

ACCEPTED MANUSCRIPT

Experimental comparison of proton CT and dual energy X-ray CT for relative stopping power estimation in proton therapy

To cite this article before publication: Georgios Dedes *et al* 2019 *Phys. Med. Biol.* in press <https://doi.org/10.1088/1361-6560/ab2b72>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2019 Institute of Physics and Engineering in Medicine.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Experimental comparison of proton CT and dual energy X-ray CT for relative stopping power estimation in proton therapy

George Dedes^{1,†}, Jannis Dickmann¹, Katharina Niepel¹, Philipp Wesp¹, Robert P. Johnson², Mark Pankuch³, Vladimir Bashkirov⁴, Simon Rit⁵, Lennart Volz^{6,7}, Reinhard W. Schulte^{4,*}, Guillaume Landry^{1,8,9,*} and Katia Parodi^{1,*}

¹ Department of Medical Physics, Ludwig-Maximilians-Universität München, Garching b. München, Germany

² Department of Physics, U. C. Santa Cruz, Santa Cruz, CA, United States of America

³ Northwestern Medicine Chicago Proton Center, Warrenville, IL, 60555, United States of America

⁴ Division of Biomedical Engineering Sciences, Loma Linda University, Loma Linda, CA, United States of America

⁵ Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69373, Lyon, France

⁶ Department of Biomedical Physics in Radiation Oncology, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

⁷ Department of Physics and Astronomy, Heidelberg University, Heidelberg, Germany

⁸ Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany

⁹ German Cancer Consortium (DKTK), Munich, Germany

* Authors contributed equally.

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

Abstract. Proton computed tomography (pCT) has been proposed as an alternative to X-ray computed tomography (CT) for acquiring relative to water stopping power (RSP) maps used for proton treatment planning dose calculations. In parallel, it has been shown that dual energy X-ray CT (DECT) improves RSP accuracy when compared to conventional single energy X-ray CT. This study aimed at directly comparing the RSP accuracy of both modalities using phantoms scanned at an advanced prototype pCT scanner and a state-of-the-art DECT scanner.

Two phantoms containing 13 tissue-mimicking inserts of known RSP were scanned at the pCT phase II prototype and a latest generation dual-source DECT scanner (Siemens SOMATOM Definition FORCE). RSP accuracy was compared by mean absolute percent error (MAPE) over all inserts. A highly realistic Monte Carlo (MC) simulation was used to gain insight on pCT image artifacts which degraded MAPE.

MAPE was 0.55% for pCT and 0.67% for DECT. The realistic MC simulation agreed well with pCT measurements (MAPE = 0.69%). Both simulation and

† Author to whom correspondence should be addressed

1
2
3 *pCT vs. DECT RSP accuracy*
4

5 experimental results showed ring artifacts in pCT images which degraded the MAPE
6 compared to an ideal pCT simulation (MAPE = 0.17%). Using the realistic simulation,
7 we could identify sources of artifacts, which are attributed to the interfaces in the five-
8 stage plastic scintillator energy detector and calibration curve interpolation regions.
9 Secondary artifacts stemming from the proton tracker geometry were also identified.

10 The pCT prototype scanner outperformed a state-of-the-art DECT scanner in
11 terms of RSP accuracy (MAPE) for plastic tissue mimicking inserts. Since artifacts
12 tended to concentrate in the inserts, their mitigation may lead to further improvements
13 in the reported pCT accuracy.
14

15
16
17
18 Submitted to: *Phys. Med. Biol.*
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Introduction

The increased use of protons for external beam radiation therapy of cancer offers the potential of sparing of healthy tissues by achieving highly conformal dose distributions to the tumor. Protons gradually slow down when traversing matter and their initial energy can be chosen such that the therapeutic proton beam stops in the tumor. The range of the protons in the patient is dependent on their energy and the stopping power of the tissue they traverse. The latter is commonly expressed as relative to water and denoted as relative stopping power (RSP). A highly accurate RSP map of the patient in treatment position is crucial for the calculation of optimal proton therapy treatment plans. Any inaccuracies in the RSP map will be translated to proton range prediction errors, thus compromising treatment plan quality by introducing range uncertainties, for which safety margins are required (Paganetti 2012).

The current practice for obtaining RSP images is based on converting X-ray linear attenuation coefficients, acquired with single energy X-ray CT (SECT) imaging, to RSP. This procedure involves a stoichiometric calibration (Schneider et al. 1996) which can result in up to 3.5% errors in the determination of RSP (Paganetti 2012, Yang et al. 2012). Proton CT (pCT) is considered as a candidate for improving RSP accuracy. Already in 1963, Cormack (1963) proposed that protons could be used for tomographic imaging. Huesman et al. (1975) described a pCT scanner concept relying on the registration of individual protons' positions and directions, as well as the residual energy behind the patient. pCT scanners effectively measure RSP line integrals which can be processed by dedicated reconstruction algorithms (Penfold et al. 2009, Penfold et al. 2010, Rit et al. 2013, Poludniowski et al. 2014, Hansen et al. 2016) to yield three-dimensional RSP maps.

Prototype pCT scanners have been designed and constructed (Takada et al. 1988, Coutrakon et al. 2013, Sadrozinski et al. 2016, Johnson et al. 2016, Pettersen et al. 2016, Esposito et al. 2018). In recent studies, the RSP accuracy achieved by some of these prototypes has been reported to be better than 1.6% for three inserts (Esposito et al. 2018) and 1.39% for seven inserts (Giacometti et al. 2017). For the same seven inserts, Volz et al. (2018) achieved accuracy better than 1% using helium ions with a pCT prototype.

An alternative for obtaining RSP images of high accuracy is dual energy X-ray CT (DECT) (Yang et al. 2010), where several studies (Hünemohr et al. 2013, Bourque et al. 2014, Hudobivnik et al. 2016, Möhler et al. 2016, Han et al. 2016, Taasti et al. 2016, Lalonde et al. 2017, Saito & Sagara 2017b, Almeida et al. 2018) have demonstrated the potential of reaching an RSP accuracy of about 1%. DECT methods for RSP estimation have been recently validated using biological tissue samples (Taasti et al. 2017, Bär et al. 2018, Xie et al. 2018, Möhler et al. 2018) and have been found to consistently outperform SECT in terms of RSP accuracy. DECT scanners are currently making their way into proton therapy clinics (Wohlfahrt et al. 2017a) and may impact clinical proton range calculation (Hudobivnik et al. 2016, Wohlfahrt et al. 2017b).

pCT vs. DECT RSP accuracy

4

While a comparison between ideal simulated pCT and DECT measurements in terms of RSP accuracy has suggested that pCT may achieve superior results (Hansen et al. 2015), there has so far been no direct comparison between the two modalities. This study aims at filling this gap.

The current study is split in two. The main part presents for the first time an experimental comparison of pCT and DECT in terms of RSP accuracy. For that purpose, two different phantoms containing a total of 13 tissue equivalent inserts of known RSP have been scanned in the phase II preclinical pCT prototype scanner and in a commercially available DECT scanner. This allowed a direct comparison of the accuracy of the resulting RSP images.

Since pCT technology is at a much earlier stage of development than state-of-the-art dual source DECT scanners, we expect that pCT images may suffer from artifacts and other effects absent in DECT. In the second part of our study, we thus supplemented our experimental work with ideal and realistic pCT detector simulations. Validation of the realistic simulations against experimental results allowed us to assess whether pCT accuracy may be impacted by fundamental limitations of using protons for imaging, or from design aspects of the pCT prototype we used. Additionally, the simulations were used to pinpoint sources of image artifacts in the pCT scanner design.

2. Material and methods

2.1. Experimental aspects

2.1.1. Known-RSP phantoms Two phantoms containing plastic tissue equivalent inserts were used in this study (see drawings in figure 2). The first phantom is the CTP404 module from the commercial multislice Catphan[®]600 phantom (The Phantom Laboratory, New York, USA), denoted for simplicity in the rest of the manuscript as CTP404 phantom. The CTP404 phantom is of cylindrical shape with a diameter of 150 mm and height of 25 mm, made of polystyrene. The phantom contained eight radially placed cylindrical inserts of 12.2 mm radius and 25 mm height. It also contained other smaller inserts of various materials (tungsten carbide wires, air and Teflon small cylindrical inserts and acrylic spherical inserts of various radii) which are not relevant to this study. During the imaging scans, two of the inserts were filled with air, while the rest contained materials whose RSP is summarized in table 4. The reference RSP was obtained by means of variable water column measurements with a 4.4 mm FWHM 310.82 MeV/u carbon ion beam (Giacometti et al. 2017).

The second phantom is a custom-made acrylic (PMMA) cylinder of 130 mm diameter and 227 mm height, called henceforth LMU phantom. The phantom body contained seven cylindrical holes of 30 mm diameter and of 55 mm height, in which different plastic tissue equivalent materials were inserted. The seven inserts were arranged in three different layers along the axis of the cylinder of the phantom body. In two of the layers one insert was placed centrally and one off-center and in a third

1
2
3 *pCT vs. DECT RSP accuracy* 5

4 layer one insert was placed centrally and two off-center. Due to the fact that the
5 LMU phantom was in the vertical direction larger than the scanner's field of view, its
6 three insert layers were scanned in three different runs. Therefore, the LMU phantom
7 is presented as three different images, hereafter called Top Supremum (*TopSup*), Top
8 Infimum (*TopInf*) and Bottom (*Bot*). The reference RSP of the seven inserts were
9
10
11 85 obtained by means of variable water column measurements with a 4.4 mm FWHM
12 310.82 MeV/u carbon ion beam (Berndt 2016, Hudobivnik et al. 2016) and are listed in
13 table 4.
14
15

16
17 *2.1.2. Proton CT scanner* The proton CT scans of this study were performed at
18 90 the Northwestern Medicine Chicago Proton Center using the phase II preclinical pCT
19 prototype scanner of the Loma Linda University and the University of California Santa
20 Cruz. The scanner hardware is described in detail in Johnson et al. (2016). It comprises
21 two tracking modules and an energy detector for the determination of the water-
22 equivalent path length (WEPL). The data acquisition system is capable of acquiring
23 broad beam proton events at a sustained rate in excess of 1 MHz.
24
25 95

26 The front and rear tracking modules (upstream and downstream of the scanned
27 object) contain in total 32 single-sided silicon strip detectors (SSD) with a strip pitch
28 of 228 μm and a thickness of 400 μm . In each tracking module, the silicon strip detectors
29 are arranged in four layers, two measuring the horizontal coordinates and two measuring
30 the vertical coordinates. The tracking system provides the capability of a four-point
31 100 measurement for each proton, allowing the estimation of the curved proton path using
32 the measured positions and calculated directions.
33
34

35 The WEPL detector (Bashkirov et al. 2016) consists of five polystyrene scintillator
36 stages (RSP ≈ 1.038), each with a thickness of 51 mm and a lateral area of $10 \times 40 \text{ cm}^2$.
37
38 105 Each scintillator stage is wrapped with 65 μm thick reflective material (VikuitTM ESR
39 film). The WEPL of a proton is deduced from the energy detector signal of the stage
40 in which the proton stopped, using a stage-specific energy-to-WEPL calibration curve.
41 The calibration procedure is described in section 2.3.
42

43 The WEPL calibration, as well as the conversion of the digitized signal from the
44 110 tracker (strip number) to physical coordinates, and the assembly of proton events from
45 information registered by the different parts of the detector are performed by dedicated
46 raw data processing software. The output of the processing software is fed to the
47 reconstruction algorithm (described in section 2.4) which produces a voxelized RSP
48 map of the scanned object.
49
50

51 115 Both the data acquisition required for the WEPL calibration and the imaging
52 runs were performed with a wobbled proton beam of 40 mm FWHM size, which was
53 magnetically deflected to sweep the field of view (FOV) of the scanner. The nominal
54 proton energy was 200 MeV as defined in the accelerator library and the variation was
55 found to be less than 0.2 MeV between runs. Finally, the phantoms used for imaging
56
57
58 120 were placed on a rotating stage and data were acquired while continuously rotating.
59 Data were subsequently binned in 360 projections, spaced at 1 degree steps. pCT scans
60

1
2
3 *pCT vs. DECT RSP accuracy* 6

4 took approximately 6 min.

7
8 *2.1.3. Dual energy CT scanner and conversion to RSP* RSP estimates based on DECT
9 images were calculated following a method proposed by Saito & Sagara (2017a) and
10 Saito & Sagara (2017b). For this purpose a PMMA calibration phantom with 150 mm
11 diameter and a central bore housing inserts from an electron density phantom (RMI 467,
12 Gammex, Middleton, USA, part of the Sun Nuclear Corporation) was scanned using
13 a dual-source DECT scanner (SOMATOM Definition FORCE, Siemens Healthineers,
14 Forchheim, Germany) at peak tube voltages of 90 kVp and 150 kVp with tin filtration.
15 The scan pitch was 0.7 and exposures of 168 mAs were set for both X-ray tubes with
16 automatic exposure control disabled. The CT dose index ($CTDI_{vol}$) was 35.7 mGy.
17 Images were reconstructed using the Q32\3s image reconstruction kernel with the
18 vendor's iterative reconstruction (ADMIRE) on a $0.39 \text{ mm} \times 0.39 \text{ mm} \times 3 \text{ mm}$ grid. Both
19 known-RSP phantoms were scanned the same way. DECT scans took 17 sec.

20
21 Using the mean CT numbers HU_k for the high-energy ($k = H$) and low-energy
22 ($k = L$) scans of the calibration phantom materials, scanner specific calibration
23 parameters were obtained by least-square fitting of the known electron densities relative
24 to water ρ_e and the ratio of effective atomic number Z_{eff} to the following functions:
25

$$26 \quad \rho_e = a \frac{(1 + \alpha)HU_H - \alpha HU_L}{1000} + b \quad (1)$$

$$27 \quad \left(\frac{Z_{eff}}{Z_{eff,w}} \right)^{3.3} - 1 = \gamma_L \left(\frac{u_L}{\rho_e} - 1 \right) + \gamma_0 \quad (2)$$

28 with the effective atomic number of water $Z_{eff,w} = 7.4774$ and reduced CT number
29 $u_L = HU_L/1000 + 1$. The offset γ_0 in equation (2) was added to Saito's original proposal
30 in order to improve the accuracy of the fit. The calibration of the mean excitation
31 energy I relative to that of water I_w was done separately for soft tissues ($Z_{eff} < 8.8$)
32 and bone tissues ($Z_{eff} > 8.8$) by fitting equation (3).
33

$$34 \quad \ln \frac{I}{I_w} = C_1 \left[\left(\frac{Z_{eff}}{Z_{eff,w}} \right)^{3.3} - 1 \right] - C_0 \quad (3)$$

35 The resulting fit parameters are summarized in table 1. The figures showing the fits on
36 the data are shown in section 1 of the *Supplementary Material*.

37 Applying these calibration parameters to the scans of the known-RSP phantoms,
38 maps of relative proton stopping power values for each voxel could be obtained according
39 to the Bethe equation:
40

$$41 \quad RSP = \rho_e \frac{\ln \left(\frac{2m_e c^2 \beta^2}{I(1-\beta^2)} \right) - \beta^2}{\ln \left(\frac{2m_e c^2 \beta^2}{I_w(1-\beta^2)} \right) - \beta^2} \quad (4)$$

42 with rest electron mass m_e , speed of light c , and proton speed relative to the speed of
43 light β . In this work, we used an I_w value of 78 eV according to the latest ICRU report
44 (Sigmund et al. 2009) and a β value of 0.4282, corresponding to a proton kinetic energy
45 of 100 MeV.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

pCT vs. DECT RSP accuracy

Table 1: Calibration parameters (CP) required to convert DECT images to RSP, together with their 95% confidence level (CL).

	CP		CL	
α	0.3452		0.02	
a	0.9928		0.01	
b	0.9929		0.004	
γ_L	9.0814		0.3	
γ_0	-0.0941		0.08	
	soft tissues		bone tissues	
	CP	CL	CP	CL
C_1	0.2020	0.1	0.0662	0.009
C_0	0.0821	0.03	0.0945	0.03

2.2. Proton CT Monte Carlo simulations

A Geant4 (version 10.03.p01) based simulation application, modelling in detail all parts of the pCT phase II prototype scanner, has been developed and published by Giacometti et al. (2017). The physics models used were the G4EmLivermorePhysics for the electromagnetic physics, the G4HadronPhysicsQGSP_BIC_HP for the inelastic interactions of hadrons, the G4HadronElasticPhysicsHP for the elastic interaction of hadrons and the G4IonBinaryCascadePhysics for the inelastic interactions of ions. In some particular cases nuclear/hadronic interactions were switched off, so as to investigate their effect on the RSP accuracy. Whenever this was the case, it is explicitly stated in the text, otherwise the full set of physics was used. The maximum step length was chosen to be 6 μm for the energy detector stages (one tenth of the wrapping material thickness) and 1 mm in the rest of the simulation geometry. For this study an amended version of the simulation code was used. In this version, the simulation includes parameterization of the non-linear response of the scintillator to the deposited energy (Birks' effect). The Birks' factor $k_B = 0.0887$ (Dickmann et al. 2019) was used to modify the deposited energy dE per step dx according to

$$\frac{dE'}{dx} = \frac{dE/dx}{1 + k_B \cdot dE/dx}. \quad (5)$$

The spatial dependence of the scintillator response (Bashkirov et al. 2016), related to the position of the hit with respect to the location of the photomultiplier tube was also implemented.

In addition, the simulation code emulates the digitization process of the real scanner, yielding raw data in the same format as the actual scanner. This means, for each proton the simulation outputs the tracker hits as strip numbers and the energy deposit in each stage of the WEPL detector as ADC numbers. This allowed the use

pCT vs. DECT RSP accuracy

8

of a unified software workflow for experimental and simulated scans. The output of the simulation (calibration and imaging runs) is processed in the same way as the experimental data.

It is important to mention here that the four silicon strip detector modules comprising each layer of a tracking module are arranged in a 1×4 matrix configuration, therefore leaving three gaps in every layer. These gaps are 0.6 mm wide vertical stripes of insensitive areas. Gaps are offset horizontally from layer to layer in order to reduce the probability that a single proton crosses more than one of them. Special care was taken to minimize these gaps, and it was estimated that, at the interface between two silicon detectors there was an opening of about 0.1 to 0.2 mm, that was partially filled with glue. Nevertheless, in the simulations, the whole 0.6 mm wide insensitive area of every gap was modelled as air, which had implications on the reconstructed image quality, as shown later.

Besides the full detector simulation resulting in raw data aiming at closely modeling experimental data, the simulation is also capable of producing idealized data. In this case, the proton's exact position, direction and energy are scored before and after the object at planes coinciding with the trackers.

For simplicity, pCT scans were simulated in step-and-shoot mode, in contrast to the continuous acquisition in experiments. 360 projections at 1 degree steps were simulated for all phantom cases.

An additional water phantom was simulated in order to investigate specific image artifacts. It was modelled after an existing water phantom consisting of a cylindrical PMMA container with outer diameter of 150.5 mm and a height of 40 mm. The wall thickness of the PMMA container was 6.35 mm and the container was filled with water. For the purposes of the current study, a larger version of the phantom (1.5 times larger in diameter) with an outer diameter of 225.75 mm was also simulated. Simulations were performed with both water cylinders centered at the imaging isocenter. To further investigate the artifact related to the tracker gap, a set of three realistic simulations of the 150.5 mm diameter water phantom was performed. In the first simulation, the phantom was centered at the isocenter and the tracker gaps were assumed to be filled with air, as in the case of all previous simulations. In the second simulation, the water phantom was placed with a lateral offset of 40 mm to the isocenter and the tracker gaps were assumed to be filled with air. Finally, in the third simulation, the water phantom was located 40 mm off the isocenter and the tracker gaps were assumed to be filled with silicon.

2.3. Proton CT scanner calibration

The signal from the five-stage detector for each proton is converted to WEPL via a calibration procedure. The concept is described in Bashkirov et al. (2016) and a detailed update based on the current calibration phantom is given in Piersimoni et al. (2017). The calibration phantom is made of a polystyrene wedge which provides a

pCT vs. DECT RSP accuracy

9

WEPL gradient due to the slopes of the wedge, as well as four polystyrene blocks, which when combined with the wedge allow sampling of the entire WEPL range in the five-stage plastic scintillator detector. The calibration runs, namely a run without any object and five runs (wedge alone and wedge plus 1-4 blocks), are performed at the beginning of a scanning session, all with 200 MeV protons. The run without any object is used in order to map the spatial dependence of the energy detector response as well as to provide a conversion from ADC counts (digitized energy detector signal) to energy. The runs with the wedge phantom are utilized for the creation of a look-up table, associating the known WEPL a proton traversed in the calibration runs to the energy deposit in the stopping stage of the five-stage detector. The known WEPL is obtained by calculating the length of the proton trajectory in the calibration phantom from the tracker information, assuming straight paths and knowing the calibration phantom's RSP and geometry. The look-up table, referred to as WEPL calibration, contains a WEPL value for each of the 340 energy deposit bins in each of the five stages (in total 1700 energy bins of 0.25 MeV bin width). To obtain an optimal calibration, fits and interpolations can be applied in regions where either a lack of sufficient data or geometrical effects distort the calibration. The standard practice so far was to make an attempt of correcting also for detector effects. This approach was based on the expectation that the calibration curve should be continuous and smooth. In this study we investigate whether this assumption is valid by toggling corrections.

All these effects and the corresponding corrections are explained in section 2.6. Subsequent imaging runs are processed using the look-up table and energy deposits of protons in the stopping stage are converted to WEPLs.

2.3.1. Calibration curve variants Alternative calibration curves (table 2) were generated to assess whether the calibration procedure may contribute to image artifacts and decrease RSP accuracy. The calibration obtained from experimental data as described previously and including all corrections is referred to as *ExpCalib1*. A variant called *ExpCalib2* was derived by omitting all applied corrections (fits and interpolations). A last experimental calibration named *ExpCalib3* was derived by omitting only the corrections related to the stage interfaces. We generated additional calibrations using the simulation platform. The *ExpCalib1*-equivalent calibration from simulations is called *SimCalib1*. In *SimCalib2*, corrections related to the stage interfaces were omitted.

2.4. Proton CT image reconstruction

The algorithm used to reconstruct the pCT images was a filtered back projection (FBP) implementation that accounts for the curved proton paths. A detailed description of the underlying principles of the algorithm is given in Rit et al. (2013). The path of every proton, curved due to multiple Coulomb scattering, is approximated by a most likely path (MLP) formulation, introduced by Schulte et al. (2008). For the determination of the curved path, the position and direction information from the tracking modules

Table 2: Variants of the WEPL calibration with their main parameters.

Calibration name	Type of data	Interpolation correction	Stage interface correction
<i>ExpCalib1</i>	Experimental	Yes	Yes
<i>ExpCalib2</i>	Experimental	No	No
<i>ExpCalib3</i>	Experimental	Yes	No
<i>SimCalib1</i>	Simulation	Yes	Yes
<i>SimCalib2</i>	Simulation	Yes	No

is necessary. To eliminate protons stemming from nuclear interactions, cuts on the energy and angular distributions were applied. The cuts rejected protons whose energy or angle were outside three standard deviations around their median energy and angle. For the cuts, protons were grouped together in $2\text{ mm} \times 2\text{ mm}$ pixels according to their position at the front tracker module. The resulting proton-per-proton data were binned in projection images with $1\text{ mm} \times 1\text{ mm}$ pixels. The projections were then filtered and back-projected. Finally, the pCT images were reconstructed as RSP maps in a grid of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$. The reconstruction was applied to both experimental and simulated data without change of parameters.

2.5. RSP accuracy quantification

The reconstructed pCT images, from simulations and experimental data as well as from the DECT experimental scans, were compared in terms of RSP accuracy. The latter was quantified for the cylindrical inserts of the phantoms as follows: in a cylindrical region of interest (ROI) in the image, concentric with the inserts, the mean RSP (RSP_{mean}) of the voxels in that region was calculated. The accuracy is then the difference of the mean RSP from the reference RSP (RSP_{ref}) in percentage:

$$\text{RSP}_{\text{acc}} = 100 \cdot \frac{\text{RSP}_{\text{mean}} - \text{RSP}_{\text{ref}}}{\text{RSP}_{\text{ref}}} \% \quad (6)$$

The ROI radius was chosen to have 50% of the radius of the cylindrical inserts. Furthermore, the ROIs spanned across 15 slices (1 mm each) for the pCT scans and across 5 slices (3 mm each) for the DECT scans. In addition to the RSP accuracy for each insert, the mean absolute percentage error (MAPE) from all inserts and for each imaging modality was calculated according to:

$$\text{MAPE} = \frac{\sum_{i=1}^n |\text{RSP}_{\text{acc},i}|}{n} \quad (7)$$

where n is the total number of inserts and $\text{RSP}_{\text{acc},i}$ is the accuracy in percent for every insert i as calculated from equation 6.

1
2
3 *pCT vs. DECT RSP accuracy*

11

4
5 *2.6. Influence of problematic WEPL intervals on image artifacts*

6
7 The RSP value of a voxel following image reconstruction is obtained from potentially
8 very different WEPL values, since it corresponds to the average of the projections'
9 values over all projection angles. This is because protons intersecting a given voxel,
10 but from different acquisition angles, traverse very different paths in the object. In the
11 special case of cylindrical objects these paths are, in two dimensions, circle's chords. The
12 length of these chords ranges from a minimum length, depending on the radial location
13 of the voxel and the acquisition angle, up to the object's diameter. For example, central
14 voxels in a cylindrical homogeneous object are crossed only by protons of the maximum
15 possible WEPL for that particular object, whereas voxels at the edge of the cylinder will
16 see a wider WEPL distribution. If a certain WEPL interval is systematically distorted
17 by the scanner, this will lead to artifacts in localized regions in the image, depending
18 on the phantom. There are two types of calibration curve regions which may introduce
19 image artifacts due to inaccurate WEPLs: stage interfaces and intra-stage calibration
20 curve kinks. Their WEPL intervals are presented in table 3.
21
22
23
24
25

26 Protons stopping near the interface of two stages of the energy detector are
27 of particular interest since ambiguities in their signal may lead to image artifacts.
28 Additionally, they can distort the calibration curve. There is a number of corrections
29 which can be applied in order to obtain a smooth curve in the region between two
30 adjacent stages. Due to a threshold of 1 MeV in the minimum energy required at the
31 stopping stage, the first four energy bins for every stage contain no WEPL value. They
32 can be arbitrarily set to the WEPL value of one of the next non-zero WEPL energy bins.
33 Furthermore, the last few highest energy bins of every stage are populated by a small
34 number of protons. In order to obtain a smooth curve there, an extrapolation correction
35 can be applied, using the values of lower energy bins. After the aforementioned
36 corrections, there is a WEPL discontinuity between the last energy bin of a stage and
37 the first energy bin of the next stage. This can be removed by using the first energy of
38 the next stage in the extrapolation correction described previously. To summarize, the
39 fact that protons might deposit part of their energy in non-active detector material at
40 the interfaces or split their energy in adjacent stages, in addition to the applied energy
41 thresholds, results in inaccuracies and uncertainties in the corresponding part of the
42 WEPL calibration.
43
44
45
46
47

48 Intra-stage calibration curve kinks correspond to a discontinuity observed in the
49 calibration curves due to the calibration phantom geometry. This discontinuity is
50 observed at about 60 MeV, in every stage except for the fifth (last) stage. The kink
51 originates from the interplay between the geometry of the calibration phantom, the
52 calibration procedure and the fact that the beam is divergent. The kink, which results
53 in a severe artifact if left uncorrected, can be mitigated to a large extent by interpolation
54 correction, using parts of the curve before and after the kink region. However residual
55 WEPL inaccuracies may remain. As opposed to the stage interface correction described
56 above, this correction is detector independent.
57
58
59
60

pCT vs. DECT RSP accuracy

12

Table 3: WEPL ranges corresponding to either interpolation of calibration curve kinks, or stage interfaces.

WEPL region	kinks	stage interfaces
	WEPL range / mm	
1	46.5 – 56.7	37.6 – 40.9
2	97.2 – 107.3	90.1 – 92.2
3	147.9 – 158.7	141.7 – 144.9
4	199.0 – 209.9	193.0 – 196.2

To investigate the impact of the WEPL regions in table 3 on the accuracy in the image, we calculated heatmaps in image domain showing the percentage contribution of a given WEPL range to a voxel of the reconstruction volume. This was done by thresholding a given slice of the reconstruction volume from the experimental pCT images to the nearest known RSP value and calculating a forward projection in parallel beam geometry. The resulting sinogram was set to 1 if its value was within the given WEPL range and to 0 otherwise. Disregarding filtering, the binary sinogram was then backprojected and divided by the number of query points in each summation. This resulted in an image with values in the range $[0, 1]$, which are 0 if the voxel is backprojected from WEPLs that are strictly outside the given WEPL range, and 1 if the voxel is backprojected from WEPLs that are exclusively inside the given WEPL range. This, however, is not to be taken quantitatively, as filtering for CT reconstruction was neglected.

2.7. Proton CT imaging dose estimation

In the case of the pCT scans, no direct dose estimation was possible. Therefore, the imaging dose was calculated with the Monte Carlo simulation code described in 2.2. A dose grid of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ was defined and the dose to material was scored in every voxel and for each projection. The dose delivered in a single scan was obtained by simply summing the doses from all projections. The number of simulated protons per scan (2.7×10^8) was chosen to be approximately equal to that in the experimental scans. The exact dose estimation was obtained by scaling the simulated dose with the factor required to match the number of protons registered by the scanner in simulations to that in each experimental scan. Only physical dose was considered.

3. Results

3.1. Proton CT – Dual energy CT comparison

3.1.1. Proton CT calibration In figure 1 the WEPL calibration is plotted for experimental data with all corrections (*ExpCalib1*), experimental data without any

pCT vs. DECT RSP accuracy

13

corrections (*ExpCalib2*) and for comparison, for simulated data with corrections (*SimCalib1*). For more details about the calibration parameters see table 2. The WEPL ranges to which corrections were applied are listed in table 3. The relative agreement between *ExpCalib1* and *SimCalib1* varied from approximately 1% or lower for stages 1-3, up to 6% for stages 4 and 5. A severe kink appeared in the calibration without interpolation correction (*ExpCalib2*), at approximately 60 MeV and was to a large extent corrected via interpolation (*ExpCalib1* and *ExpCalib3*). Removing the additional corrections concerning the stage interfaces and high energy deposits (*ExpCalib2*) also led to a dramatic distortion of the calibration curve between adjacent energy detector stages. Nevertheless, although the corrections related to stage interfaces yielded a smoother curve, they exacerbated artifacts and led to reduced RSP accuracy. The RSP MAPE achieved with pCT when using *ExpCalib1* was 0.87% for experimental data and 0.86% for simulations. Maximum errors exceeded 1.5%. The optimal calibration applied to experimental and realistic pCT simulations was obtained with interpolation at the kink region, but without stage interface corrections. Therefore, unless mentioned otherwise, *ExpCalib3* was used for experimental pCT data and *SimCalib2* for simulated pCT data. Detailed RSP accuracy results from this optimized calibration variant are presented in 3.1.2, together with results for DECT.

3.1.2. RSP accuracy The reconstructed RSP images from experimental pCT and experimental DECT are presented in figure 2. As explained in section 2.5, the RSP accuracy was quantified in cylindrical ROIs. The RSP accuracy as a function of the reference RSP is shown in figure 3 (top) for the experimental pCT and experimental DECT data, and in figure 3 (bottom) for the realistic simulation pCT and experimental DECT data. The experimental pCT RSP accuracy showed a tendency towards underestimation, which was well reproduced by the realistic simulation. For experimental pCT, inserts of PMP, Delrin and Teflon had errors exceeding 1% (1.08%, 1.16% and 1.31% respectively). In the case of realistically simulated pCT, all three central inserts had an RSP accuracy worse than 1% due to a detector modeling effect which will be discussed later. DECT values were more evenly distributed, and only cortical bone and Teflon had errors larger than 1% (1.17% and 2.38% respectively). These results are also summarized in table 4, in addition to the RSP accuracy obtained from ideal pCT simulations. The RSP MAPE achieved with phase II preclinical prototype scanner was 0.55%. For realistic pCT simulations it was 0.69%, dominated by the central inserts. Without the central inserts, the realistic pCT simulation MAPE was 0.50%, in good agreement with measurements. The RSP accuracy for DECT was 0.67%. In the case of ideal pCT simulations the RSP MAPE was below 0.2%.

The noise (one standard deviation) in a 25 mm diameter circular homogeneous ROI of the LMU phantom (outside of inserts) was 2.1×10^{-2} for experimental pCT and 5.0×10^{-3} for experimental DECT. For the CTP404 phantom, the standard deviations in a similar ROI were 2.6×10^{-2} and 5.1×10^{-3} .

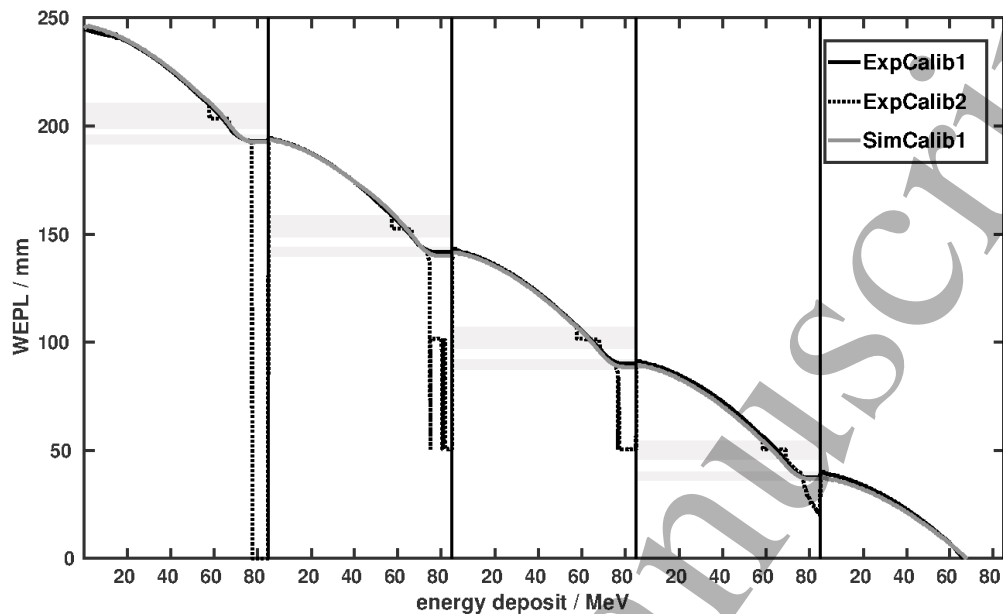


Figure 1: Calibration curves for the experimental (*ExpCalib1* and *ExpCalib2*) and simulated (*SimCalib1*) calibration runs. Vertical lines denote the transition between stages. Shaded areas indicate the WEPL ranges listed in table 3. The shallowest stage is on the left, and the deepest on the right. The step-like features located at approximately 60 MeV for the first four stages are the kink regions. *ExpCalib3* is not shown for clarity but would overlap with *ExpCalib2*, except for the kink regions where it would overlap with *ExpCalib1*.

390 *3.1.3. Proton CT imaging dose* The imaging dose in the pCT simulated scans was approximately 1.5 mGy for all phantoms, with variations less than 0.2 mGy in different inserts. This value was obtained considering that 7.5×10^5 protons were simulated per projection for a total of 2.7×10^8 protons in a scan with 360 projections. Scaling the dose calculated from simulations, as described in section 2.7, we estimated the dose in
 395 the experimental scans to vary from 1.5 mGy to 1.9 mGy, depending on the phantom. This is compatible to the dose measured with an ionization chamber during scans with the pCT phase II prototype scanner and reported by Johnson et al. (2017).

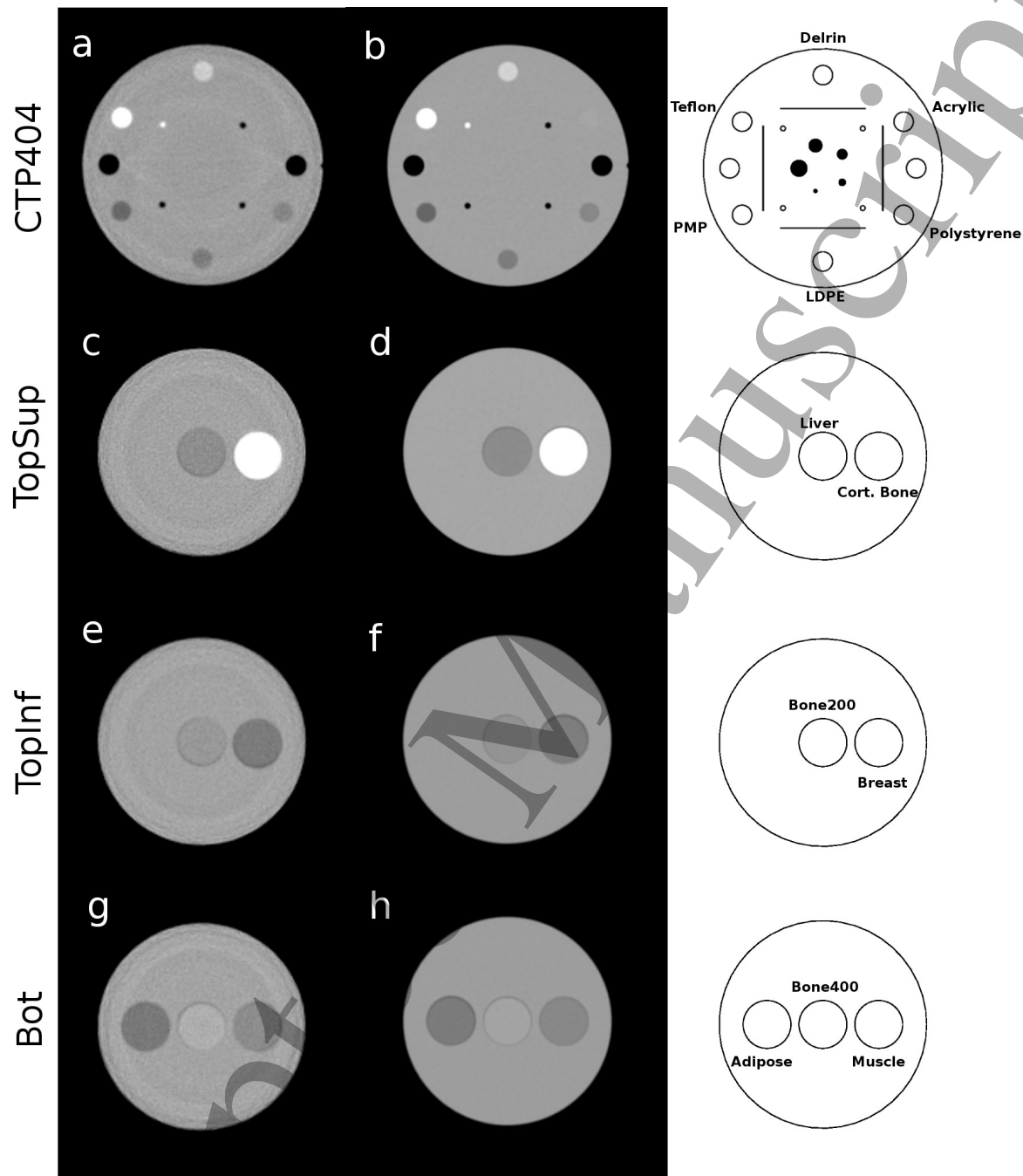


Figure 2: Reconstructed experimental pCT and DECT images. The left column contains images reconstructed from experimental pCT data and the middle column from measured DECT data. An RSP level of 1.0 and window of 1.5 were applied on the images for display purposes. The right column contains drawings of the phantoms with the insert materials labeled.

pCT vs. DECT RSP accuracy

16

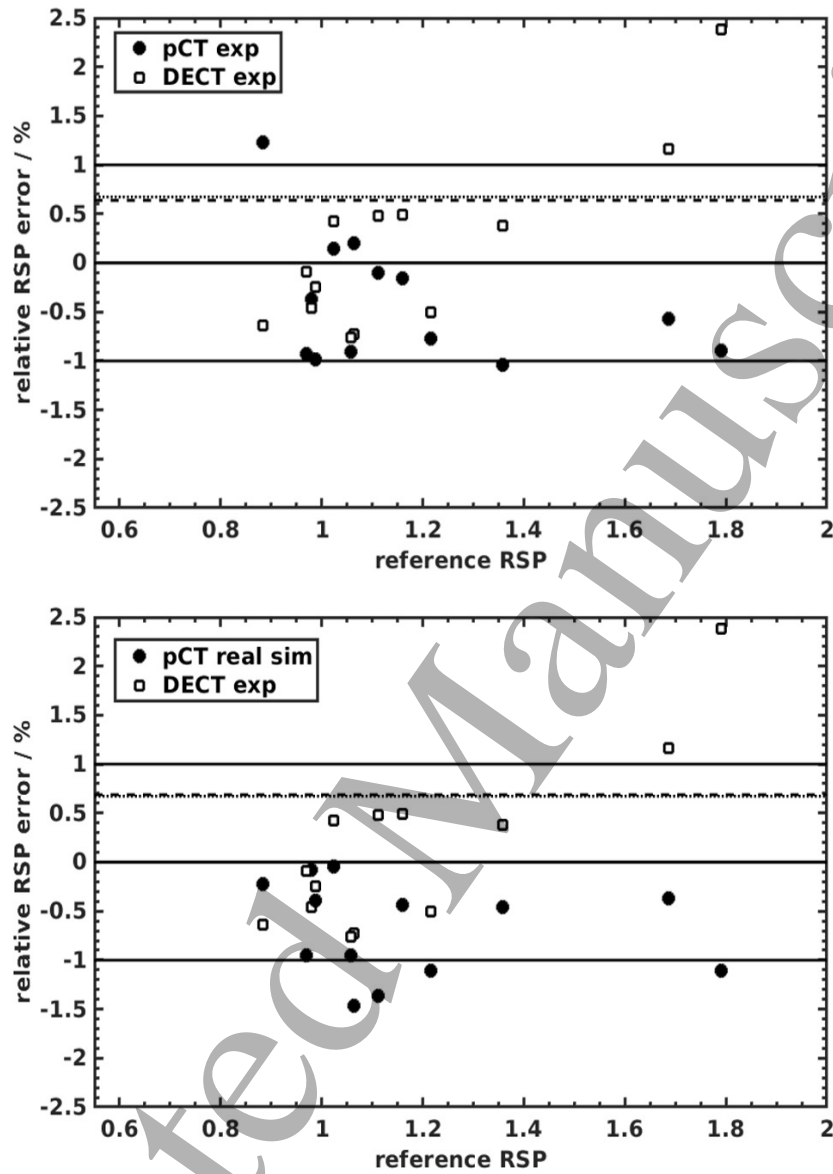


Figure 3: The RSP accuracy as a function of the reference RSP from (top) experimental pCT and (bottom) realistically simulated pCT is indicated with black circles. The RSP accuracy from experimental DECT is indicated with open square markers. The dashed and dotted horizontal lines indicate the MAPE for pCT and DECT, respectively.

pCT vs. DECT RSP accuracy

17

Table 4: RSP accuracy from experimental pCT (pCT_{exp}), realistically simulated pCT ($pCT_{\text{sim}}^{\text{real}}$), ideally simulated pCT ($pCT_{\text{sim}}^{\text{ideal}}$) and experimental DECT scans. The inserts are ordered in increasing reference RSP values. The standard error of the mean was used to express the uncertainty on the estimated RSP accuracy. The ROI size for the CTP404 phantom was 319 pixels for pCT and 985 pixels for DECT. For the LMU phantom, it was 1773 pixels for pCT and 6285 pixels for DECT. The mean absolute percentage error (MAPE) is shown for each simulation mode and imaging modality.

Insert	Phantom	RSP _{ref}	pCT_{exp} %	$pCT_{\text{sim}}^{\text{ideal}}$ %	$pCT_{\text{sim}}^{\text{real}}$ %	DECT %
PMP	CTP404	0.88	1.08 ± 0.11	-0.07 ± 0.09	-0.22 ± 0.11	-0.64 ± 0.02
Adipose	LMU	0.97	-0.14 ± 0.04	-0.36 ± 0.03	-0.95 ± 0.04	-0.09 ± 0.01
LDPE	CTP404	0.98	-0.49 ± 0.11	-0.18 ± 0.08	-0.08 ± 0.10	-0.46 ± 0.02
Breast	LMU	0.99	-0.52 ± 0.04	0.05 ± 0.03	-0.39 ± 0.04	-0.25 ± 0.01
Polystyrene	CTP404	1.02	-0.04 ± 0.10	0.02 ± 0.08	-0.04 ± 0.10	0.43 ± 0.02
Muscle	LMU	1.06	-0.12 ± 0.04	-0.44 ± 0.03	-0.95 ± 0.03	-0.76 ± 0.01
Liver*	LMU	1.06	0.04 ± 0.03	-0.17 ± 0.03	-1.47 ± 0.03	-0.73 ± 0.01
Bone200*	LMU	1.11	-0.41 ± 0.03	-0.14 ± 0.03	-1.36 ± 0.03	0.48 ± 0.01
Acrylic	CTP404	1.16	-0.30 ± 0.10	-0.10 ± 0.07	-0.44 ± 0.09	0.49 ± 0.01
Bone400*	LMU	1.22	-0.84 ± 0.03	-0.44 ± 0.03	-1.11 ± 0.03	-0.50 ± 0.01
Delrin	CTP404	1.36	-1.16 ± 0.09	-0.01 ± 0.07	-0.45 ± 0.09	0.38 ± 0.02
Cort. Bone	LMU	1.69	-0.73 ± 0.02	-0.21 ± 0.02	-0.37 ± 0.02	1.17 ± 0.01
Teflon	CTP404	1.79	-1.31 ± 0.05	-0.06 ± 0.05	-1.11 ± 0.05	2.38 ± 0.01
MAPE %			0.55	0.17	0.69	0.67

* central insert

1
2
3 *pCT vs. DECT RSP accuracy*

18

4
5 *3.2. Simulation investigations of pCT artifacts*

6
7 *3.2.1. Appearance of artifacts in proton CT images* In figure 4 the reconstructed pCT
8 images of the CTP404 and the LMU phantom are shown. A narrow window (see figure
9 caption) was chosen in order to highlight the image artifacts, along with averaging
10 15 slices to reduce noise. Similar ring artifacts were observed in experimental scans
11 (figures 4 a, d, g, j) and realistic simulations (figures 4 b, e, h, k), but not in ideal
12 simulations (figures 4 c, f, i, l). The right column of figure 4 shows a comparison
13 of line profiles through the phantoms for the experimental and realistically simulated
14 pCT scans. In the case of experimental scans the strongest artifacts exceeded 2%
15 in RSP and appeared mostly as RSP overestimation. In the case of the realistic
16 simulations, the strongest artifacts reached up to 2% in RSP and appeared mostly
17 as RSP underestimation. In the CTP404 phantom, the artifacts appeared distorted by
18 the inserts of higher or lower RSP, deviating from the appearance of conventional ring
19 artifacts.
20
21
22
23

24 Nuclear interactions of protons in the scanned object can produce secondary protons
25 or heavier ions, which have mostly lower energy than the primary protons traversing
26 the same material and experience only electromagnetic interactions. These secondary
27 particles, if not efficiently removed by the cuts, will result in an overestimation of the
28 RSP. Switching off nuclear interactions in the simulations did not entirely remove these
29 artifacts (results not shown), confirming our hypothesis that detector and calibration
30 effects are the main source of artifacts.
31
32
33

34
35 *3.2.2. Proton CT water phantom simulations* All images presented in this section were
36 obtained with the realistic pCT simulation of water phantoms detailed in section 2.2
37 and were averaged over 15 slices. In figure 5 simulations of a pCT scan of the 150.5 mm
38 diameter water phantom are shown for different modelling of the tracker gaps and
39 placements of the phantom. This allowed to identify artifacts originating from the
40 tracker. In the current implementation of the tracker geometry in the simulation, these
41 gaps were overestimated by the assumption that they were filled solely with air. Protons
42 traversing a gap experience a slightly lower WEPL (by approximately 0.8 mm). This
43 results in their arrival to the five-stage energy detector with higher energy than nearby
44 protons that lose some additional energy by going through an additional tracker layer.
45 As shown in figure 5 (a) there is lower RSP artifact (dark spot), located at the center
46 of the water phantom when the latter is placed at the isocenter. When the water
47 phantom was laterally displaced from the isocenter, as shown in figure 5 (b), the dark
48 spot remained at the isocenter, not coinciding anymore with the center of the phantom.
49 The dark spot almost disappears from figure 5 (c), when filling the tracker gaps with
50 silicon instead of modelling air.
51
52
53
54

55
56 The images of the 150.5 mm and the 225.75 mm diameter water phantoms are shown
57 in figure 6, where we observed that the location of the ring artifacts varied with the object
58 diameter, an observation consistent with rings originating from specific, problematic
59
60

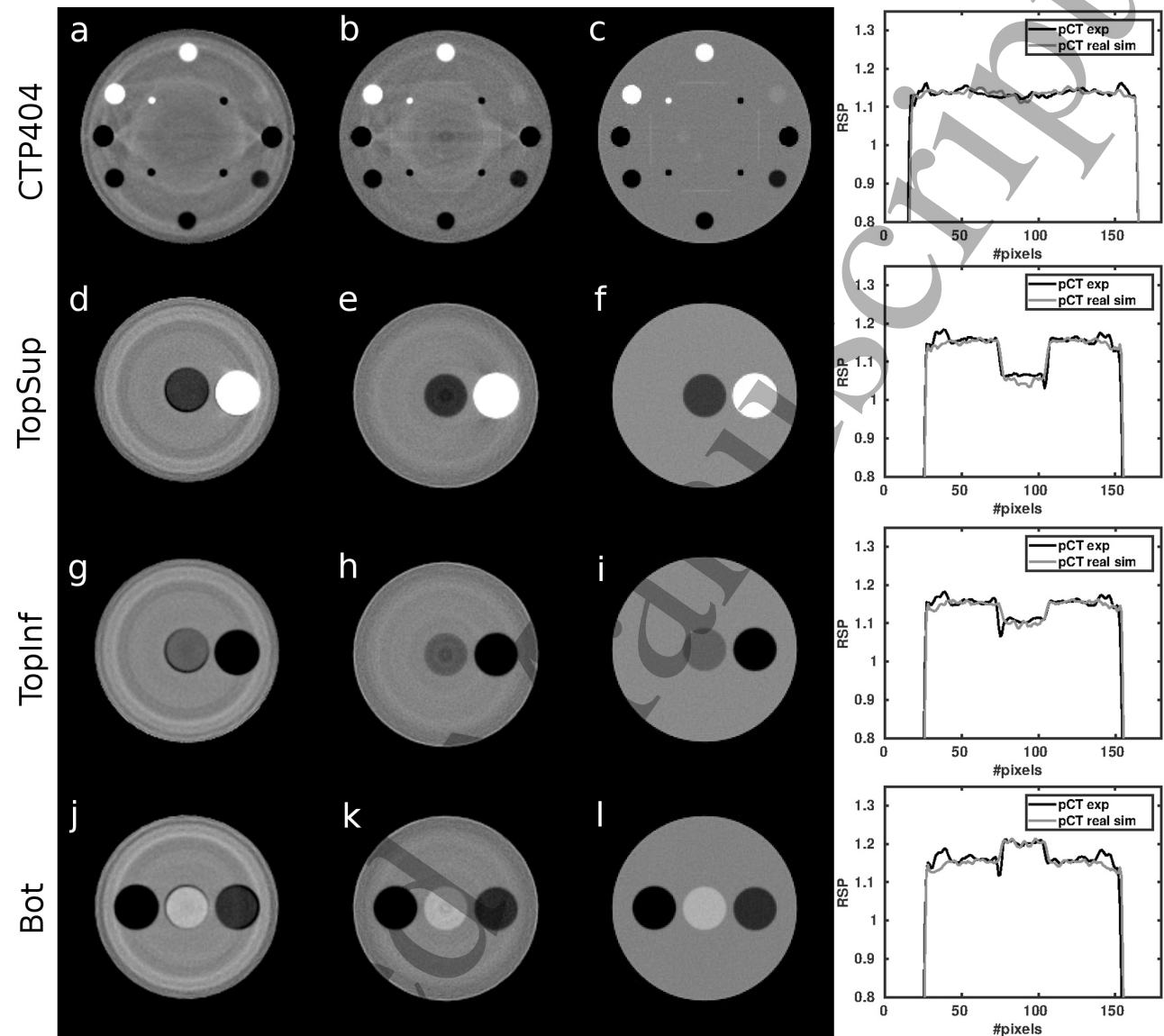
pCT vs. DECT RSP accuracy

Figure 4: Reconstructed pCT images. The first column contains images reconstructed from experimental data, the second column from realistic simulations and the third column from ideal simulations. An RSP level of 1.15 and window of 0.3 were applied on the images in order to highlight the pCT image artifacts. The fourth column contains line profiles for the experimental and realistically simulated pCT images. For the LMU *TopSup*, *TopInf* and *Bot* the profiles were obtained along the vertical diameter. For the LMU *CTP404* phantom the profile was obtained along the diameter that has a -30° angle with respect to the vertical and does not cross any visible insert. For all images and profiles 15 slices were averaged to better display artifacts.

WEPL ranges. Artifacts observed in the water phantom simulations were consistent in terms of amplitude with these from the other phantoms, both in experimental and realistically simulated pCT scans. For example in figures 6 (a) and (b) they reached up to 1.5%.

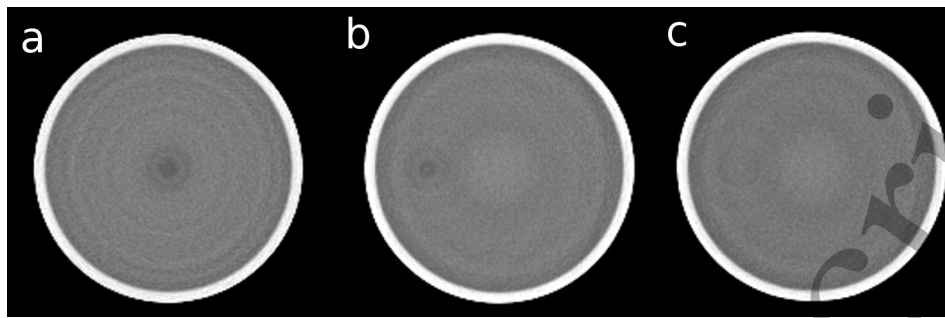


Figure 5: pCT images from realistic simulations processed with the *SimCalib2* calibration. In (a), the water phantom was aligned to the isocenter, while in (b) and (c) it was shifted by 40 mm with respect to the isocenter. In (a) and (b) the tracker layer gaps were filled with air. In (c), the tracker layer gaps were filled with silicon. An RSP level of 1.0 and window of 0.3 were applied on the images for display purposes.

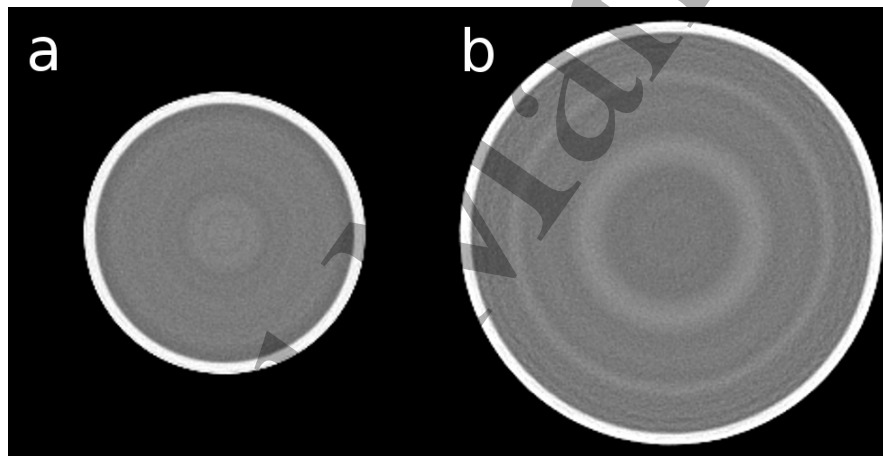


Figure 6: pCT images from realistic simulation processed with the *SimCalib2* calibration and tracker gaps filled with silicon. In (a) the 150.5 mm diameter water phantom and in (b) the 225.75 mm diameter water phantom. An RSP level of 1.0 and window of 0.3 were applied on the images for display purposes.

3.2.3. *WEPL analysis* Following the methods described in section 2.6, we have identified WEPL value ranges (see table 3) that correspond to ambiguities and uncertainties in the calibration due to the four stage interfaces and to the kinks (see figure 1). Figure 7 contains maps displaying for each pixel what fraction of the total number of protons which intersected that pixel had WEPLs within the ranges listed in table 3. The value of 1 in the scale (bright yellow - "hot" regions) indicates pixels in which all protons from all projections had WEPLs in these ranges. As it can be deduced from figures 7 (a-d), areas with high fraction of lower accuracy WEPLs were overlapping with many of the inserts, and are in good qualitative agreement with the artifacts seen in figure 4 and 6.

pCT vs. DECT RSP accuracy

21

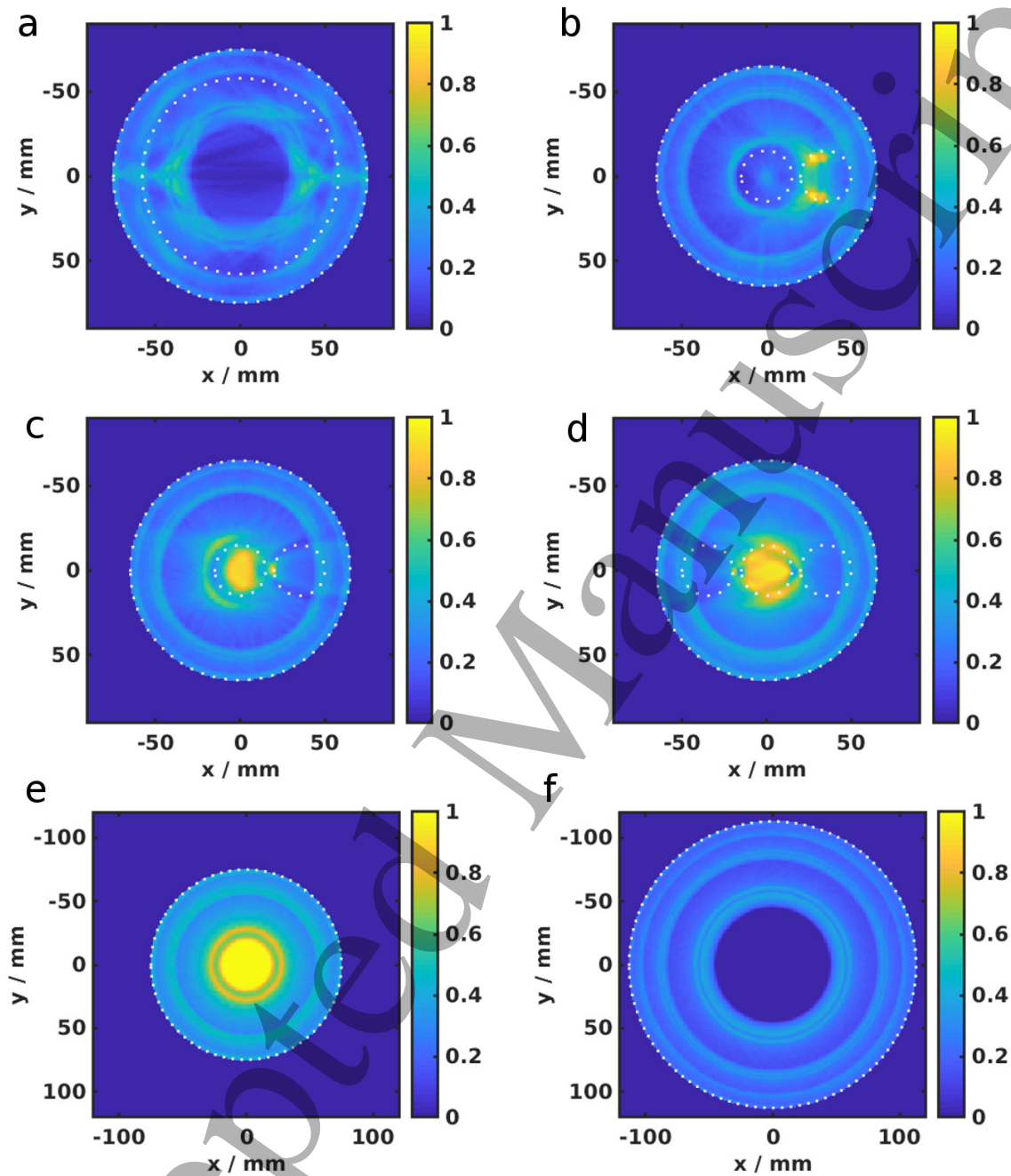


Figure 7: Fraction of the number of protons with certain WEPLs (see table 3) crossing a pixel for (a) CTP404 phantom, (b) LMU *TopSup*, (c) LMU *TopInf*, (d) LMU *Bot*, (e) water phantom 150.5 mm diameter and (f) water phantom 225.75 mm diameter. The outer dotted circles denote the hull of the cylindrical phantoms. The inner dotted circle in (a) indicates the radius at which the centers of the inserts of the CTP404 phantom are located. For (b-f), the inner dotted circles delineate the inserts of the LMU phantom. The images were produced by thresholding experimental pCT images to the nearest reference RSP values.

1
2
3 *pCT vs. DECT RSP accuracy*

22

4. Discussion

4.1. Proton CT – Dual energy CT comparison

4
5
6
7
8
9 The RSP MAPE achieved with an optimized calibration of the pCT phase II preclinical
10
11 455 prototype scanner was below 1%, at a physical dose of 1.5 mGy to 1.9 mGy, as
12 summarized in table 4 and shown as a function of the reference RSP of the inserts in
13 figure 3. In specific, for experimental pCT it was 0.55%. This dose to noise relationship
14 is in agreement to Schulte et al. (2005) and Dickmann et al. (2019). The RSP MAPE
15 accuracy from realistic pCT simulations was slightly worse, 0.69%. This is mainly due
16
17 460 to the exaggerated tracker gaps in the simulation, resulting in high RSP errors for
18 central inserts. When enforcing a smoother connection of calibration curves between
19 adjacent stages, RSP MAPE deteriorated to 0.87%. At the stage interfaces, different
20 effects such as low numbers of protons with very high energy deposits, sharing their
21 energy to adjacent stages, depositing energy to inactive material and imposed minimum
22
23 465 energy threshold can contribute to a higher ambiguity in that region of the calibration.
24 Ignoring the interpolation of the kink region as in *ExpCalib2* led to markedly worse RSP
25 MAPE. The RSP MAPE accuracy achieved with DECT was 0.67%, at an imaging dose
26 of 35.7 mGy. The relatively high DECT imaging dose (about 20 times that of pCT) was
27 used to provide a robust benchmark to compare pCT and resulted in noise in DECT
28
29 470 being 4 to 5 times lower than in pCT. However, we did not aim at reducing the DECT
30 imaging dose, and equivalent accuracy can be expected for lower exposures (Landry
31 et al. 2016). Nevertheless, below a certain threshold noise is expected to impact DECT
32 (Lee et al. 2019). Since the DECT reconstruction uses an iterative algorithm to reduce
33 noise, while pCT uses an analytical reconstruction, it is outside the scope of this work
34
35 475 to further discuss the noise differences. For both imaging modalities the maximum RSP
36 error exceeded 1%. For pCT the maximum RSP error was 1.31% for the Teflon insert
37 of the CTP404 phantom, while for DECT the highest RSP error was 2.38% also for the
38 Teflon insert of the CTP404 phantom and the second highest was 1.17% for the cortical
39 bone insert of the LMU phantom. Excluding the Teflon insert, the RSP MAPE for pCT
40
41 480 and DECT were 0.49% and 0.53%, respectively.

4.2. Proton CT artifacts

42
43
44
45
46
47
48 For experimental and realistically simulated pCT scans, the RSP image contained
49 artifacts whose amplitude in some cases exceeded 2% in RSP. As seen in figure 4 (a, d,
50 g, j) and (b, e, h, k), realistic simulations and experimental scans show both artifacts in
51
52 485 forms of rings and approximately at the same locations. Nevertheless, in experimental
53 pCT scans the artifacts were expressed mostly as RSP overestimation, contrary to what
54 happened in realistically simulated pCT scans. The presence of similar artifacts in
55 realistic simulations and experiment was the result of the detailed modeling of the
56 scanner geometry and detection effects (Birks' effect), as already manifested in the
57
58 490 agreement of the respective calibration curves shown in figure 1.

In cylindrical phantoms, ring artifacts usually appear when the error in the WEPL determination of protons that traverse specific WEPLs is higher than for other protons. For the case of the pCT phase II prototype scanner these can be attributed to the kink and stage interfaces regions. The ring artifacts were mostly expected to occur at specific WEPL value ranges (see table 3). Therefore they appeared at different radii in different phantoms, depending on the radial distance from the center at which protons traversed chord lengths corresponding to the previously mentioned WEPL ranges. This was confirmed by the realistic simulations of a water phantom with two different radii shown in figure 6. In the image of the 225.75 mm diameter phantom, the same ring artifacts are observed as in the 150.5 mm diameter phantom, but at larger radii.

Qualitative spatial maps of lower WEPL accuracy, using the WEPL ranges listed in table 3 were presented in figure 7. The image pixels which are sampled by a large fraction of the protons having low accuracy WEPLs were expected to suffer from lower RSP accuracy. This is indeed the case, as the areas with high fraction in each phantom seem to overlap well with the artifacts shown in figure 4 and 6, confirming our assumptions about the source of the problem. A relevant observation for this study is that several regions of low accuracy WEPLs were located inside the tissue mimicking inserts, thus affecting the achievable RSP accuracy. The strong correlation between the level of that fraction in an insert and the quantified RSP accuracy, can be appreciated by examining the example of the central inserts of the LMU phantom for experimental pCT. In the central insert of the LMU *TopSup* (Liver), the fraction was lower compared to that in the LMU *TopInf* (Bone200), and much lower than that in the LMU *Bot* (Bone400). The reverse trend, as we would expect, was observed in terms of RSP accuracy, which was in experimental pCT 0.04% for Liver, -0.41% for Bone200 and -0.84% for Bone400.

In addition to the above-mentioned artifacts occurring at specific WEPL ranges, other types of artifacts were also observed. The most prominent being a strong lower RSP artifact at the center of the images (dark spot) which was present in all images of realistic pCT simulations and is mostly visible in figures 4 (b, h and k). This artifact was caused by the gaps in each tracker layer, described in section 2.2. When the simulation models air in the tracker gaps, the position-fixed RSP artifact is visible and always located at the isocenter. When the tracker gaps are filled with the same material as the active areas (silicon), the main position-fixed artifact disappears. Less prominent position-fixed artifacts are still visible in the image. It is possible that some minor effect of the tracker gaps is also present in experimental pCT images, but given the fact that these inactive areas are mostly filled with silicon or glue, this effect should be small. The tracker gaps artifact (dark spot) is also responsible for the fact that in the realistically simulated pCT, where the gaps are exaggerated, the correlation between the fraction of protons with low WEPL accuracy and RSP accuracy was not preserved. In contrast to the experimental pCT results, for realistic simulations all three central inserts of the LMU phantom suffered, as expected, from a significant RSP underestimation. The RSP accuracy was -1.47% , -1.36% and -1.11% for Liver, Bone200 and Bone400. An example of the realistic pCT simulation with the tracker gaps filled is shown results in section 2

1
2
3 *pCT vs. DECT RSP accuracy*

24

4 of the *Supplementary Material*.

5
6 As depicted in figure 4 (c, f, i, l), artifacts were not present in pCT images
7
8 535 reconstructed from ideal detector simulations, i.e. from ideal proton energies and
9 positions. Therefore, it was concluded that they were not inherent to the applied FBP
10 reconstruction algorithm. The overall RSP accuracy in ideal pCT simulations was better
11 than 0.2%, on par with past ideal simulation based studies (Hansen et al. 2016), and
12 suggesting that successful artifact mitigation is required to fully exploit the phase II
13 pCT prototype's potential for high accuracy RSP estimation. For three inserts of the
14 540 LMU phantom (Adipose, Muscle and Bone400) the RSP accuracy achieved with ideal
15 pCT simulations was worse (approximately at 0.4%) than for all other inserts. These
16 inserts, with relatively large differences amongst their reference RSP, were located in the
17 same layer of the LMU phantom and moreover were aligned along one line. Therefore,
18 we hypothesized that, in some projections, nuclear interaction and large angle scattering
19 events might not be efficiently filtered from the data with the current cuts. To confirm
20 this hypothesis, an ideal pCT simulation of that phantom, with the nuclear interaction
21 545 physics switched off, was performed. The RSP accuracy of that simulation was below
22 0.1% for all three inserts, showing that for some material and geometry configurations
23 more efficient filtering of nuclear interaction and large angle scattering events might be
24 necessary.
25
26
27
28
29

30 31 **5. Conclusion**

32
33 In this first direct experimental comparison of RSP accuracy between a state-of-the-
34 art DECT scanner and the phase II pCT prototype, we have demonstrated that both
35 555 modalities can currently achieve an RSP accuracy better than 1%. The pCT phase
36 II prototype scanner yielded better RSP MAPE (0.55%) than the commercial DECT
37 scanner (0.67%). We could demonstrate, using a realistic simulation, that characteristic
38 artifacts cause the ideal pCT RSP accuracy of 0.17% MAPE to be degraded to 0.55%.
39 Mitigating these artifacts is thus critical to further improve pCT RSP accuracy.
40
41
42
43

44 45 560 **6. Acknowledgements**

46
47 This work was supported by the Bavaria-California Technology Center (BaCaTeC),
48 by the German Research Foundation (DFG) project #388731804 "Fluence modulated
49 proton computed tomography: a new approach for low-dose image guidance in particle
50 therapy", by the DFG's Cluster of Excellence Munich-Centre for Advanced Photonics
51 565 (MAP) and the DFG's Research Training Group GRK2274 Advanced Medical Physics
52 for Image-Guided Cancer Therapy project B4. We thank Dr. Valentina Giacometti and
53 Dr. Pierluigi Piersimoni for sharing their initial implementation of the pCT simulation
54 platform.
55
56
57
58
59
60

References

- Almeida, I. P., Schyns, L. E. J. R., Vaniqui, A., van der Heyden, B., Dedes, G., Resch, A. F., Kamp, F., Zindler, J. D., Parodi, K., Landry, G. & Verhaegen, F. (2018). Monte Carlo proton dose calculations using a radiotherapy specific dual-energy CT scanner for tissue segmentation and range assessment., *Phys Med Biol* **63**: 115008.
- Bashkirov, V. A., Schulte, R. W., Hurley, R. F., Johnson, R. P., Sadrozinski, H. F.-W., Zatserklyaniy, A., Plautz, T. & Giacometti, V. (2016). Novel scintillation detector design and performance for proton radiography and computed tomography., *Med Phys* **43**: 664–674.
- Berndt (2016). *DECT based tissue segmentation as input to Monte Carlo simulations for proton treatment verification using PET imaging*, Master's thesis, Department of Medical Physics, Ludwig–Maximilians–University of Munich, Garching, Germany.
- Bourque, A. E., Carrier, J.-F. & Bouchard, H. (2014). A stoichiometric calibration method for dual energy computed tomography., *Phys Med Biol* **59**: 2059–2088.
- Bär, E., Lalonde, A., Zhang, R., Jee, K.-W., Yang, K., Sharp, G., Liu, B., Royle, G., Bouchard, H. & Lu, H.-M. (2018). Experimental validation of two dual-energy CT methods for proton therapy using heterogeneous tissue samples., *Med Phys* **45**: 48–59.
- Cormack, A. M. (1963). Representation of a function by its line integrals, with some radiological applications, *J Appl Phys* **34**(9): 2722–2727.
- Coutrakon, G., Bashkirov, V., Hurley, F., Johnson, R., Rykalin, V., Sadrozinski, H. & Schulte, R. (2013). Design and construction of the 1st proton CT scanner, *AIP Conference Proceedings* **1525**, 327 (2013).
- Dickmann, J., Wesp, P., Rädler, M., Rit, S., Pankuch, M., Johnson, R. P., Bashkirov, V. A., Schulte, R. W., Parodi, K., Landry, G. & Dedes, G. (2019). Prediction of image noise contributions in proton computed tomography and comparison to measurements, *Phys Med Biol*, *Accepted manuscript*.
URL: <http://iopscience.iop.org/10.1088/1361-6560/ab2474>
- Esposito, M., Waltham, C., Taylor, J. T., Manger, S., Phoenix, B., Price, T., Poludniowski, G., Green, S., Evans, P. M., Allport, P. P., Manolopoulos, S., Nieto-Camero, J., Symons, J. & Allinson, N. M. (2018). PRaVDA: The first solid-state system for proton computed tomography., *Physica Med* **55**: 149–154.
- Giacometti, V., Bashkirov, V. A., Piersimoni, P., Guatelli, S., Plautz, T. E., Sadrozinski, H. F.-W., Johnson, R. P., Zatserklyaniy, A., Tessonnier, T., Parodi, K., Rosenfeld, A. B. & Schulte, R. W. (2017). Software platform for simulation of a prototype proton CT scanner., *Med Phys* **44**: 1002–1016.
- Han, D., Siebers, J. V. & Williamson, J. F. (2016). A linear, separable two-parameter model for dual energy CT imaging of proton stopping power computation., *Med Phys* **43**: 600.
- Hansen, D. C., Sangild Sørensen, T. & Rit, S. (2016). Fast reconstruction of low dose proton CT by sinogram interpolation., *Phys Med Biol* **61**: 5868–5882.
- Hansen, D. C., Seco, J., Sørensen, T. S., Petersen, J. B. B., Wildberger, J. E., Verhaegen, F. & Landry, G. (2015). A simulation study on proton computed tomography (CT) stopping power accuracy using dual energy CT scans as benchmark, *Acta Oncol* **54**(9): 1638–1642.
- Hudobivnik, N., Schwarz, F., Johnson, T., Agolli, L., Dedes, G., Tessonnier, T., Verhaegen, F., Thieke, C., Belka, C., Sommer, W. H., Parodi, K. & Landry, G. (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.
- Huesman, R., Rosenfeld, A. & Solmitz, F. (1975). Comparison of heavy charged particles and x-rays for axial tomographic scanning, *Technical Report*.
- Hünemohr, N., Krauss, B., Dinkel, J., Gillmann, C., Ackermann, B., Jäkel, O. & Greilich, S. (2013). Ion range estimation by using dual energy computed tomography, *Z Med Phys* **23**(4): 300–313.
- Johnson, R. P., Bashkirov, V., Coutrakon, G., Giacometti, V., Karbasi, P., Karonis, N., Ordoñez,

- C., Pankuch, M., Sadrozinski, H.-W., Schubert, K. et al. (2017). Results from a prototype proton-CT head scanner, *Phys Procedia* **90**: 209–214.
- Johnson, R. P., Bashkirov, V., DeWitt, L., Giacometti, V., Hurley, R. F., Piersimoni, P., Plautz, T. E., Sadrozinski, H. F.-., Schubert, K., Schulte, R., Schultze, B. & Zatserklyaniy, A. (2016). A fast experimental scanner for proton CT: Technical performance and first experience with phantom scans, *IEEE Trans Nucl Sci* **63**(1): 52–60.
- Lalonde, A., Bär, E. & Bouchard, H. (2017). A Bayesian approach to solve proton stopping powers from noisy multi-energy CT data, *Med Phys* **44**(10): 5293–5302.
- Landry, G., Gaudreault, M., van Elmpt, W., Wildberger, J. E. & Verhaegen, F. (2016). Improved dose calculation accuracy for low energy brachytherapy by optimizing dual energy CT imaging protocols for noise reduction using sinogram affirmed iterative reconstruction, *Z Med Phys* **26**(1): 75–87.
- Lee, H. H., Li, B., Duan, X., Zhou, L., Jia, X. & Yang, M. (2019). Systematic analysis of the impact of imaging noise on dual-energy CT-based proton stopping-power-ratio estimation, *Med Phys* **46**: 2251–2263.
- Möhler, C., Wohlfahrt, P., Richter, C. & Greilich, S. (2016). Range prediction for tissue mixtures based on dual-energy CT, *Phys Med Biol* **61**(11): N268.
- Möhler, C., Russ, T., Wohlfahrt, P., Elter, A., Runz, A., Richter, C. & Greilich, S. (2018). Experimental verification of stopping-power prediction from single- and dual-energy computed tomography in biological tissues., *Phys Med Biol* **63**: 025001.
- 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. & Schubert, K. E. (2009). A more accurate reconstruction system matrix for quantitative proton computed tomography., *Med Phys* **36**: 4511–4518.
- Penfold, S. N., Schulte, R. W., Censor, Y. & Rosenfeld, A. B. (2010). Total variation superiorization schemes in proton computed tomography image reconstruction., *Med Phys* **37**: 5887–5895.
- Pettersen, H. E. S., Alme, J., Biegun, A., van den Brink, A., Chaar, M., Fehlker, D., Meric, I., Odland, O. H., Peitzmann, T., Rocco, E., Wang, H., Yang, S., Zhang, C. & Röhrich, D. (2016). Proton tracking in a high-granularity digital tracking calorimeter for proton CT purposes, *Nucl. Instr. Meth. Phys. Res. A* **860** 51–61.
- Piersimoni, P., Ramos-Méndez, J., Geoghegan, T., Bashkirov, V., Schulte, R. & Faddegon, B. (2017). The effect of beam purity and scanner complexity on proton CT accuracy, *Med Phys* **44**(1): 284–298.
- Poludniowski, G., Allinson, N. M. & Evans, P. M. (2014). Proton computed tomography reconstruction using a backprojection-then-filtering approach., *Phys Med Biol* **59**: 7905–7918.
- Rit, S., Dedes, G., Freud, N., Sarrut, D. & Létang, J. M. (2013). Filtered backprojection proton CT reconstruction along most likely paths., *Med Phys* **40**: 031103.
- Sadrozinski, H. F.-W., Geoghegan, T., Harvey, E., Johnson, R. P., Plautz, T. E., Zatserklyaniy, A., Bashkirov, V., Hurley, R. F., Piersimoni, P., Schulte, R. W., Karbasi, P., Schubert, K. E., Schultze, B. & Giacometti, V. (2016). Operation of the preclinical head scanner for proton CT., *Nucl. Instr. Meth. Phys. Res. A* **831**: 394–399.
- Saito, M. & Sagara, S. (2017a). A simple formulation for deriving effective atomic numbers via electron density calibration from dual-energy CT data in the human body, *Med Phys* **44**(6): 2293–2303.
- Saito, M. & Sagara, S. (2017b). Simplified derivation of stopping power ratio in the human body from dual-energy CT data., *Med Phys* **44**: 4179–4187.
- Schneider, U., Pedroni, E. & Lomax, A. (1996). The calibration of CT Hounsfield units for radiotherapy treatment planning., *Phys Med Biol* **41**: 111–124.
- Schulte, R. W., Bashkirov, V., Klock, M. C. L., Li, T., Wroe, A. J., Evseev, I., Williams, D. C. & Satogata, T. (2005). Density resolution of proton computed tomography., *Med Phys* **32**: 1035–1046.

- 670 Schulte, R. W., Penfold, S. N., Tafas, J. T. & Schubert, K. E. (2008). A maximum likelihood proton path formalism for application in proton computed tomography., *Med Phys* **35**: 4849–4856.
- 675 Sigmund, P., Schinner, A. & Paul, H. (2009). Errata and addenda for ICRU report 73, stopping of ions heavier than helium, *J ICRU* **5**(1): 1–10.
- 680 Taasti, V. T., Michalak, G. J., Hansen, D. C., Deisher, A. J., Kruse, J. J., Krauss, B., Muren, L. P., Petersen, J. B. B. & McCollough, C. H. (2017). Validation of proton stopping power ratio estimation based on dual energy CT using fresh tissue samples., *Phys Med Biol* **63**: 015012.
- Taasti, V. T., Petersen, J. B. B., Muren, L. P., Thygesen, J. & Hansen, D. C. (2016). A robust empirical parametrization of proton stopping power using dual energy CT., *Med Phys* **43**: 5547.
- Takada, Y., Kondo, K., Marume, T., Nagayoshi, K., Okada, I. & Takikawa, K. (1988). Proton computed tomography with a 250 MeV pulsed beam, *Nucl. Instr. Meth. Phys. Res. A* **273**(1): 410–422.
- 685 Volz, L., Piersimoni, P., Bashkirov, V. A., Brons, S., Collins-Fekete, C.-A., Johnson, R. P., Schulte, R. W. & Seco, J. (2018). The impact of secondary fragments on the image quality of helium ion imaging., *Phys Med Biol* **63**: 195016.
- 690 Wohlfahrt, P., Möhler, C., Hietschold, V., Menkel, S., Greilich, S., Krause, M., Baumann, M., Enghardt, W. & Richter, C. (2017a). Clinical implementation of dual-energy CT for proton treatment planning on pseudo-monoenergetic CT scans, *Int J Radiat Oncol Biol Phys* **97**(2): 427–434.
- 695 Wohlfahrt, P., Möhler, C., Stützer, K., Greilich, S. & Richter, C. (2017b). Dual-energy CT based proton range prediction in head and pelvic tumor patients, *Radiother Oncol* **125**(3): 526–533.
- Xie, Y., Ainsley, C., Yin, L., Zou, W., McDonough, J., Solberg, T. D., Lin, A. & Teo, B.-K. K. (2018). Ex vivo validation of a stoichiometric dual energy CT proton stopping power ratio calibration., *Phys Med Biol* **63**: 055016.
- 700 Yang, M., Virshup, G., Clayton, J., Zhu, X., Mohan, R. & Dong, L. (2010). Theoretical variance analysis of single-and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues, *Phys Med Biol* **55**(5): 1343.
- 705 Yang, M., Zhu, X. R., Park, P. C., Titt, U., Mohan, R., Virshup, G., Clayton, J. E. & Dong, L. (2012). Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration., *Phys Med Biol* **57**: 4095–4115.