used that models the complete geometry of the detector. The platform was described and validated for its RSP fidelity in Giacometti *et al* (2017a). It is based on the GEANT4 framework, version 10.2.p01, as presented in Agostinelli *et al* (2003). The reference physics list `QGSP_BIP_HP` was used for the simulation of the interaction of particles with matter. For a highly accurate description of electromagnetic interactions, the `G4EmLivermorePhysics` model was used for electrons and photons. The cut for secondary particle production in the energy detector was defined at 80 μm while for the rest of the scanner at 1 mm.

In this work, we extended the platform to model non-linear effects of light production in the scintillator (section 2.6) and to incorporate a realistic beam model (section 2.7). Moreover, we aimed at quantifying the accuracy of image noise (sections 2.8 and 2.9). All phantoms mentioned in the previous section were simulated as analytical phantoms based on their known geometry and materials, except for the pediatric head phantom, for which a high-resolution voxelized implementation based on x-ray CT scans as presented in Giacometti *et al* (2017b) existed. For all phantoms, RSP accuracy was ensured by fine-tuning the I-value in the simulation such that agreement to reference values was better than 0.1% for protons with an energy of 150 MeV. The I-value of water was 78.0 eV, according to the latest ICRU recommendation.

2.5. Verification of noise reconstruction

In order to calculate image noise maps, we used a noise reconstruction method developed for use with distance-driven binning by Rädler *et al* (2018). For every bin p with a given mean WEPL in the distance-driven projection, the expected standard deviation of the mean for n protons intersecting that bin was calculated as

$$\sigma_p = \sigma_{\text{WEPL}}/\sqrt{n}, \tag{1}$$

where σ_{WEPL} is the per-proton standard deviation of the underlying set of WEPLs and σ_p is the expected standard deviation of the mean of n WEPLs. This yielded a three-dimensional distance-driven noise projection, which was reconstructed after convolving it with a special noise filter, consisting of the square of the reconstruction filter (e.g. the filter of Ramachandran and Lakshminarayanan (1971) in Rädler *et al* (2018)). As described in Rädler *et al* (2018), effects of projection value interpolation (as part of filtered backprojection) on the image noise can be calculated per voxel or as an effective mean over the entire volume. In this study, the second option was chosen as it is computationally more efficient and the high-frequency voxel-wise components are not relevant for a future application using fluence modulation, since they are smaller than any clinical proton pencil beams.

In the publication of Rädler *et al* (2018), the method was only applied to simulated ideal pCT data with a parallel beam. Accuracy was only evaluated using annular ROIs, and not on a voxel-by-voxel basis, which is standard (see for example Wunderlich and Noo (2008)). Therefore, in a first noise reconstruction validation study, we simulated $N = 40$ independent noise realizations for a tomographic scan of the water phantom. Each of the noise realizations was reconstructed independently, which allowed us to calculate a voxel-wise standard deviation over N samples, which is referred to as batch method. This served as the ground truth to be compared to the distance-driven noise reconstruction of a single noise realization.

2.6. Non-linearities in the detection process

When irradiated by high-LET particles, such as protons of low energy, the light production in the scintillators becomes a non-linear function of the deposited energy due to the effect of quenching as investigated by Birks (1951). To allow for a quantitative prediction of the measurement uncertainty, the simulation needed to account for this non-linearity. In the case of the five-stage energy detector and using the theory of Birks (1951), a proton stopping in stage n and with a residual range R_n within that stage will produce the distorted energy measurement

$$E'_n(R_n) = S_n \cdot \int_{R_n}^0 dx \, \frac{dE/dx}{1 + k_B \cdot dE/dx}, \tag{2}$$

where k_B is the empirical Birks' factor and S_n is a scaling factor specific for each stage, which is fixed during the calibration. In appendix A in (A.5), we calculate the scaling factor taking into account the specific calibration procedure. We find a formulation for $E'_n(\mathrm{WEPL})$ as a function of the WEPL instead of the range. This formulation depends only on the Birks' factor k_B as well as the thickness of each scintillating stage l_{stage} and the residual range in the detector material at the entrance of the detector R_0. The required material specific stopping power dE/dx was calculated as a function of the residual range R for the correct material composition using the GEANT4 platform.

At the same time, we were able to obtain energy measurements $E_n^*(\mathrm{WEPL})$ from experimental data using the wedge phantom, which covers the whole dynamic range $\mathrm{WEPL} \in [0, 254]\,\mathrm{mm}$ of the detector. For each proton the energy deposit E_n^* was calculated as well as the WEPL based on the tracking information and the known geometry. For the whole dataset, this yielded a 2D-histogram of E_n^* against WEPL, from which the relationship $E_n^*(\mathrm{WEPL})$ was deduced by finding the most frequent energy deposit for each WEPL.

This allowed us to find the estimation of the three unknowns k_B, R_0 and l_{stage} by minimizing the cost function

$$C = \sum_n \left[\sum_{\mathrm{WEPL}} \left(E'_n(\mathrm{WEPL}) - E_n^*(\mathrm{WEPL}) \right)^2 \right]^2 \tag{3}$$

using the quasi-Newton method of Broyden (1970). The Birks' factor k_B was then used to modify each incremental energy deposit in the simulation as

$$\frac{d\,\mathrm{ADC}}{dx} \propto \frac{dE'}{dx} \propto \frac{dE/dx}{1 + k_B \cdot dE/dx}, \tag{4}$$

where $d\,\mathrm{ADC}$ is the increment of the simulated ADC number. The scaling factor S_n was neglected here due to the arbitrary scaling of the ADC number.

2.7. Realistic beam model

To realistically model the initial positions and directions of protons in the simulation, we exploited the fact that such measurements are directly available from experimental data. Each proton's position was projected to a point at 400 mm in front of the isocenter, and, therefore, 232.8 mm in front of the first tracking module, along a straight line according to its direction vector. To avoid interplay-effects between the proton placement and the tracking strip, the position information was blurred with a random number normal-distributed around zero and with a standard deviation equal to the distance between two tracking strips.

The experimental dataset, on which this study was based, used different field widths for the calibration and the subsequent phantom acquisitions. Modeling noise of phantom acquisitions correctly required to use the smaller field width of these measurements. Unfortunately, degrader-free runs were only available for the calibration runs with the larger field width. However, acquisitions with a the water phantom in the beam path were used to create a beam model. We therefore accounted for the additional attenuation due to the phantom by randomly selecting protons with a probability anti-proportional to the transmittance.

The initial energy of each proton could not be measured directly, because in the degrader-free run the calibration forces it to be $E_0 = \sum_n E_n^{\mathrm{G4}} + \Delta E_{\mathrm{tracker}} = 200\,\mathrm{MeV}$, where $\Delta E_{\mathrm{tracker}} = 3.65\,\mathrm{MeV}$ is the assumed mean energy loss in the tracking modules and air. It was, however, possible to determine the beam energy spread σ_{beam}, which is likely to have a strong impact on the resulting image noise. Assuming that electronic readout noise is negligible compared to σ_{beam}, we performed degrader-free simulations without attenuation correction and for values of $\sigma_{\mathrm{beam}} \in \{0.5, 0.6, \ldots, 1.0\}\,\mathrm{MeV}$. We calculated the resulting spread σ_{E_5} of the energy deposit in the last stage using Gaussian fits. Since the simulations considered the optimal Birks' factor as described in the last section, it was possible to compare this σ_{E_5} to the one obtained from a measurement and choose σ_{beam} such that σ_{E_5} matches the experimental value.

2.8. Comparison to experimental data in the projection domain and noise contributions

To ascertain that the modeling in sections 2.6 and 2.7 resulted in accurate noise predictions, we simulated a projection of the wedge-shaped calibration phantom with zero to three bricks and compared them to experimental data in terms of standard deviation. Due to the larger extent of the phantom compared to reconstructions, projections were binned to 250×50 bins laterally and at 250 depths longitudinally. The voxel size and all other parameters are described in section 2.2.

After applying the WEPL calibration and cuts, two sets of distance-binned projections were generated from the data. The first was the average calibrated WEPL. The second was the per-proton WEPL standard deviation $\sigma_{\mathrm{WEPL}} = \sigma_p \cdot \sqrt{\bar{n}}$, where σ_p is the standard deviation according to (1) and $\bar{n}$ is the average fluence in the

Table 1. Calculation of the five noise contributions as a difference of the five scoring techniques. Numbers in brackets refer to the listing of scoring techniques in section 2.8.

Noise contribution	Difference of scoring techniques
Scattering	(i)
Energy straggling (ES) in the object	(ii)–(i)
Tracking	(iii)–(ii)
Energy detection process	(iv)–(iii)
Beam energy spread	(v)–(iv)

projection. The normalization to a fluence of one proton per bin makes results comparable to, e.g. Bashkirov *et al* (2016). After rejecting data outside the hull of the wedge, this resulted in a list of WEPL-noise-pairs, with every pixel in the projection giving one entry in the list. For each binning depth, these pairs were binned by WEPL to multiples of 1.5 mm and for each bin the median noise value was calculated.

We hypothesize that noise in the projection can be attributed to five different contributions as listed in table 1. In order to study these contributions to image noise, noise calculations were performed for simulations using the following scoring methods, which disentangle each of these effects.

(i) *WEPL scoring.* Dynamic scoring in GEANT4 of each proton's exact WEPL by multiplying the stepping length with the current material's RSP calculated for the current proton energy. Exact coordinates when crossing the tracking planes were recorded. Noise is caused only by scattering of protons with different paths into the same distance-driven bin.

(ii) *Energy scoring.* Scoring of the proton's energy at tracking planes before and after the phantom and conversion of the energy loss to WEPL. Exact coordinates when crossing the tracking planes were recorded. Noise is caused by multiple Coulomb scattering as well as energy straggling within the object.

(iii) *Energy scoring (realistic position).* Energy scoring, but using the position information estimated by the tracking modules. Noise is increased due to the additional uncertainty of the tracking information.

(iv) *Realistic scoring ($\sigma_E = 0$ MeV).* Fully realistic simulation, but setting the beam energy spread to 0 MeV. Noise is increased compared to *energy scoring* due to energy straggling in the detector and the calibration process, but it is lower compared to *realistic scoring*, which includes the beam energy spread, and its impact can be quantified.

(v) *Realistic scoring.* Fully realistic simulation, including the described beam model with the determined beam energy spread. This model contains all known noise contributions.

To quantify the relative noise contribution of scattering, energy straggling in the object, tracking, the energy detection process and the beam energy spread, we calculated the difference in projection noise values for the five scoring techniques as indicated in table 1. Since the cuts of proton histories depend on the underlying standard deviation, the efficiency of data filtering may vary across datasets and impair this analysis as a different set of protons was used to calculate each dataset. However, considerable differences of the cuts efficiency were only observed at the edge of the phantom and only for the *WEPL scoring* technique, where the analyses including this scoring technique may not be quantitative. To correctly assess the relative noise contributions, this part investigates variance values, which are the square of the standard deviation values we calculated so far. This is because variance contributions add up linearly, while standard deviation contributions need to be added quadratically. The variance difference between the scoring techniques was calculated for each WEPL bin and normalized to the variance using *realistic scoring*.

The previous analyses were performed on very homogeneous data of the continuous part from the wedge-shaped calibration phantom. To demonstrate the impact of heterogeneities, we calculated standard deviation profiles for the steep edge of the calibration phantom. The evaluation of this dataset is shown in appendix B.

2.9. Comparison to experimental data with heterogeneous and anthropomorphic phantoms

To investigate the performance of the Monte Carlo simulation for predicting image noise maps, we simulated pCT data for the water phantom, the CTP404 phantom as well as the pediatric head phantom (see section 2.3) using the *realistic scoring* technique. For the same phantoms, experimental data were acquired using the prototype pCT scanner. We reconstructed both RSP and noise images. All image noise maps were normalized to an average projection fluence of $f_0 = 20$ mm^{-2}. For the water phantom and the CTP404 phantom, which are symmetric in the z-direction, 16 slices were averaged after noise reconstruction.

In a first step, we manually selected two corresponding slices and registered the 2D-RSP image of the simulation onto the experimental reconstruction using a rigid registration allowing translation and rotation.

This registration was then applied to the corresponding slices of the image noise maps. In the following evaluation, only pixels inside the object's outer hull as determined by an RSP threshold of 0.15 were considered.

The registration allowed a calculation of the pixel-wise relative error map $\Delta\sigma$ in terms of an estimation of the standard deviation as

$$\Delta\sigma = \frac{\sigma_{\mathrm{RSP,sim}} - \sigma_{\mathrm{RSP,exp}}}{\sigma_{\mathrm{RSP,exp}}}, \tag{5}$$

where σ_{RSP} is the registered data of the image noise reconstruction for experimental and simulated data, respectively.

Using the *WEPL scoring* technique (see section 2.8), we reconstructed images showing the simulated scatter-only standard deviation $\sigma_{\mathrm{scatter}}$. The non-scatter standard deviation $\sigma_{\mathrm{non\text{-}scatter}}$, i.e. the expected standard deviation in absence of scatter, was calculated as

$$\sigma_{\mathrm{non\text{-}scatter}} = \sqrt{\sigma_{\mathrm{RSP}}^2 - \sigma_{\mathrm{scatter}}^2}, \tag{6}$$

where σ_{RSP} is the total standard deviation. Again, a varying cuts efficiency may impair this analysis in particular at the edges of objects.

Using standard GEANT4 functionality of the simulation platform, the imaging dose was scored for every projection angle and then summed. Mean imaging dose values for individual slices were calculated within the phantom.

As a sanity check, we calculated RSP and standard deviation histograms of the whole volume of the head phantom. Results of this verification are presented in appendix C.

2.10. Application: a bow-tie filter for proton CT

To demonstrate the feasibility of using Monte Carlo simulated noise data for fluence modulation, we calculated a fluence modulation profile for the homogeneous water phantom which has the same effect as a bow-tie filter in x-ray CT. The simplest concept of an x-ray bow-tie filter aims at homogeneous noise at the detector as described in Harpen (1999) or Graham *et al* (2007). While in x-ray CT this can be achieved by aiming at homogeneous fluence at the detector, this does not hold true for pCT because of the impact of multiple Coulomb scattering as discussed in Rädler *et al* (2018) for an idealized scanner. Instead, a modulated fluence is required that yields a constant noise level at the detector, which in turn will make noise flat in the image, as can be derived from the variance reconstruction formula presented in Rädler *et al* (2018).

The artificial fluence modulation profile was calculated solely based on simulated data. It was then applied to experimental data by randomly selecting protons with an acceptance probability $p(u)$ as a function of the lateral coordinate u (no modulation was needed orthogonal to that due to the symmetric phantom). Given a simulated image noise profile $\sigma_{p,\mathrm{sim}}(u)$ in the projection binned at the isocenter, we calculated a fluence modulation

$$p(u) = \min\left(\frac{\sigma_{p,\mathrm{sim}}^2(u)}{\sigma_0^2}, 1\right) \tag{7}$$

that forces the standard deviation in the projection to the value σ_0. Note that the fluence modulation is proportional to variance and, therefore, to the square of the standard deviation. In the given case, it was only possible to reduce fluence, and, therefore, $\sigma_0 \geqslant \max_u \sigma_{p,\mathrm{sim}}(u)$. However, this condition could also be violated at the cost of a resulting non-homogeneous fluence where noise is highest in the original acquisition. A violation is required at the object's hull, where noise tends to be elevated. In these regions, an unreasonably high skin dose would result from forcing the noise level to be the same as inside the object. In this study, we prescribed a projection standard deviation of $\sigma_0 = 5.48$ mm, and, therefore, a variance of $\sigma_0^2 = 30$ mm².

3. Results

3.1. Verification of noise reconstruction

Figure 2 shows the standard deviation ground truth calculated with $N = 40$ noise realizations as well as the noise reconstruction for the water phantom simulation. Both image noise maps agreed well. The residual difference was caused by the approximation of the effect of projection value interpolation in our noise reconstruction and displayed as a star-shaped increase in noise that was centered at the center of rotation and that spanned the whole reconstruction volume. In the relative difference map shown in (c) the star-shaped pattern reoccurred as an underrepresentation of the noise reconstruction. The agreement between the ground truth and the noise reconstruction was good with the mean error over the whole phantom being -2.5% and the root-mean-square error 4.1%. Also the profile plots in (d) revealed that there was a slight underestimation of noise from the noise reconstruction.
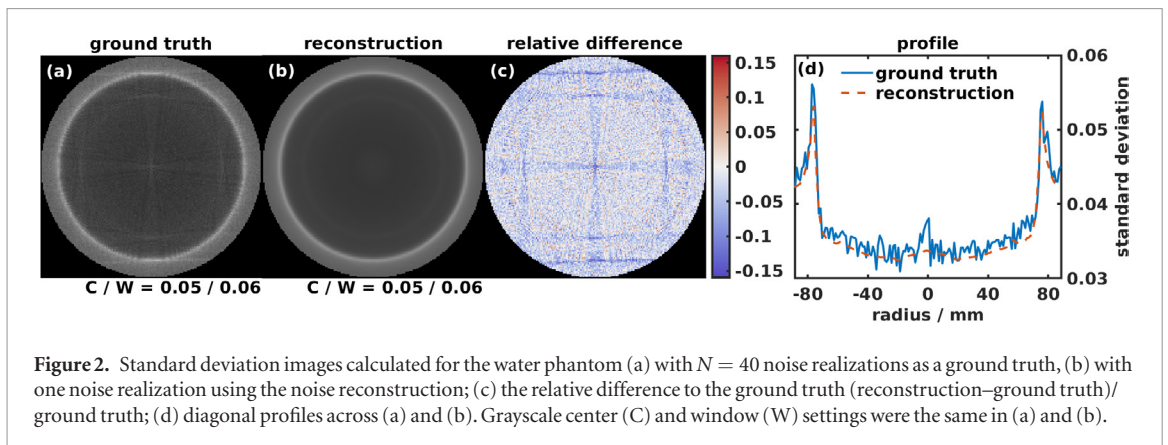
Figure 2. Standard deviation images calculated for the water phantom (a) with $N = 40$ noise realizations as a ground truth, (b) with one noise realization using the noise reconstruction; (c) the relative difference to the ground truth (reconstruction–ground truth)/ ground truth; (d) diagonal profiles across (a) and (b). Grayscale center (C) and window (W) settings were the same in (a) and (b).

3.2. Non-linearities in the detection process

Figure 3(a) shows a two-dimensional histogram of the relative proton counts for WEPL versus energy deposit in the stopping stage. Data were obtained from an experimental acquisition of the wedge-shaped calibration phantom, for which the WEPL was geometrically calculated for every proton. Each of the five lines in the histogram corresponds to one of the five stopping stages. For example, the line at the lowest WEPL values corresponds to the fifth stage as protons with a low WEPL can penetrate all other stages before stopping in the last one. With increasing WEPL, the energy deposit decreases until suddenly the stopping stage changes and the energy deposit is high again.

For each stage, there was some WEPL range at multiples of the WEPL of one brick $WEPL_{brick} = l \cdot RSP = 52.3$ mm for which counts decreased and which is visible in the histogram as a vertical line of lower intensity. At this point, data need to be merged between protons going through n bricks and the center of the wedge, and protons going through $n + 1$ bricks, but missing the wedge (see figure 1). These flat regions of the calibration phantom and in particular divergent protons leaving the lateral edge of bricks cause the observed change in statistics. However, since each WEPL bin was evaluated individually to obtain the relation-ship E_n^*(WEPL), this did not affect the evaluation.

Table 2 reports the three free parameters, which were optimized by minimizing the cost-function in (3). Note that these parameters were found solely from experimental data.

In figure 3(b), the same histogram is shown for a simulated dataset and for $k_B = 0$ mm MeV^{-1} and, there-fore, neglecting quenching effects. For a given WEPL, energy deposits were considerably higher compared to the experimental data. Figure 3(c) shows simulated data using the optimized Birks' factor. The energy deposits decreased and agreed with experimental data shown in figure 3(a). For experimental data, there was a cut-off related to the measurement process below an energy deposit of around 20 MeV for the first stage (WEPL around 250 mm), which was not present in the simulated data.

Figure 4 shows calibration curves that were used to calculate a WEPL given the energy deposit to the stopping stage for each individual proton. These curves were obtained for each stage during the calibration process. Three calibrations were calculated based on data shown in figures 3(a)–(c). As expected from the raw data, the calibra-tion curve of the simulation without quenching disagreed with the one of the experimental data and the simulation using the optimized Birks' factor.

3.3. Realistic beam model

Figure 5 shows the resulting spread of the last stage's calibrated energy deposit σ_{E_5} as a function of σ_{beam} together with a quadratic fit to this relationship. In experimental data, where σ_{beam} cannot be measured directly, we determined the energy spread of the last stage to be $\sigma_{E_5} = (3.47 \pm 0.02)$ MeV. According to the quadratic fit, this corresponded to a beam energy spread of $\sigma_{beam} = (0.66 \pm 0.02)$ MeV. The uncertainty of σ_{E_5} is given by a one standard deviation confidence interval of the fit parameter and the uncertainty of σ_{beam} was calculated as an error propagation thereof through the quadratic fitting function. Both uncertainties are visualized in figure 5 as shaded areas.

3.4. Comparison to experimental data in the projection domain and noise contributions

Figures 6(a)–(c) shows the WEPL standard deviation as a function of the WEPL for projection data of the wedge-shaped calibration phantom. Data are shown at three different binning depths: at the front tracker position (a), at the isocenter (b), and at the rear tracker position (c). While the first five color-coded curves show the noise level for simulations with different scoring techniques at increasing complexity, the last curve was obtained from experimental data. See section 2.8 for the definition of terms printed in italic type in the following paragraphs.
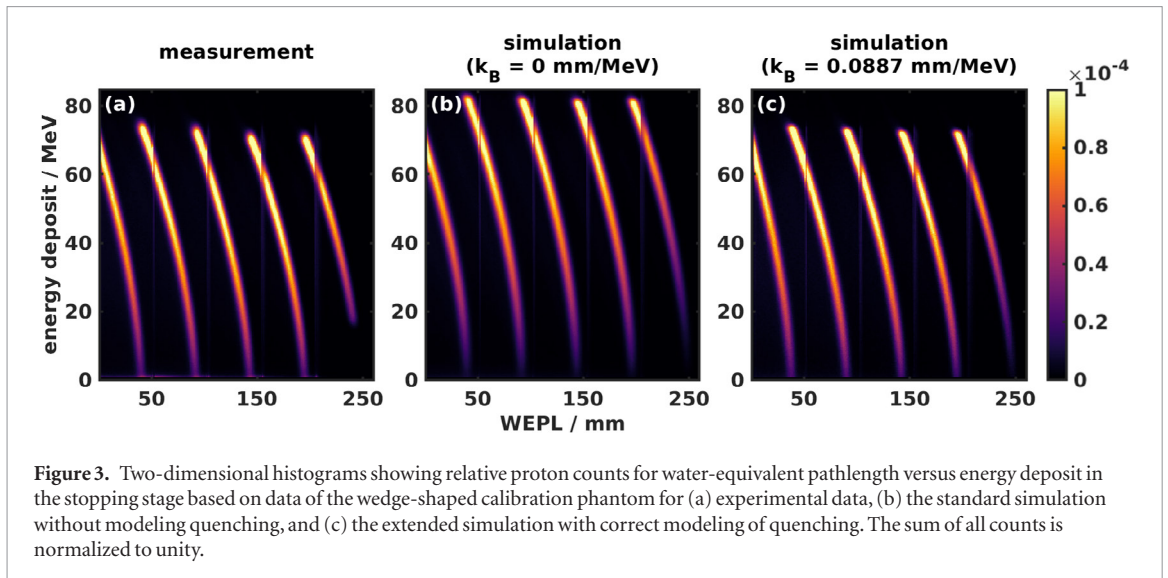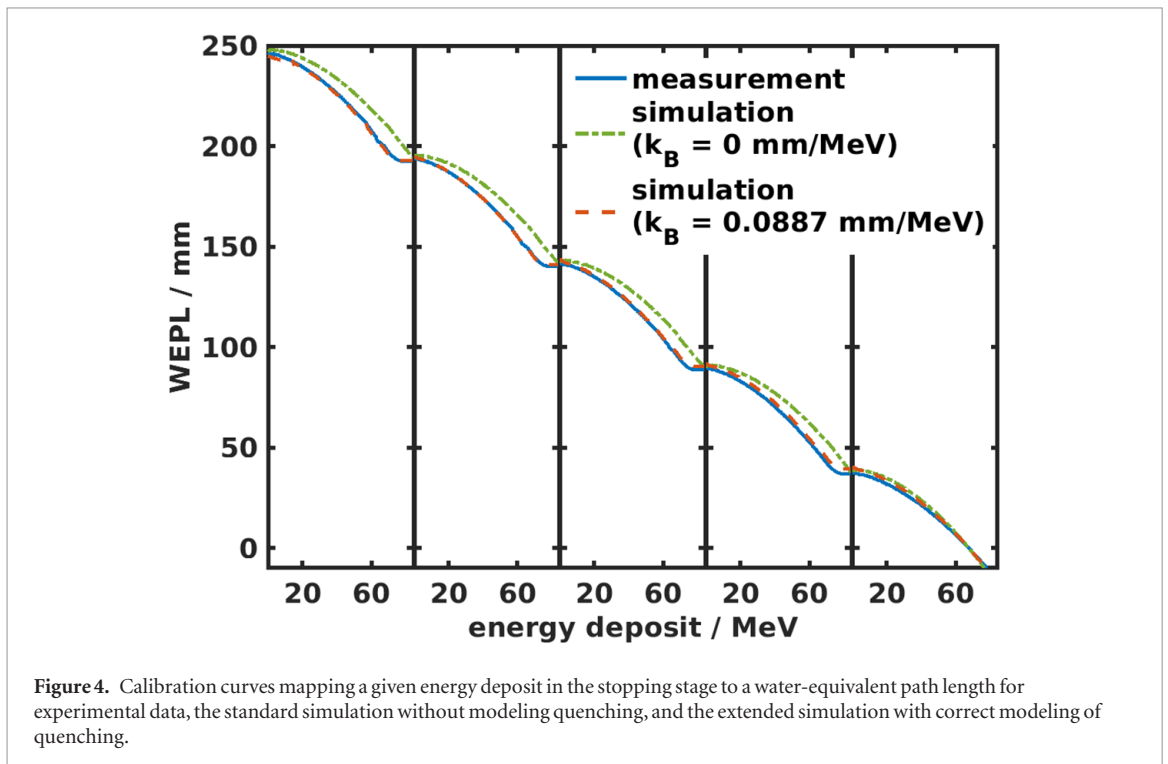
Figure 3. Two-dimensional histograms showing relative proton counts for water-equivalent pathlength versus energy deposit in the stopping stage based on data of the wedge-shaped calibration phantom for (a) experimental data, (b) the standard simulation without modeling quenching, and (c) the extended simulation with correct modeling of quenching. The sum of all counts is normalized to unity.

Table 2. Optimization result of the three free parameters to model quenching effects in the detector.

Parameter	Fit estimate
Birks' coefficient k_B	0.0887 mm MeV^{-1}
Residual range at entrance R_0	237.9 mm
Stage thickness l_{stage}	49.7 mm

Figure 4. Calibration curves mapping a given energy deposit in the stopping stage to a water-equivalent path length for experimental data, the standard simulation without modeling quenching, and the extended simulation with correct modeling of quenching.

The lowest noise level was recorded for *WEPL scoring*. While for binning at the front tracker standard deviation was below 0.2 mm, it increased to values between 1 mm and 2 mm for binning at the rear tracker with a general increase of noise with the WEPL. Additionally, at multiples of the brick thickness (indicated by the dashed lines in figure 6), an abrupt decrease of standard deviation of about 0.25 mm was observed for binning at the isocenter and at the rear tracker, which will be explained in section 4.4.

For *energy scoring*, noise increased with increasing WEPL and pronounced effects at multiples of the brick thickness were only observed for rear tracker binning. Use of *energy scoring (realistic position)* had only a minor effect on the standard deviation compared to *energy scoring*.

For *realistic scoring*, using the same calibration method as in a measurement, the dependency of the standard deviation on WEPL became less pronounced. Table 3 presents mean WEPL standard deviation over the whole
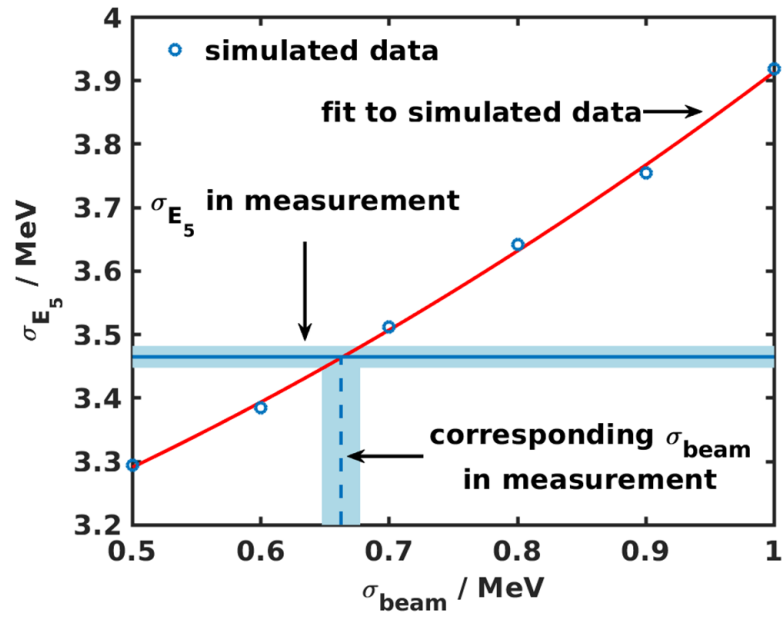
Figure 5. Relationship between the beam energy spread σ_{beam} and the spread of the energy deposit of the last stage σ_{E_5} for a degrader-free measurement. Circles indicate simulated data to which a quadratic model was fitted. Shaded areas indicate a one standard deviation confidence interval. The intersection of this fit with the σ_{E_5} of a measurement was chosen to be the correct σ_{beam} to be used in simulations.
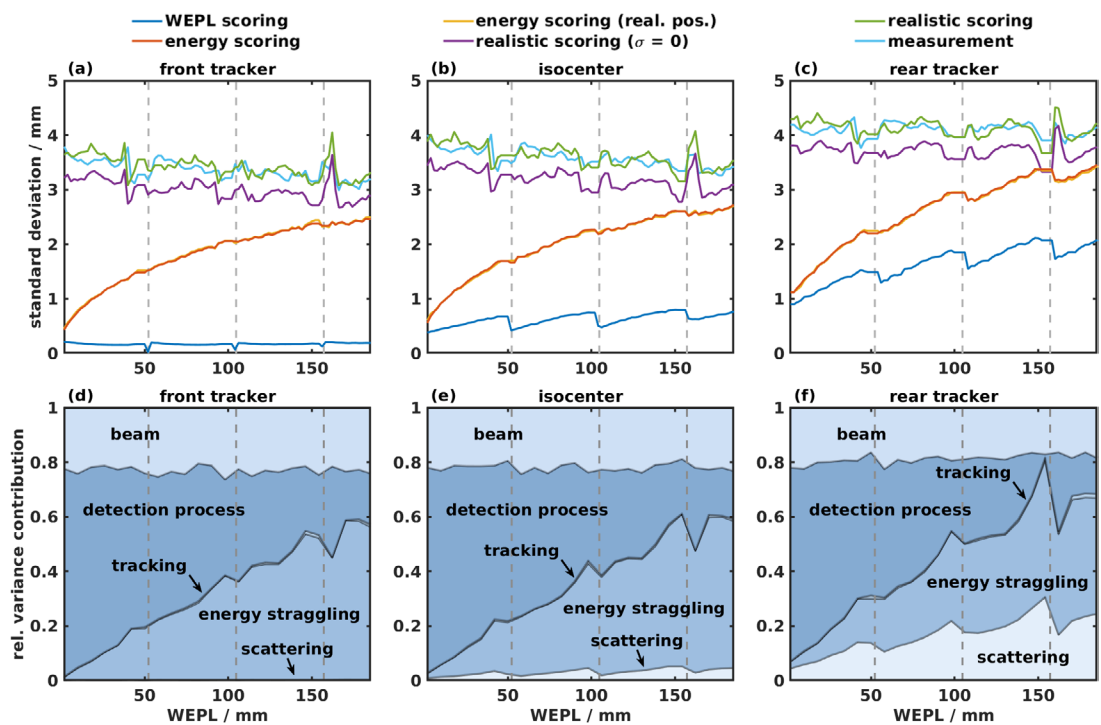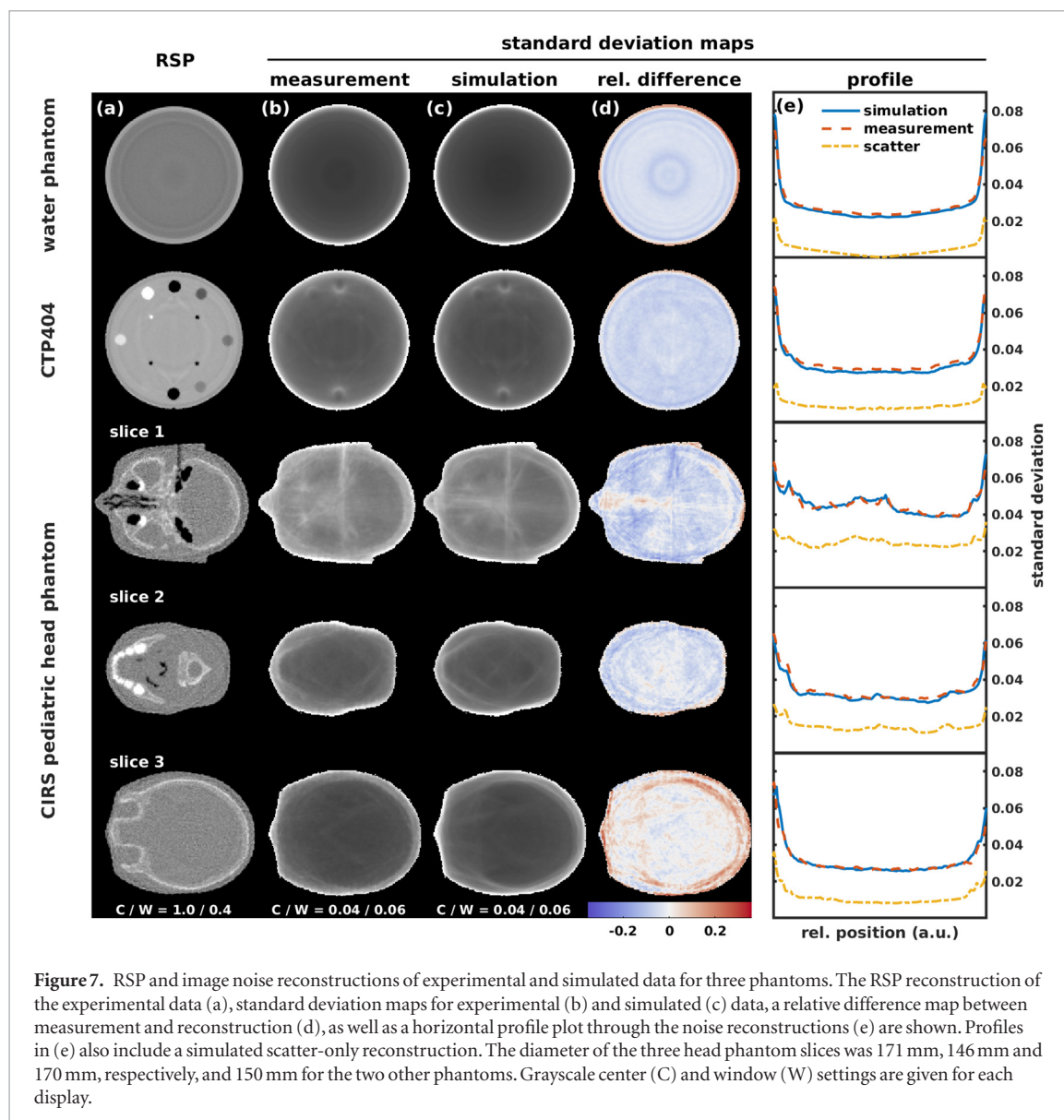

Figure 6. (a)–(c) Standard deviation level in the projection domain for different scoring techniques used in a simulation and comparison to the experimentally determined noise level; (d)–(e) relative variance contributions attributed to physical effects as calculated by subtraction of data in (a). All data are shown at three different depths of the distance-driven binning: at the front tracker for (a) and (d), the isocenter for (b) and (e) and the rear tracker position for (c) and (f). Dashed lines indicate multiples of the brick thickness. Note that the legend in (a)–(c) spans across all three plots.

WEPL range. There is a considerable increase in standard deviation due to the beam energy spread. The curve for the fully realistic simulation, as well as its mean value, agreed within their uncertainty with the experimental data.

In figures 6(d)–(f), the relative contributions to WEPL variance are shown, and can be attributed to the following effects (see section 2.8): scattering, energy straggling in the object, tracking, the energy detection process and the beam energy spread. While scattering was negligible at the front tracker (d), the accumulated contribution at the rear tracker (f) reached about 20%. The contribution of the tracking uncertainty was negligible over

Table 3. Mean WEPL standard deviation over the 0 mm–184 mm WEPL range for two different scoring techniques as well as for experimental data binned at three different depths.

Scoring	Mean standard deviation/mm		
	Front tracker	Isocenter	Rear tracker
Realistic scoring ($\sigma_E = 0$)	2.99 ± 0.18	3.19 ± 0.19	3.71 ± 0.14
Realistic scoring	3.41 ± 0.20	3.61 ± 0.21	4.12 ± 0.16
Measurement	3.35 ± 0.19	3.57 ± 0.18	4.09 ± 0.11

Figure 7. RSP and image noise reconstructions of experimental and simulated data for three phantoms. The RSP reconstruction of the experimental data (a), standard deviation maps for experimental (b) and simulated (c) data, a relative difference map between measurement and reconstruction (d), as well as a horizontal profile plot through the noise reconstructions (e) are shown. Profiles in (e) also include a simulated scatter-only reconstruction. The diameter of the three head phantom slices was 171 mm, 146 mm and 170 mm, respectively, and 150 mm for the two other phantoms. Grayscale center (C) and window (W) settings are given for each display.

the whole WEPL range. The main source of noise in the projection of the wedge phantom was the detection process as well as energy straggling in the object. Both contributions had an approximately constant sum, while the contribution of energy straggling in the object increased proportionally to the WEPL. The variance added by the beam energy spread had a constant value of about 20% independent of the WEPL.

3.5. Comparison to experimental data with heterogeneous and anthropomorphic phantoms

In figure 7, RSP and image noise reconstructions of experimental and simulated data, for three different phantoms are shown: the homogeneous water phantom, the sensitometric CTP404 phantom with tubular inserts as well as three slices of the CIRS pediatric head phantom. Figure 7(a) shows the RSP reconstruction of the experimental data. Standard deviation maps shown in figures 7(b) and (c) all show an increased noise level at the hull of the object. In general, experimental and simulated data agreed well visually and subtle noise features near heterogeneities of the CTP404 phantom and the head phantom were captured.

Table 4. Relative mean and root mean square noise prediction errors for the phantom slices shown in figure 7. Errors are the pixel-wise difference between the simulated standard deviation versus the experimental data, relative to the experimental data.

Phantom (slice)	Mean error/%	Root mean square error/%
Water phantom	−2.9	6.7
CTP404 phantom	−5.6	6.4
Head phantom (1)	−4.6	6.8
Head phantom (2)	−3.6	5.3
Head phantom (3)	3.2	6.2

Table 5. Noise levels (mean standard deviation) within a centered circular region and mean imaging doses calculated using the simulation platform.

Phantom (slice)	Noise level (simulation)	Noise level (experiment)	Imaging dose/mGy (simulation)
Water phantom	0.022 ± 0.001	0.024 ± 0.001	0.85 ± 0.04
CTP404 phantom	0.027 ± 0.001	0.029 ± 0.001	0.86 ± 0.04
Head phantom (1)	0.044 ± 0.003	0.046 ± 0.002	0.82 ± 0.04
Head phantom (2)	0.030 ± 0.001	0.030 ± 0.001	0.84 ± 0.04
Head phantom (3)	0.026 ± 0.001	0.027 ± 0.001	0.88 ± 0.04

In the RSP images of the homogeneous water phantom, ring artifacts were apparent at the center and one larger radius. These were caused by inaccuracies of the calibration at multiples of the brick thickness and are known to degrade RSP accuracy as reported in Sadrozinski *et al* (2016). These ring artifacts became more smeared out in the more heterogeneous phantoms such as the CTP404 and are barely visible in the head phantom, except for the very homogeneous region in the brain (slice 3).
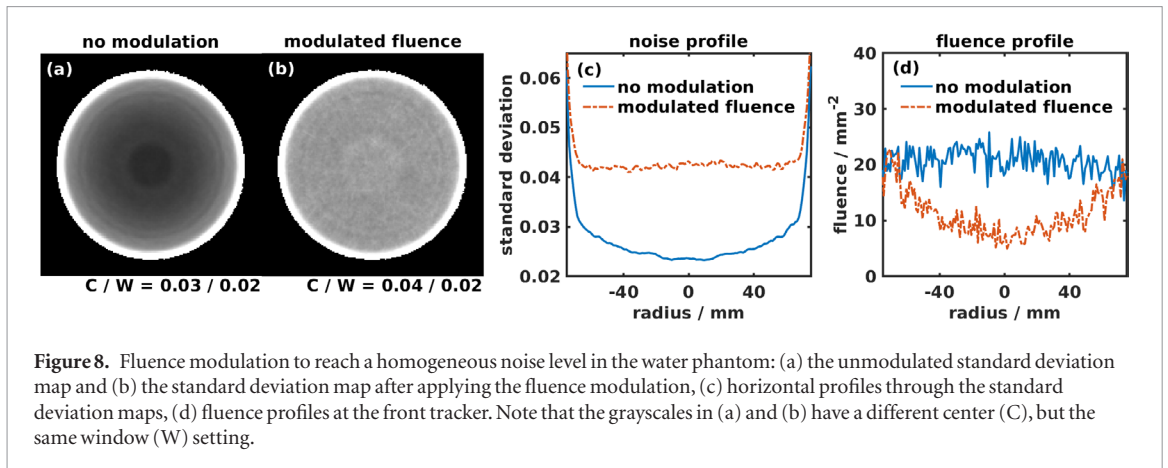
The overall good agreement between simulation and measurement was supported by the difference maps relative to the noise level of the experimental data in figure 7(d). The relative error was increased primarily in regions where ring artifacts also impact the RSP accuracy, as it was seen for the water phantom, but also for the CTP404 phantom and the homogeneous part of the head phantom (slice 3). Table 4 summarizes the mean and the root mean square error for the simulated versus the experimentally acquired standard deviation maps. While the water phantom and the CTP404 phantom both had a negative offset, the head phantom showed both slices where the simulated noise maps over-and under-estimated the experimental level. Root mean square errors are below 10% for all acquisitions.

In figure 7(e), profile plots across the standard deviation maps going horizontally through the isocenter are shown for simulation and measurement. The position coordinate was normalized from the phantom entrance to the exit point. For all phantoms, subtle spatial fluctuations in noise level were captured. The absolute noise level in the center was lower for the homogeneous water phantom and slice 3 of the head phantom (around 0.025). For the more heterogeneous CTP404 phantom and slice 2 of the head phantom, the standard deviation was increased to about 0.03, while for the most heterogeneous slice 1, the noise level exceeded 0.05, i.e. a two-fold increase compared to the homogeneous phantoms. Noise levels as the mean standard deviation within a centered circular region with a radius of 20 mm are summarized in table 5. The scatter-only noise contribution obtained with *WEPL scoring* and shown as a third profile, behaves correspondingly. For the homogeneous water phantom the scatter contribution went to zero at the center and increased towards the edges. For the CTP404 phantom and slice 3 of the head phantom it was 0.01 at the center and had a similar, but less pronounced increase towards the edges. While for slice 2 of the head phantom it reached 0.015, for slice 1 the scatter-only standard deviation was above 0.025. The non-scatter contribution $\sigma_{\text{non-scatter}}$ according to (6) was mostly constant for all phantoms. In the central region excluding the edges it was $\sigma_{\text{non-scatter}} = 0.028 \pm 0.004$, where the uncertainty is calculated over all phantoms.

To allow for a comparison of noise levels in other and future studies, table 5 reports average imaging doses. Experimental imaging doses are expected to be similar as the number of protons was matched and the effect of pile-up was found to be small in previous studies (Dedes *et al* 2018). Reconstruction parameters that impact image noise, such as the voxel size, are given in section 2.2.

3.6. Application: a bow-tie filter for proton CT

In figure 8(a), standard deviation maps are shown for experimental data of the water phantom without post-processing fluence modulation, and in figure 8(b) with a fluence modulation according to (7) that aims at equalizing noise in the projection domain and, therefore, also in the image. Since this part was based on experimental data, fluence modulation was done by rejecting events and only fluence reduction was possible.

Figure 8. Fluence modulation to reach a homogeneous noise level in the water phantom: (a) the unmodulated standard deviation map and (b) the standard deviation map after applying the fluence modulation, (c) horizontal profiles through the standard deviation maps, (d) fluence profiles at the front tracker. Note that the grayscales in (a) and (b) have a different center (C), but the same window (W) setting.

Therefore, the modulated image noise map in figure 8(a) had a higher average noise level compared to the unmodulated map in figure 8(b). While standard deviation was clearly lower in the center of figure 8(a) compared to the outside, the virtual bow-tie modulation succeeded in having a homogeneous noise level. At the outer hull of the phantom, noise is purposely kept high as discussed in section 2.10. In figure 8(c), horizontal profile plots across the image noise maps are shown, which demonstrated again the homogeneous profile that can be reached with the virtual bow-tie filter. Figure 8(d) shows the fluence at the front tracker. While fluence was homogeneous in the unmodulated data, it was reduced in the center of the object in order to achieve homogeneous image noise. RSP accuracy of the fluence-modulated reconstruction was not degraded compared to the non-modulated reconstruction.

4. Discussion

4.1. Verification of noise reconstruction

Using a ground truth generated from independent noise realizations, we were able to show that noise reconstruction predicts the image standard deviation with a root-mean-square error of only a few percent. Accurate image noise maps for FMpCT may be calculated very efficiently at computational costs similar to a standard filtered backprojection reconstruction and without the need to simulate multiple independent noise realizations for every fluence pattern. The remaining star-shaped discrepancy between the noise reconstruction and the ground truth is systematic and due to approximation of interpolation effects, which was extensively discussed in Rädler *et al* (2018). This pattern is independent of the imaged object and did not impact evaluations in this study, as the same reconstruction algorithm was applied to both experimental and simulated data and a systematic error would cancel out in a relative comparison. Furthermore, for FMpCT, such high-frequency components were not relevant in the given scope, as fluence can only be modulated within the extension of one pencil beam.

4.2. Non-linearities in the detection process

By fitting Birks' law to experimental data, we obtained a Birks' coefficient allowing the simulation of quenching effects in the energy detector's five scintillators. The initial range of protons at the entrance of the detector and the stage thickness were also fitting parameters. The CSDA range of 200 MeV protons in polystyrene is 250.0 mm as calculated from the *PSTAR* database (Berger *et al* 2005). This value needs to be reduced by an equivalent of the physical tracker thickness of 2×0.4 mm in silicon (see Johnson *et al* (2016)) and the energy loss in air, and, therefore, the fitting result given in table 2 were reasonable. The physical stage thickness was 51 mm and thus also this fitting result was in the correct range. Uncertainties in the knowledge of the precise beam energy, the trackers' and the scintillators' RSP, as well as the impact of wrapping materials around the scintillator stages, did not allow to precisely calculate these two values. Given that the fit resulted in values close to the expected ones, it confirmed that leaving them as free parameters is a valid procedure.

The determined Birks' factor in table 2 multiplied by the density of polystyrene ($\rho = 1.06$ g cm^{-3}) was $k_B \cdot \rho = 9.4 \times 10^{-3}$ gMeV^{-1} cm^{-2}, which agreed with values for polystyrene published in literature of $k_B \cdot \rho = 9 \times 10^{-3}$ gMeV^{-1} cm^{-2} in Tretyak (2010) or $k_B \cdot \rho = 14 \times 10^{-3}$ gMeV^{-1} cm^{-2} in Reichhart *et al* (2012).

Using the optimized Birks' factor, the agreement between the simulation and experimental data increased considerably. A subtle remaining difference between the simulation and the experimental data was that no protons are recorded with an energy deposit of less than 20 MeV in the first stage. This was because the first stage is needed for triggering recording of an event in experiments and, therefore, protons below this threshold are generally not considered. In the simulation, each proton was treated individually and a threshold was not needed

nor considered. This, however, was no problem, because none of the phantoms in this study had WEPLs above 200 mm.

The parametrization of Birks' law came at almost no additional computational cost for data simulation while a full-scale simulation of the optical photon transport, which is possible with GEANT4, would be unfeasible in terms of computational speed considering the amount of data needed for pCT applications.

In particular, the shape of the calibration curve was highly impacted by quenching. For WEPL accuracy, the exact shape of the calibration curve is irrelevant, since in any case the energy deposit gets assigned the correct WEPL. Therefore, in Giacometti *et al* (2017a) a high RSP fidelity was achieved even without a quenching model. However, non-linearities in the detection process may impact accuracy of the WEPL image noise prediction. This is because the change in WEPL caused by a given change in the energy deposit (e.g. due to range straggling in a prior stage) is proportional to the inclination of the calibration curve, which is incorrect when neglecting quenching. Furthermore, having calibrated energy deposits of the simulation agree with those of a measurement, allowed to fine-tune the beam energy spread in the next section.

4.3. Realistic beam model

By exploiting tracking information from experimental data, we created a virtual beam model to be used in simulations. The estimated beam energy spread given in section 3.3 was lower than the value typically expected for a proton treatment facility of about 0.5%–1% of the initial beam energy as described in Schippers (2017), which would be around 1 MeV–2 MeV in the case of 200 MeV protons. However, in order to reduce fluence to a level that can be handled by the pCT scanner, momentum slits of the accelerator needed to be closed down to a minimum which is atypical for treatment operation, and which may explain the lower energy spread.

4.4. Comparison to experimental data in the projection domain and noise contributions

Using different scoring techniques, we disentangled noise contributions in simulated data. For the continuous part of the wedge-shaped calibration phantom, we observed that the scatter-only component was low for front tracker binning and highly elevated for rear tracker binning. This was because of the lateral widening of the beam envelope as protons traverse the phantom. Due to the geometry of the calibration phantom shown in figure 1 with the wedge facing the front tracker, for a given front tracker location, protons will go through a small distribution of WEPLs. However, for a given rear tracker location, protons can have scattered from a larger cone due to the drift in air (a passage with a low scattering probability) after the phantom and thus the WEPL distribution was broader. Because the accumulated mean scattering angle increases with the WEPL there was an increased WEPL uncertainty towards higher WEPLs.

Moreover, an abrupt decrease of the standard deviation was observed at multiples of the brick thickness. This has a similar geometrical reason as the region of reduced statistics described in section 3.2. Assume a proton going through the center of the wedge phantom. By adding a brick and moving to the periphery of the wedge, the WEPL stays almost constant and so does the accumulated mean scattering angle. However, the geometrical distance of the drift in air is reduced and, therefore, noise decreases (in figure 1 compare the upper proton with the lower proton with one additional brick).

The main contribution of projection variance came from energy straggling, which occurs both within the imaging object as well as in the scintillating detector. Energy straggling in the object increased proportionally to the WEPL. Energy straggling in the detector decreased accordingly such that the total contribution of energy straggling stayed constant. Apart from energy straggling in the scintillator, noise in the detection process was also increased due to the impact of the calibration curve, which itself is noisy and of varying slope. Further noise contributions to the detection process, such as the energy resolution of the scintillator and electronic noise, are assumed to be small and independent of the WEPL. They were discussed in Bashkirov *et al* (2016), where the design of the detector was chosen such that WEPL noise is close to the energy straggling limit. Their remaining contribution may be covered by the beam energy spread estimated in section 2.7.

We showed that the contribution of the tracking uncertainty due to the finite resolution of the location and direction measurement is negligible compared to other sources of uncertainty. A considerable contribution, however, came from the beam energy spread, which contributed a fraction of about 20% despite the fact that in section 2.7 it was estimated to be lower than the energy spread during treatment. This completes the work of Rädler *et al* (2018) and is the first noise simulation of a realistic detector system considering all relevant contributions.

4.5. Comparison to experimental data with heterogeneous and anthropomorphic phantoms

By calculating image noise maps for various phantoms with different amounts of heterogeneities and comparing them to experimental data, we demonstrated the feasibility of using Monte Carlo simulated data to predict image noise in complex geometries. Remaining errors were below 10% and, therefore, small compared to the

overall fluctuation of image noise for the anthropomorphic head phantom which varied within a factor of 2.3 in experimental data (see appendix C). Therefore, fluence modulation based on the simulated noise maps should be feasible.

Furthermore, we showed that the absolute image noise level depends on the heterogeneity of the phantom. The increase of the noise level in heterogeneous phantoms is driven by multiple Coulomb scattering along heterogeneities, which were shown to be considerably different even between two slices of the same (anthropomorphic) phantom. This means that in a clinical setting, the imaging dose advantage of pCT over conventional x-ray CT might be less than expected, as the previous study investigating the density resolution of pCT (Schulte *et al* 2005) used a phantom with homogeneous materials and thus may have predicted a reduced noise level compared to a clinically relevant geometry.

The non-scatter noise contribution was found to be comparable for all phantoms (excluding the edge). This agrees with section 4.4 and experiments of Bashkirov *et al* (2016), who found WEPL-variance to be constant over the whole WEPL-range for a homogeneous phantom without heterogeneities. The impact of a single heterogeneity in a controlled setting was demonstrated in appendix B.

4.6. Application: a bow-tie filter for proton CT

We calculated the profile of a virtual bow-tie filter, which makes noise flat at the detector level. The modulation profile was found based on the simulated noise prediction. The resulting noise profile in image domain was flat as desired and mean RSP accuracy was maintained. This result showed the feasibility of using Monte Carlo simulated noise predictions to modulate the fluence in experimental scans according to a given noise prescription. The virtual bow-tie fluence profile was fundamentally different compared to the fluence profile expected in x-ray CT, which is high in the center of the patient to compensate for the stronger attenuation. For proton CT instead, fluence must be lower in the center and higher at the periphery to compensate for noise introduced due to scattering at the object's hull. However, unlike in fan-beam x-ray CT, where imaging is based on the straight path of primary photons, we cannot expect the shape of a bow-tie filter based on a homogeneous water-equivalent model to be generally valid for proton CT. As we showed, noise maps for the pediatric head phantom differ considerably from the ones observed for the water phantom. Therefore, this virtual bow-tie filter is only valid for a given phantom and is presented here to demonstrate the capability of predicting, prescribing and creating a specific noise pattern.

5. Conclusion

We demonstrated the feasibility of using image noise reconstruction to generate tomographic image noise maps using the cone-beam geometry of a prototype pCT scanner by comparing data to a simulated ground truth. By modeling quenching effects, we were able to match calibrated energy deposits of the five-stage energy detector to Monte Carlo simulated data. This allowed us to determine the beam energy spread of the given incident proton beam. Together with experimental tracking data, this allowed us to create an accurate beam model to be used in the simulation that matches the experimental beam in terms of positions, direction vectors, energy and energy spread. In the projection domain, we compared the Monte-Carlo-predicted noise levels with experimental data and disentangled noise contributions. While the contribution of scattering to noise is negligible in the center of a homogeneous phantom, it becomes a dominating source of noise around heterogeneities for an anthropomorphic head phantom. This is a novel finding for a realistic detector model compared to a previous study (Rädler *et al* 2018) that investigated the impact of multiple Coulomb scattering at the edge of a homogeneous phantom for an idealized detector.

In conclusion, we improved a Monte Carlo simulation to yield accurate image noise maps, which we compared to experimental data. The accuracy of predicted noise is better than the expected fluctuations within a head phantom. Therefore, fluence-modulated proton CT based on Monte Carlo simulated image noise maps should be feasible. At the same time, it was shown that a simple ray-tracing model mapping a WEPL to its uncertainty would not be feasible. To demonstrate the use of calculated image noise maps, we obtained a virtual fluence modulation profile to achieve a bow-tie-like homogeneous noise level in the image. While this was successful for a homogeneous water phantom, such fluence modulation profiles will depend on the patient geometry and cannot be generalized. This suggests that patient-specific fluence-modulation is a crucial component for dose-efficient pCT imaging.

Acknowledgments

tomography: a new approach for low-dose image guidance in particle therapy' and the DFG's Cluster of Excellence Munich-Centre for Advanced Photonics (MAP), by the Bavaria-California Technology Center (BaCaTeC) and by the Franco–Bavarian university cooperation center (BayFrance).

Appendix A. Derivation of the distorted energy deposit

The scaling factor S_n in (2) is fixed during the calibration procedure where for a degrader-free run the energy deposits of all stages are forced to be equal to values $E_n^{\mathrm{G4}} = \{25.25, 28.01, 32.76, 42.62, 67.71\}$ MeV pre-calculated by GEANT4 (see details in Bashkirov *et al* (2016)). Therefore we find

$$E_n^{\mathrm{G4}} = S_n \cdot \int_{R_n^i}^{R_n^f} \mathrm{d}x \, \frac{\mathrm{d}E/\mathrm{d}x}{1 + k_B \cdot \mathrm{d}E/\mathrm{d}x}, \tag{A.1}$$

where R_n^i and R_n^f are the proton's residual ranges at the entrance and at the exit of the stage, respectively. Now, let l_{stage} be the thickness of each scintillating stage and R_0 be the residual range in the detector material of a proton at the entrance of the detector without any degrader. Then we find for the degrader-free calibration run

$$R_n^i = R_0 - (n-1) \cdot l_{\mathrm{stage}}, \tag{A.2}$$

$$R_n^f = R_0 - n \cdot l_{\mathrm{stage}}. \tag{A.3}$$

For a proton that additionally passed through a WEPL during the calibration with the wedge-shaped phantom or during a measurement, we can calculate

$$R_n = R_n^i - \mathrm{WEPL}/\mathrm{RSP}_{\mathrm{det}}, \tag{A.4}$$

where $\mathrm{RSP}_{\mathrm{det}} = 1.038$ is the relative stopping power of the detector material (see Bashkirov *et al* (2016)). By using (2) and (A.1)–(A.4) this yields

$$E_n'(\mathrm{WEPL}) = E_n^{\mathrm{G4}} \cdot \frac{\displaystyle\int_{R_n^i - \mathrm{WEPL}/\mathrm{RSP}_{\mathrm{det}}}^{0} \mathrm{d}x \, \frac{\mathrm{d}E/\mathrm{d}x}{1 + k_B \cdot \mathrm{d}E/\mathrm{d}x}}{\displaystyle\int_{R_n^i}^{R_n^f} \mathrm{d}x \, \frac{\mathrm{d}E/\mathrm{d}x}{1 + k_B \cdot \mathrm{d}E/\mathrm{d}x}}, \tag{A.5}$$

where $E_n'(\mathrm{WEPL})$ is the expected energy measurement for a given Birks' factor k_B.

Appendix B. Noise contributions at a discontinuity

As a side study to investigations in section 2.8, and to demonstrate the impact of heterogeneities, we calculated standard deviation profiles for the steep edge of the calibration phantom with two bricks (upper part of the phantom as seen in figure 1). There, the WEPL rapidly drops from 101.8 mm to zero. The same scoring techniques as for the smooth part of the phantom were exploited. Figure B1 shows a lateral profile of standard deviation for the steep edge of the phantom. In the vicinity of the discontinuity located at 104.5 mm from the isocenter, standard deviation is increased to 27 mm in the measurement, which is an increase by a factor of 8.1 compared to the homogeneous region. *WEPL scoring* standard deviation representing the scatter-only contribution is at 16 mm and thus more than half of the total standard deviation. Agreement between the measurement and *realistic scoring* is satisfactory.

This shows that a discontinuity of 101.6 mm increased the WEPL standard deviation by a factor of 8.1. This distinct increase of noise was driven by the scatter-only component, which was negligible for the homogeneous phantom in figure 6, but contributed a large noise fraction at the heterogeneity. Therefore, heterogeneities are likely to dominate image noise in certain regions and need to be considered using a precise patient model. To achieve the same noise level at the edge as for the homogeneous wedge phantom, the dose would need to be increased by a factor of 65. However, this is an extreme case and neglects the smoothing effect of interpolation and filtering during the reconstruction. Moreover, this does not mean that the dose increase would be required considering the low diagnostic value of RSP at the skin. The Monte Carlo simulation modeled scattering correctly, as the resulting noise increase is captured well when compared to the measurement.

Appendix C. Full-volume RSP and noise histograms

As a sanity check, we calculated RSP and standard deviation histograms of the whole volume of the pediatric head phantom. For this, the two acquisitions of the superior and the inferior part of the phantom needed to be merged. The phantom's outer hull was determined per slice as described in section 2.9 and values outside of it
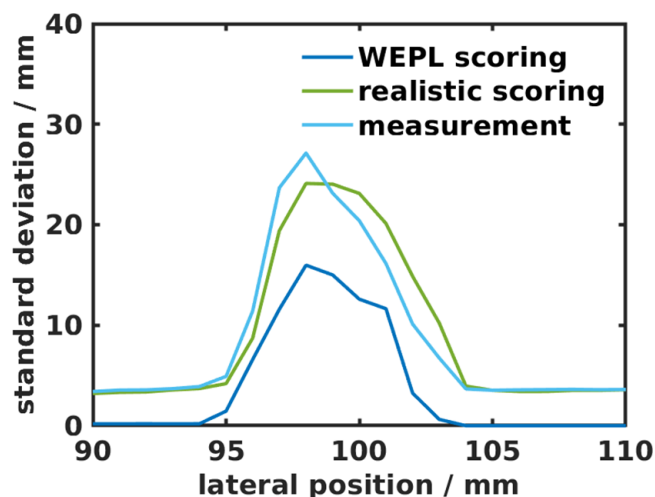
Figure B1. Standard deviation profile along the lateral coordinate for a discontinuity of two bricks (ΔWEPL = 101.6 mm) and for front tracker binning. The discontinuity in WEPL is located at 104.5 mm from the isocenter. Note that the scale is different by a factor of 10 compared to figure 6(a).
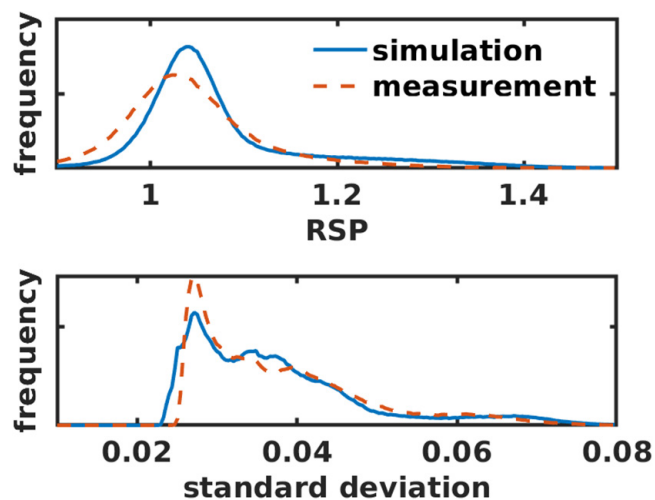

Figure C1. RSP and standard deviation histograms for simulated and experimental data of all slices of the pediatric head phantom.

were neglected. Slices close to the detector's upper and lower edges were also excluded, because noise increases dramatically there.

Figure C1 shows histograms of RSP and standard deviation. While the RSP distribution showed a Gaussian shape around a mean value slightly above 1, the distribution of standard deviation values exhibited a broader structure. For simulated data, the standard deviation 5- to 95-percentile range was between 0.026 and 0.064 while for measurements it was between 0.026 and 0.061. This is an increase from the lower to the higher percentile value by a factor of 2.5 or 2.3, respectively. Note that the lower percentile value is close to the non-scatter contribution $\sigma_{\text{non-scatter}}$ in section 3.5.

In conclusion, over the complete volume of the pediatric head phantom, agreement of standard deviation histograms was comparable to the agreement of RSP histograms.

ORCID iDs

J Dickmann https://orcid.org/0000-0001-8410-3995
V Bashkirov https://orcid.org/0000-0002-7850-3105
R W Schulte https://orcid.org/0000-0002-7892-2756

References

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

Alaei P and Spezi E 2015 Imaging dose from cone beam computed tomography in radiation therapy *Phys. Medica* **31** 647–58

Arbor N, Dauvergne D, Dedes G, Létang J M, Parodi K, Quiñones C T, Testa E and Rit S 2015 Monte Carlo comparison of x-ray and proton CT for range calculations of proton therapy beams *Phys. Med. Biol.* **60** 7585–99

Artikov A, Budagov J, Chirikov-Zorin I, Chokheli D, Lyablin M, Bellettini G, Menzione A, Tokar S, Giokaris N and Manousakis-Katsikakis A 2005 Properties of the Ukraine polystyrene-based plastic scintillator UPS 923A *Nucl. Instrum. Methods Phys. Res.* A **555** 125–31

Bartolac S and Jaffray D 2013 Compensator models for fluence field modulated computed tomography *Med. Phys.* **40** 121909

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

Berger M, Coursey J, Zucker M and Chang J 2005 ESTAR, PSTAR, and ASTAR: computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions (version 1.2.3) (https://doi.org/10.18434/T4NC7P)

Birks J B 1951 Scintillations from organic crystals: specific fluorescence and relative response to different radiations *Proc. Phys. Soc.* A **64** 874–7

Broyden C G 1970 The convergence of a class of double-rank minimization algorithms *IMA J. Appl. Math.* **6** 222–31

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

Coutrakon G, Bashkirov V, Hurley F, Johnson R, Rykalin V, Sadrozinski H and Schulte R 2013 Design and construction of the 1st proton CT scanner *AIP Conf. Proc.* **1525** 327–31

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, 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

Esposito M *et al* 2018 PRaVDA: the first solid-state system for proton computed tomography *Phys. Medica* **55** 149–54

Gang G J, Siewerdsen J H and Stayman J W 2017 Task-driven optimization of fluence field and regularization for model-based iterative reconstruction in computed tomography *IEEE Trans. Med. Imaging* **36** 2424–35

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

Giacometti V, Guatelli S, Bazalova-Carter M, Rosenfeld A and Schulte R 2017b Development of a high resolution voxelised head phantom for medical physics applications *Phys. Medica* **33** 182–8

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

Hansen D C, Sangild Sørensen T and Rit S 2016 Fast reconstruction of low dose proton CT by sinogram interpolation *Phys. Med. Biol.* **61** 5868–82

Harpen M D 1999 A simple theorem relating noise and patient dose in computed tomography *Med. Phys.* **26** 2231–4

Huck S, Fung G, Parodi K and Stierstorfer K 2019 Optimized intensity modulation for a dynamic beam attenuator in x-ray computed tomography *Proc. SPIE Med. Imaging* **10948** 1094824

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

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

Li T, Liang Z, Singanallur J V, Satogata T J, Williams D C and Schulte R W 2006 Reconstruction for proton computed tomography by tracing proton trajectories: a Monte Carlo study *Med. Phys.* **33** 699–706

Mao A, Gang G J, Shyr W, Levinson R, Siewerdsen J H, Kawamoto S and Webster Stayman J 2018 Dynamic fluence field modulation for miscentered patients in computed tomography *J. Med. Imaging* **5** 043501

Meyer S, Gianoli C, Magallanes L, Kopp B, Tessonnier T, Landry G, Dedes G, Voss B and Parodi K 2017 Comparative Monte Carlo study on the performance of integration- and list-mode detector configurations for carbon ion computed tomography *Phys. Med. Biol.* **62** 1096–112

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

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

Penfold S N, Rosenfeld A B, Schulte R W and Schubert K E 2009 A more accurate reconstruction system matrix for quantitative proton computed tomography *Med. Phys.* **36** 4511–8

Penfold S N, Schulte R W, Censor Y and Rosenfeld A B 2010 Total variation superiorization schemes in proton computed tomography image reconstruction *Med. Phys.* **37** 5887–95

Pettersen H E S *et al* 2017 Proton tracking in a high-granularity digital tracking calorimeter for proton CT purposes *Nucl. Instrum. Methods Phys. Res.* A **860** 51–61

Piersimoni P, Ramos-Méndez J, Geoghegan T, Bashkirov V A, Schulte R W and Faddegon B A 2017 The effect of beam purity and scanner complexity on proton CT accuracy *Med. Phys.* **44** 284–98

Poludniowski G, Allinson N M and Evans P M 2014 Proton computed tomography reconstruction using a backprojection-then-filtering approach *Phys. Med. Biol.* **59** 7905–18

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

Ramachandran G N and Lakshminarayanan A V 1971 Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms *Proc. Natl Acad. Sci.* **68** 2236–40

Reichhart L *et al* 2012 Quenching factor for low energy nuclear recoils in a plastic scintillator *Phys. Rev.* C **85** 065801

Rinaldi I, Brons S, Gordon J, Panse R, Voss B, Jäkel O and Parodi K 2013 Experimental characterization of a prototype detector system for carbon ion radiography and tomography *Phys. Med. Biol.* **58** 413–27

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

Sadrozinski H F *et al* 2016 Operation of the preclinical head scanner for proton CT *Nucl. Instrum. Methods Phys. Res.* A **831** 394–9

Schippers J M 2017 Beam-transport systems for particle therapy *Proc. CAS-CERN Accel. Sch. Accel. Med. Appl.* **1** 241–52

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

Shunhavanich P, Hsieh S S and Pelc N J 2019 Fluid-filled dynamic bowtie filter: description and comparison with other modulators *Med. Phys.* **46** 127–39

Stayman J W, Mathews A, Zbijewski W, Gang G, Siewerdsen J, Kawamoto S, Blevis I and Levinson R 2016 Fluence-field modulated x-ray CT using multiple aperture devices *Proc. SPIE Med. Imaging* **9783** 97830X

Szczykutowicz T and Mistretta C 2013a Design of a digital beam attenuation system for computed tomography. Part II. Performance study and initial results *Med. Phys.* **40** 21906

Szczykutowicz T P and Mistretta C A 2013b Design of a digital beam attenuation system for computed tomography. Part I. System design and simulation framework *Med. Phys.* **40** 21905

Szczykutowicz T P, Hermus J, Geurts M and Smilowitz J 2015 Realization of fluence field modulated CT on a clinical TomoTherapy megavoltage CT system *Phys. Med. Biol.* **60** 7245–57

Tretyak V I 2010 Semi-empirical calculation of quenching factors for ions in scintillators *Astropart. Phys.* **33** 40–53

Volz L, Piersimoni P, Bashkirov V A, Brons S, Collins-Fekete C A, Johnson R P, Schulte R W and Seco J 2018 The impact of secondary fragments on the image quality of helium ion imaging *Phys. Med. Biol.* **63** 195016

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

Wunderlich A and Noo F 2008 Image covariance and lesion detectability in direct fan-beam x-ray computed tomography *Phys. Med. Biol.* **53** 2471–93

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