Experimental validation of two dual-energy CT methods for proton therapy using heterogeneous tissue samples

Esther Bär,[1,2] Arthur Lalonde,[3] Rongxiao Zhang,[4] Kyung-Wook Jee,[4] Kai Yang,[4] Gregory Sharp,[4] Bob Liu,[4] Gary Royle,[2] Hugo Bouchard,[3] and Hsiao-Ming Lu[4]

[1] *Acoustics and Ionising Radiation Team, National Physical Laboratory, Hampton Road, Teddington TW11 0LW, United Kingdom*

[2] *Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom**

[3] *Department of Physics, Université de Montréal, 2900 boul. Édouard-Montpetit, Montréal QC H3T 1J4, Canada*

[4] *Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114, USA*

Purpose: The purpose of this work is to evaluate the performance of dual-energy CT (DECT) for determining proton stopping power ratios (SPRs) in an experimental environment and to demonstrate its potential advantages over conventional single-energy CT (SECT) in clinical conditions.

Methods: Water equivalent range (WER) measurements of 12 tissue-equivalent plastic materials and 12 fresh animal tissue samples are performed in a 195 MeV broad proton beam using the dose extinction method. SECT and DECT scans of the samples are performed with a dual-source CT scanner (Siemens SOMATOM Definition Flash). The methods of Schneider *et al.* (1996), Bourque *et al.* (2014) and Lalonde *et al.* (2017) are used to predict proton SPR on SECT and DECT images. From predicted SPR values, the WER of the proton beam through the sample is predicted for SECT and DECT using Monte Carlo simulations and compared to the measured WER.

Results: For homogeneous tissue-equivalent plastic materials, results with DECT are consistent with experimental measurements and show a systematic reduction of SPR uncertainty compared to SECT, with root mean square errors of 1.59% versus 0.61% for SECT and DECT respectively. Measurements with heterogeneous animal samples show a clear reduction of the bias on range predictions in the presence of bones, with -0.88% for SECT versus -0.58% and -0.14% for both DECT methods. An uncertainty budget allows isolating the effect of CT number conversion to SPR and predicts improvements by DECT over SECT consistently with theoretical predictions, with 0.34% and 0.31% for soft tissues and bones in the experimental setup compared to 0.34% and 1.14% with the theoretical method.

Conclusion: The present work uses experimental measurements in a realistic clinical environment to show potential benefits of DECT for proton therapy treatment planning. Our results show clear improvements over SECT in tissue-equivalent plastics materials and animal tissues. Further work towards using Monte Carlo simulations for treatment planning with DECT data and a more detailed investigation of the uncertainties on I-value and limitations on the Bragg additivity rule could potentially further enhance the benefits of this imaging technology for proton therapy.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.12175

Keywords: proton therapy, dual-energy CT, range prediction

I. INTRODUCTION

The advantage of proton beam radiotherapy lies in the favorable dose deposition pattern and the sharp dose fall-off at the end of the proton range, leading to a high dose conformity [1, 2]. To exploit the advantages of protons for tumor treatment, an accurate knowledge of the properties of tissues in the beam path is required to precisely control the position of the Bragg peak. In clinical practice, treatment planning systems (TPS) calculate the range of the beam typically based on a map of proton stopping power ratios (SPRs). Currently, these SPR maps are mainly obtained from a single-energy computed tomography (SECT) scan of the patient.

The conversion from CT numbers (in Hounsfield units, HU) to SPRs is often performed using the stoichiometric calibration method proposed by Schneider *et al.* [3]. The CT scanner is calibrated for each specific X-ray spectrum and CT scan protocol by acquiring the CT numbers of a calibration phantom composed of plastic materials of known composition and density. The CT-number-to-SPR calibration curve is established for each CT scan protocol by applying the calibrated model on a set of human reference tissues [4]. This calibration method was validated experimentally by Schaffner and Pedroni [5]. They performed a SECT scan of different biological samples (liver, muscle, spleen, heart, brain, adipose, kidney, blood) and applied the CT-number-to-SPR conversion to the samples. From the resulting SPR maps, they calculated the water-equivalent range (WER) of a proton beam traveling through the samples and compared it to measured values of the WER. This study reports range errors caused by CT scanning and calibration of 1.1% in soft tissues and 1.8% in bones, taking beam hardening into account (expected 1% without beam hardening artifacts).

The uncertainties arising from CT imaging are, among setup and dose calculation uncertainties, taken into account during treatment planning [6, 7]. To ensure target coverage, a range margin is applied to the target volume, increasing the amount of irradiated healthy tissue.

Dual-energy CT (DECT) was proposed in literature to potentially increase the accuracy of SPR predictions. Yang *et al.* [8] performed a theoretical study comparing SECT and DECT determined SPRs of standard human biological tissues. They proved the theoretical superiority of DECT over SECT, especially when the electron density (ED) and elemental compositions of the investigated tissue vary from reference tissues.

Several formalisms for converting CT numbers from DECT to SPR were proposed in literature. Most of these formalisms are parameter-based and aim at the extraction of ED and effective atomic number (EAN) [9–15]or

*Electronic address: esther.baer@npl.co.uk

photon absorption cross section [16], which are used to estimate the mean excitation energy (I-value) or the stopping number, respectively. Alternative methods parameterize the I-value [17] or the SPR [18] directly, while other groups focus on calibrating pseudo-monoenergetic images derived from DECT to predict the SPR of tissues [19]. Another class of formalisms aims at extracting elemental compositions as inputs for Monte Carlo simulations [20–23]. In a previous study, we performed a theoretical study on the potential of DECT to reduce proton range uncertainties [24]. For a virtual phantom filled with human reference tissues, we found that DECT can reduce beam range uncertainties by about 0.4% in soft tissues and up to 1 mm (as reported for a 5 cm thick slab) for therapeutic energies in bones. Additional studies [25, 26] performed treatment planning on phantom and patient data sets, comparing treatment plans calculated using SECT-derived and DECT-derived SPRs. Zhu and Penfold [26] demonstrate a dose difference of up to 8% between an SECT-based and a DECT-based plan. Hudobivnik *et al.* [25] evaluated range differences between SECT-based and DECT-based plans for five head trauma patients, concluding group median relative range differences of -1.4%.

The presented study aims at validating DECT for estimating SPR values in biological tissues. We utilize real animal tissue samples consisting of single organs and a variety of different animal bones to measure the WER of the samples in a proton beam using the dose extinction method proposed by Zhang *et al.* [27]. The measured WER is then compared to a calculated WER predicted from SECT and two different DECT formalisms. These formalisms include one ED-EAN method (Bourque *et al.*) and one eigentissue decomposition method (Lalonde *et al.*).

II. MATERIALS AND METHODS

A. Sample preparation

In this study, 12 types of tissue samples are used. The samples are collected fresh from the butcher's on the day of the experiment and fitted into plastic containers (5 cm x 5 cm x 15 cm). The remaining space in the container is filled up with saline water, with a concentration of 9 g NaCl per liter of water. The amount of saline water that is added per container varies between 50 mL and 100 mL, depending on the sample size. Fig. 1 shows a picture of the samples in the container, and a list of all samples can be found in tab. I.

We measure the WER of the real tissues using the dose extinction method, as described in the paper by Zhang *et al.* [27], and briefly illustrated here. The water-equivalent range measurements are performed in a double-scattering beam line with gantry (IBA International, Louvain-La-Neuve, Belgium), at a gantry angle of 0°. A Matrixx Evolution ion chamber array (IBA Dosimetry, Bartlett, TN, USA) with 1020 ionization chambers (chamber volume 0.08 cm^3, diameter 4.5 mm, 7.62 mm center to center distance, $24.4 \times 24.4\,cm^2$ active area) is positioned on the treatment table. For measurement of the WER, we operate the Matrixx detector in movie mode, collecting 1 frame per second, in this way collecting a dose profile as a function of time. The plastic containers with the biological samples are positioned on top of the detector array. A water filled tank is placed on top of the samples. The setup is illustrated in fig. 2.

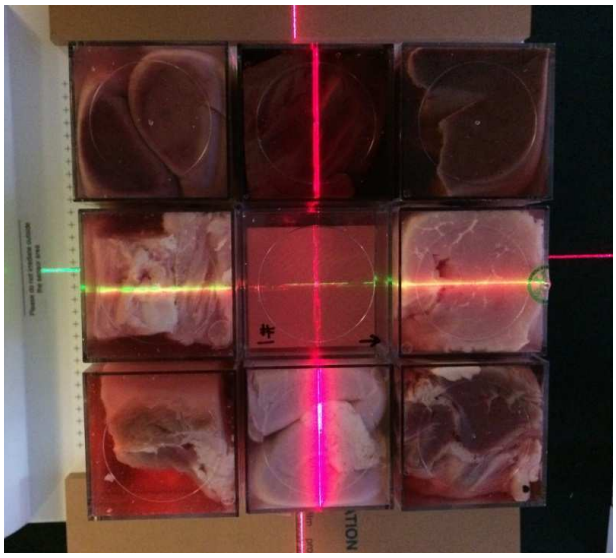

FIG. 1: Animal organs individually packed into plastic containers, view from the top along the beam direction. After packing the organs into containers with saline water, the samples are placed in the beam line for measurements of the WER.

TABLE I: List of animal tissues investigated in this study with the statistics of their ED and EAN values extracted using Bourque et al., as well as their low- and high-kVp CT numbers. The means and standard deviations are calculated over all voxels of an ROI of 308 mm^3.The ROI is placed into a single slice within the samples and is chosen to represent ED, EAN and HU of the respective tissues.

		ED		EAN		HU_l		HU_h	
	Tissue type	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	Pig stomach	1.039	0.018	7.33	0.61	43.7	10.0	42.6	11.4
2	Pig blood	1.043	0.016	7.33	0.55	48.0	7.9	47.0	9.3
3	Pig muscle	1.056	0.019	7.41	0.61	64.0	14.1	61.4	13.1
4	Cow muscle	1.057	0.020	7.30	0.61	60.8	14.1	60.6	15.1
5	Veal brain	1.036	0.019	7.37	0.69	42.5	11.8	41.2	11.6
6	Pig kidney	1.027	0.028	7.16	0.70	26.2	32.2	29.5	26.3
7	Pig liver	1.053	0.017	7.43	0.62	62.0	9.2	59.0	10.1
8	Pig leg	1.452	0.409	10.82	2.95	853.5	758.2	616.2	532.7
9	Cow tail	1.185	0.167	9.45	2.22	341.5	344.5	257.8	243.3
10	Pig rib	1.227	0.168	10.94	1.70	486.1	335.2	345.4	240.6
11	Pig vertebra	1.296	0.079	11.81	0.57	614.2	140.1	443.2	103.8
12	Pig scapula	1.262	0.314	10.14	2.68	526.6	591.7	375.8	425.0

B. Measurement of the WER using the dose extinction method

We continuously irradiate the setup with a broad proton beam (circular field with a diameter of 25 cm). The energy of the beam is 195 MeV, corresponding to a beam range of 25 cm in water, with a modulation of 20 cm. During irradiation, we slowly drain the water from the tank, thus increasing the range of the beam in the samples. The tank and the pump are calibrated to reduce the water level by 0.4 mm per second.

The dose measured as a function of time is converted into dose as a function of water height in the column. For every detector pixel of interest, we find the water height corresponding to a dose fall off of 80% of the maximum dose.

From this water height H_{80}, we calculate the WER of the samples as

$$\mathrm{WER}_{\mathrm{exp}} = R_{80} - H_{80} - R_{\mathrm{offset}} \tag{1}$$

with R_{80} the beam range as measured in the dose extinction setup, at the 80% dose fall off behind the Bragg peak. $\mathrm{R}_{\mathrm{offset}}$ includes the water tank bottom and the Matrixx build up material, and is measured independently for this setup using a pinpoint chamber (PinPoint TN31006, PTW Freiburg, Germany). The value of $\mathrm{R}_{\mathrm{offset}}$ is 13.85 mm, and the uncertainty is determined in repeated measurements as 0.35% (68th percentile), leading to a negligible impact on WER measurements (i.e., less than 0.05 mm). Additionally, we account for non-uniformity of the beam across the field by applying a map of correction factors, as described in Zhang *et al.*. We find that the beam range heterogeneity is around 1.5 mm across the field. The uncertainty of the resulting correction factor is estimated to be 0.2% with repeated measurements, leading to a negligible impact on WER measurements (i.e., less than 0.01 mm). Accounting for the chamber response reproducibility (±0.20%), the overall uncertainty of WER measurements with the dose extinction method is estimated to be 0.20% (68th percentile).

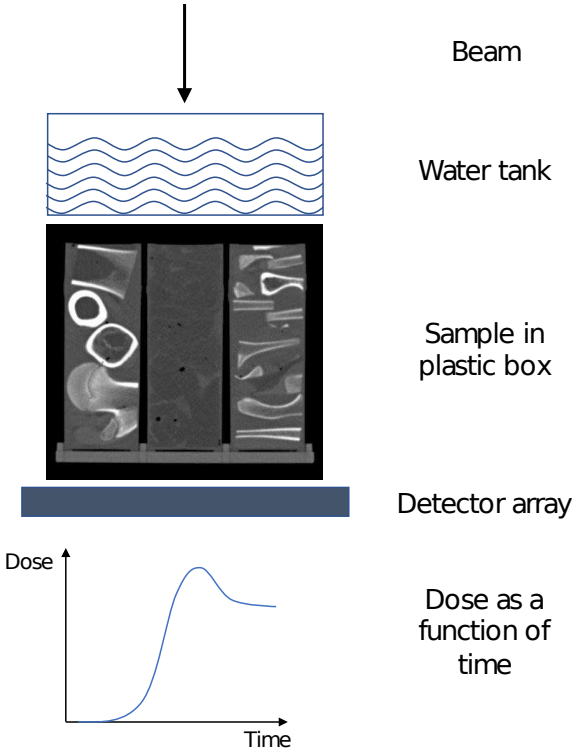

FIG. 2: Illustration of the measurement setup for the dose extinction method. The range of the broad beam is varied by continuously decreasing the water level in the tank. The beam passes the samples and the dose behind the samples is recorded by the Matrixx detector, operated in movie mode. This way we measure the dose as a function of time. A CT slice of the samples is shown here (Window: 1500 HU; Level: 300 HU).

C. Estimation of the WER using computed tomography

1. CT acquisition

To perform SECT and DECT scans of the tissues, a Siemens Somatom Flash (Siemens Sector Healthcare, Forchheim, Germany) dual-source CT scanner is used. The scan parameters for the DECT and SECT scans are summarized in table II. To estimate SPRs from CT numbers, a calibration procedure must be performed. Therefore, a tissue characterization phantom Gammex RMI-467 (Gammex, Inc., Middleton, WI, USA) is scanned using the quoted SECT and DECT protocols. The phantom disk has a diameter of 33 cm, the insert diameter is 2.8 cm and the insert height is 7 cm. To perform the calibration, the CT numbers of the tissue-equivalent inserts are measured in circular ROIs (226 mm^3) and averaged over 20 slices within the phantom.

Both protocols (SECT and DECT) are chosen to represent clinically applicable protocols. As a result, the DECT protocol has a higher $CTDI_{Vol}$, indicating a higher dose to the patient. We decide not to adapt SECT scan parameters to match the $CTDI_{Vol}$ of the DECT scan. Although the image quality of the SECT in terms of noise would improve, we have shown in an earlier study [24] that the SPR prediction based on Schneider *et al.* is fairly robust to noise and an improvement in image quality would only have a minimal effect on the predicted SPR.

TABLE II: List of scan parameters for the SECT and DECT scans. The image noise (1σ) was measured in lucite.

Parameter	SECT	DECT low	DECT high
Tube voltage	140 kV	100 kV	140 kV Sn
Exposure	300 mAs	300 mAs	232 mAs
Slice thickness	1 mm	1 mm	1 mm
Pixel width	0.46 mm	0.46 mm	0.46 mm
$CTDI_{Vol}$	29.51 mGy	23.29 mGy	23.29 mGy
Image noise (ΔHU)	9.1 HU	13.4 HU	13.0 HU
Reconstruction Kernel	B30f	B30f	B30f
CTDI phantom size	32 cm	32 cm	32 cm

2. CT-number-to-SPR conversion

The SECT calibration is performed using the stoichiometric method proposed by Schneider *et al.* [3]. The CT numbers of the tissue-equivalent inserts are used to find the energy fit parameters (K^{ph}, K^{coh}, K^{incoh}). With the found parameters it is possible to calculate the CT numbers of a set of human reference tissues [4]. Theoretical SPR values of these tissues are calculated from composition data using the Bethe-Bloch formula for an energy of 195 MeV. I-values of the tissues are calculated from elemental mass fractions of the tissues using the Bragg additivity rule. The elemental I-values are taken from ICRU Report 37 [28]. The calibration curve is then obtained by performing piecewise linear fits for three tissue regions (Lung: -1000 HU to -40 HU, soft tissues: -40 HU to 150 HU, bony tissues: 150 HU to 2000 HU). The calibration curve is applied voxel wise to obtain a map of SPR values.

The DECT data is processed using two different calibration methods. In a previous study [24], we evaluated the theoretical accuracy of both methods. The method by Bourque *et al.* estimates the EAN from DECT images using a parametric fit to the dual-energy index (DEI) of tissue substitutes. From the EAN estimation, we can solve the system for the ED and estimate the I-value; both quantities are necessary to estimate the SPR using the Bethe-Bloch equation.

In the same study [24], we tested the eigentissue decomposition method by Lalonde and Bouchard [22], which predicts elemental compositions and electron densities of the tissues. The SPR values are calculated from elemental compositions using the Bragg additivity rule as well as the Bethe-Bloch formula. In this study, we investigate an adaptation of this method especially made for SPR prediction from noisy multi-energy CT data [23].

To calibrate the method by Bourque *et al.*, we use the CT numbers measured in the tissue-equivalent inserts to find the fit parameters b_m, as described in [13], eq. 28. From the DEI of the tissue substitutes, we find the c_k values by fitting the DEI to theoretically calculated values for the EAN, as described in [13], eq. 35. Bourque *et al.* suggest a polynomial order of 5 for the DEI-EAN fit. When single pixels are considered, it is possible that the observed DEI is outside the calibration domain due to noise. To be more robust to noise, we add two virtual materials with $Z = 3$ and $Z = 20$ to the DEI-EAN calibration. We find the DEI values of these two virtual materials by linear extrapolation from the original DEI-EAN fit. This strategy ensures the behavior of the calibration curve outside the calibration domain.

For the method of Lalonde *et al.* [23], we use the measured CT numbers of the phantom to calibrate the Z-space coefficients, as proposed in [22], eq. 24. Once the Z-space coefficients are calibrated, it is possible to decompose CT data with optimal materials following a method called Bayesian eigentissue decomposition, or Bayesian ETD. Theses optimal materials, called eigentissues, are principal components of human tissue elemental compositions found in literature. By using a prior function that penalizes solutions with unlikely eigentissue fractions, Bayesian ETD extracts the maximum *a posteriori* composition in each voxel. The advantage of ETD over most DECT formalisms is the direct voxelwise estimation of elemental compositions of tissues, without performing an ED-EAN fit first or applying tissue segmentation. A suggested future application is the use of these elemental compositions for accurate Monte Carlo treatment planning. However, this study aims at the validation of SPR inputs for current treatment planning systems rather than the validation of MC input parameter, which can be the subject of future studies.

The here applied CT-number-to-SPR conversion methods are designed to predict the SPR values in human tissues. The containers however are made out of plastic. We ensure that the use of these methods on plastic HUs does not introduce a bias to the predicted WER. Therefore, we measure the WER of the container top and bottom in a water tank using a pinpoint chamber. The resulting WER (3.5 mm) is compared to each CT predicted WER of the container

walls, and the difference is added to or subtracted from the CT predicted WER. We follow the same procedure with the trays holding the containers in place. We obtain three different maps for the WER: based on SECT images ($\mathrm{WER_{SECT}}$) and based on DECT images ($\mathrm{WER_{Bourque}}$ and $\mathrm{WER_{Lalonde}}$). In figure 3, examples of the resulting 2D maps are shown for SECT, one DECT method and the measured WER map from dose extinction.

3. Monte Carlo simulation

To take into account the inevitable beam degradation due to lateral inhomogeneities on WER calculations, Monte Carlo (MC) simulations are performed using TOPAS (TOol forPArticle Simulation) version 3.1. Because MC calculations cannot be performed directly on the SPR maps, results obtained by each method are converted to MC inputs in a two steps procedure. First, the built-in CT number to MC inputs tool of TOPAS based on the method from Schneider *et al.* [29] is used to construct a reference CT number to SPR look-up table. Then, the inverse of this look-up table is used to convert SPR volumes predicted by each method into synthetic input CT geometries. This way, we ensure that the SPR distributions seen by TOPAS correspond to the ones originally predicted by each method. The WER associated to each SPR volume is calculated using a mono-energetic proton beam of 195 MeV, with a total number of 1.2×10^7 histories. The mean energy of protons exiting the animal samples is scored in a surface detector placed at the effective point of measurement of the Matrixx detector. Conversion from proton energy to WER is done using the PSTAR database [30]. The co-registration between the MC detector and the Matrixx is performed using metallic markers routinely used in image based treatment planning simulation, that are attached to the sample containers before the CT scan. Before each WER measurement, a radiograph of the samples and the Matrixx detector positioned on the treatment table is taken. On this radiograph, it is possible to identify the markers and thus calculate their position relative to the center of the Matrixx. This information is used to place the surface detector in the MC simulation.

4. Statistical analysis

For the purpose of clarifying our WER analysis, we define four statistical quantities and specify the formulas used to calculate them. For WER values obtained experimentally and predicted with a specific CT-based method, we define the WER error as follows:

$$\Delta\mathrm{WER} \equiv \frac{\mathrm{WER_{CT}} - \mathrm{WER_{exp}}}{\mathrm{WER_{exp}}}, \tag{2}$$

where $\mathrm{WER_{CT}}$ is the CT-based WER value (e.g., $\mathrm{WER_{SECT}}$, $\mathrm{WER_{Bourque}}$ or $\mathrm{WER_{Lalonde}}$) and $\mathrm{WER_{exp}}$ is the WER measurement with the dose extinction method as described in equation 1. ΔWER is determined for the number N of detector elements. From the statistical sample of N ΔWER values, we calculate the following quantities:

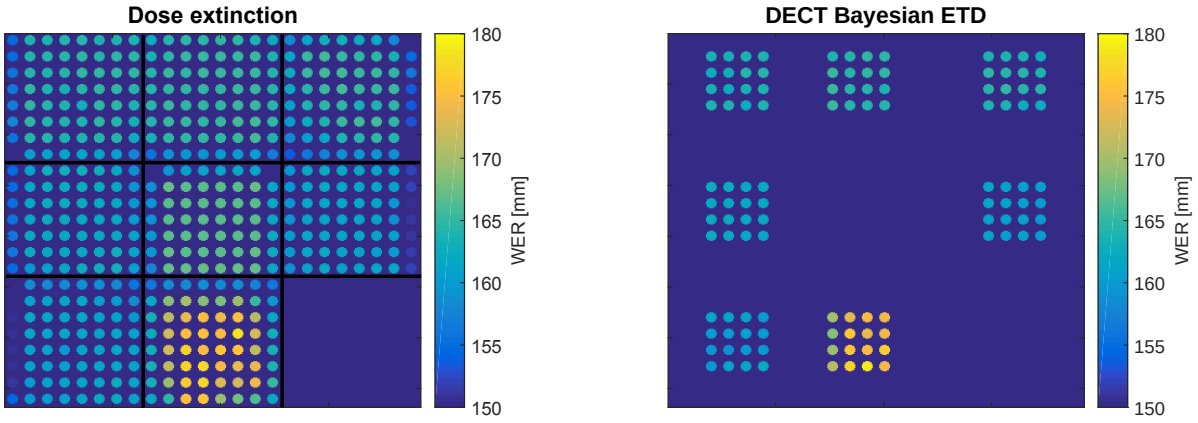

FIG. 3: Examples of the 2D WER maps that are compared in this study. On the left are the measured WER maps using the dose extinction method; on the right are the WER maps obtained with Monte Carlo from the ETD method. The image on the right indicated which chambers are used for analysis. The containers in the displayed images are filled with (from top to bottom, left to right): Pig liver, pig muscle, cow muscle, pig blood, a Lucite block, pig stomach, pig kidney, pig vertebra, pig leg bone. The pig leg bone measurement of this particular sample is not included in the analysis since it did not range out during the measurement. The measurement was repeated with less bone in the beam path.

1. The *mean WER error*, defined as

$$\overline{\Delta\mathrm{WER}} \equiv \frac{1}{N}\sum_{i=1}^{N}\Delta\mathrm{WER}_i. \tag{3}$$

For each CT-based method, this quantity corresponds to the bias in WER prediction. We can show that the expectation value of the WER error is $\mathrm{E}(\overline{\Delta\mathrm{WER}}) = \mu$ (the true mean).

2. The *standard deviation of the WER error*, defined as

$$s_{\Delta\mathrm{WER}} \equiv \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N}\Delta\mathrm{WER}_i^2 - \overline{\Delta\mathrm{WER}}^2\right]}. \tag{4}$$

For each CT-based method, this quantity represents the distribution spread of WER errors with respect to its mean. We can show that the expectation value of the estimator is $\mathrm{E}(s^2_{\Delta\mathrm{WER}}) = \sigma^2$ (the true variance).

3. The *root mean square error* (RMS error), defined as

$$\mathrm{RMS}_{\Delta\mathrm{WER}} \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\Delta\mathrm{WER}_i^2}. \tag{5}$$

We can show that the expectation value of the RMS error squared is $\mathrm{E}(\mathrm{RMS}^2_{\Delta\mathrm{WER}}) = \sigma^2 + \mu^2$. The square root of this quantity corresponds approximately to the limit of the symmetric interval (i.e., $[-\sqrt{\sigma^2+\mu^2}, \sqrt{\sigma^2+\mu^2}]$) containing 68% of a Gaussian distribution with mean μ and variance σ^2. Therefore, this quantity is interpreted as the overall uncertainty, i.e. $u \equiv \sqrt{\sigma^2+\mu^2}$, with a statistical significance of k=1.

We can show quite trivially that the variance of the mean error is the following quadratic sum

$$\sigma^2 = \sigma^2_{\mathrm{WER}} + \sigma^2_{\mathrm{exp}} \tag{6}$$

where σ_{WER} and σ_{exp} are the standard deviations (i.e., type A uncertainties) of the WER estimated with the overall CT-based method and the dose extinction method, respectively. We can also show that for each CT-based method, the expectation value of the RMS error squared is the quadratic sum of the type A uncertainty in predicting WER with the CT method (σ_{WER}), the bias of the result (μ) and the type A uncertainty in predicting WER experimentally with the dose extinction method (σ_{exp}). That is,

$$\mathrm{E}(\mathrm{RMS}^2_{\Delta\mathrm{WER}}) = \sigma^2_{\mathrm{WER}} + \mu^2 + \sigma^2_{\mathrm{exp}}. \tag{7}$$

In this relation, the value of σ_{exp} is set to 0.20% and the assumption is made that the dose extinction method is unbiased. This assumption is based on results found during the validation of the workflow using tissue substitutes, see figure 5. In equation 3 we define the mean WER error as a relative quantity, we can show that the expectation value of the average error corresponds to the ratio of the expectation of the range difference over the expected range. This approximation is valid since the probability distribution of the range difference can be assumed narrow with respect to the average range and also far from the singularity of the denominator being zero. This leads to our following definition of uncertainty of the WER estimation (with statistical significance of k=1):

$$u_{\mathrm{WER}} = \sqrt{\mathrm{RMS}^2_{\Delta\mathrm{WER}} - \sigma^2_{\mathrm{exp}}} \tag{8}$$

which is an unbiased estimator of the quadratic sum of the WER standard deviation from CT only and the mean WER error, that is the expectation value of the estimator is

$$\mathrm{E}(u^2_{\mathrm{WER}}) = \sigma^2_{\mathrm{WER}} + \mu^2. \tag{9}$$

The uncertainty of WER estimation from CT only u_{WER} includes the uncertainties from range degradation due

to multiple Coulomb scattering u_{deg}. Also included into u_{WER} are the uncertainties from the I-value u_I and the uncertainty from the CT calibration an conversion u_{CT}. Hence we can define

$$u^2_{\mathrm{WER}} = u^2_{\mathrm{deg}} + u^2_I + u^2_{\mathrm{CT}}. \tag{10}$$

The contribution from range degradation u_{deg} for soft tissues is assumed to be negligible since the samples are fairly homogenous. For the bones, we assume an uncertainty of 0.14%, based on the value reported by Paganetti (2012).

The contribution of the I-value uncertainty u_I to range uncertainty was estimated previously [6, 31–33]. In this study, we expect to observe a combined uncertainty of the contributions from I-value and CT calibration and conversion, defined as

$$u^2_{\mathrm{comb}} = u^2_I + u^2_{\mathrm{CT}}. \tag{11}$$

D. Validation of the workflow using Gammex RMI-467 tissue substitutes

Our implementation of the dose extinction method is validated using tissue substitutes from a Gammex RMI-467 phantom. The reference SPR values are measured using a pinpoint ionization chamber (PinPoint TN31006, PTW Freiburg, Germany) in a broad beam (25 cm range, 20 cm modulation). We use 12 tissue-equivalent inserts of the Gammex RMI-467 phantom, as listed in table III. To measure the WER of the Gammex inserts with the dose extinction method, we remove the inserts from the phantom disk and place them on the Matrixx. In this work, we additionally validate our SECT and DECT calibrations using the SPR values from the pinpoint chamber. After calibrating SECT and DECT with the Gammex RMI-467 phantom, we apply the calibration to the phantom images and measure the SPR throughout the phantom inserts. The SECT calibration curves for tissues and substitutes as well as the curves to convert the effective atomic number into I-values are displayed in figure 4. It should be noted that the SECT calibration curve for this validation study is not established using the method by Schneider *et al.* (1996). To establish a calibration curve, we apply a piecewise linear fit between measured CT numbers in the Gammex RMI-467 phantom inserts and measured reference SPR. This curve does not represent human tissues, but is suitable for the use in tissue-equivalent plastics. The displayed results for SECT hence represent the errors coming from the fit procedure itself. Similarly, we proceed with the DECT method by Bourque *et al.*. This method uses a polynomial fit to relate the EAN to I-values based on the idea of Yang *et al.* [8]. The differences between the approaches are the definitions of EAN and the use of continuous function instead of a piecewise fit. In their paper, Yang *et al.* suggest to perform the fit based on theoretical I-values calculated from Woodard and White [4] composition data, representing human tissues. Hence, to make this method suitable for the use in tissue substitute materials, we perform the fit based on

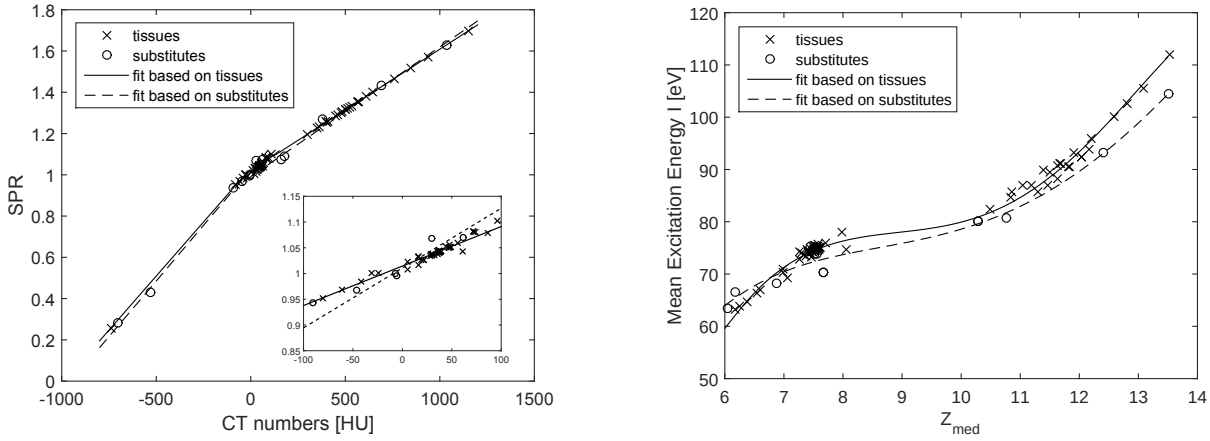

FIG. 4: Calibration technique on tissue-equivalent plastics: (a) calibration curves for the SECT methods are shown, where the solid line represents the fit for the human reference tissues and the dotted line represents the fit for tissue substitutes; the soft tissue region is shown as inlet; (b) the curves to convert effective atomic number into I-values for the Bourque method, where the solid line represents the fit for the human reference tissues and the dotted line represents the fit for tissue substitutes.

I-values calculated from compositions of the substitutes. We do not adapt the method of Lalonde and Bouchard to plastic materials as the whole method is designed to describe human tissues.

TABLE III: List of tissue-equivalent materials used for validation of our measurement techniques. Material-specific SPR values are given for: i) a reference measurement with a pinpoint chamber, ii) the dose extinction measurement, ii) the SECT calibration and iv) the DECT calibration by Bourque *et al.*. The uncertainty of the reference SPRs values is estimated from repeated measurements to be 0.20% and the uncertainty of SPR values measured with the dose extinction method is estimated to be 0.20%. Uncertainties of SECT and DECT are estimated by calculating the standard deviation of the mean over an ROI of 243 mm^3.

	Tissue-equivalent material	SPR Reference	SPR dose extinction	SPR SECT	SPR DECT
1	Water	1.000	1.000	1.006 (0.001)	0.998 (0.002)
2	LN300 Lung	0.286 (0.001)	0.286 (0.001)	0.287 (0.001)	0.275 (0.006)
3	AP6 Adipose	0.951 (0.002)	0.950 (0.002)	0.941 (0.001)	0.943 (0.002)
4	BR12 Breast	0.978 (0.002)	0.977 (0.002)	0.958 (0.001)	0.971 (0.002)
5	Solid Water	1.004 (0.002)	1.005 (0.002)	1.007 (0.001)	0.996 (0.002)
6	LV1 Liver	1.078 (0.002)	1.080 (0.002)	1.088 (0.001)	1.080 (0.002)
7	SR2 Brain	1.069 (0.002)	1.068 (0.002)	1.050 (0.001)	1.063 (0.002)
8	CB2 - 30% $CaCO_3$	1.267 (0.003)	1.266 (0.003)	1.232 (0.001)	1.262 (0.003)
9	CB2 - 50% $CaCO_3$	1.431 (0.003)	1.428 (0.003)	1.421 (0.001)	1.435 (0.002)
10	SB3 Cortical Bone	1.621 (0.003)	1.614 (0.004)	1.636 (0.001)	1.628 (0.002)
11	B200 Mineral Bone	1.095 (0.002)	1.100 (0.002)	1.106 (0.001)	1.099 (0.002)
12	IB3 Inner Bone	1.076 (0.002)	1.076 (0.002)	1.096 (0.001)	1.073 (0.003)

III. RESULTS

A. Validation of the workflow using Gammex RMI-467 tissue substitutes

In the validation measurement, we show good agreement between SPRs measured by dose extinction and our reference SPR with a root mean square (RMS) error of 0.29%. We consider this value as part of the overall uncertainty of the experimental method to determine WER. The resulting differences in SPR values determined with dose extinction, SECT and DECT are summarized in fig. 5. Table IV summarizes the mean error and standard deviation as well

as RMS errors on SPR for each investigated method. Our results clearly demonstrate the limitations of the use of a standard SECT calibration curve in agreement with earlier studies [26, 34]. We report an RMS error on SPR of 1.59% between SECT-determined values and reference SPR. For DECT, we report RMS errors of 0.61%. This value is in accordance with the original values published in Bourque *et al.*, who report an RMS error of 0.67% (for the same energy couple) on experimentally determined SPR values.

TABLE IV: Mean error, standard deviation and RMS error in % between reference SPR values (pinpoint chamber), dose extinction measured SPR values and SECT and DECT predicted SPR values. Quoted values are relative to water.

Method	Mean	Std	RMS error
Dose extinction	-0.06	0.28	0.29
SECT	-0.25	1.57	1.59
DECT $\rho_e - Z$	-0.28	0.55	0.61

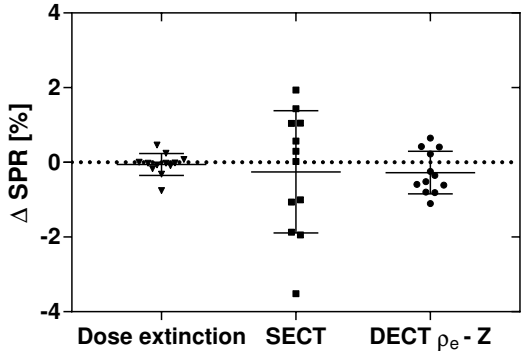

FIG. 5: Validation of our measurement methods. We compare SPR values from tissue-equivalent materials of the Gammex RMI-467 phantom determined with 1) dose extinction (left), 2) an empirical SECT calibration curve (middle) and 3) the stoichiometric DECT calibration by Bourque *et al.* (right) to the SPR measured in a water tank. Quoted values are relative to water. Also presented are mean errors and the standard deviations.

B. Real tissue samples

For each investigated tissue type, we collect 16 data points from adjacent ionization chambers of the array, presented in figures 6 and 7. With 16 data points per sample, we extract mean error and standard deviation for the WER per organ, resulting in a total number of detector elements of $N = 112$ for soft tissues and $N = 80$ for bones. To determine if the WER data come from a normally distributed population, we apply a Lilliefors test. The test result showed that all data with few exceptions are normally distributed, therefore we can apply a paired two-tailed t-test on most of the data. For the not normally distributed exceptions, we apply a Wilcoxon signed-rank test. We perform the appropriate statistical test on each pair of observed WER differences (SECT vs. DECT $\rho_e - Z$, SECT vs. DECT ETD, DECT $\rho_e - Z$ vs. DECT ETD). For most pairings, we observe significant differences ($p<0.05$) between WER predictions. Whenever two distributions are not found to be significantly different, it is explicitly indicated in the graphs of figures 6 and 7. Additionally, table V shows an overall uncertainty on the WER estimation. We observe

uncertainties of 0.53% and 1.37% for SECT (soft tissues and bones respectively), 0.19% and 1.06% for DECT using the method by Bourque *et al.* (DECT $\rho_e - Z$) and 0.38% and 1.06% for the DECT ETD method. A paired t-test is performed on the overall distributions of WER errors for soft tissues and bones, showing significant differences ($p<0.05$) for all pairings.

TABLE V: Statistics of the WER errors in % for soft tissues and bones from CT imaging only. The bias is the mean WER error (Eq. 3). The type A uncertainty is the standard deviation of the WER error (Eq. 4) with a quadratic subtraction of the experimental uncertainty (σ_{exp}=0.20%). The overall uncertainty includes range degradation, I-value and CT calibration, and it is equal to the quadratic sum of the mean WER error and the standard deviation of the WER error (Eq. 8). Uncertainties are reported with a statistical significance of k=1.

	Soft tissues			Bones		
Method	Bias	Type A	Overall	Bias	Type A	Overall
SECT Schneider *et al.*	-0.44	0.29	0.53	-0.88	1.06	1.37
DECT $\rho_e - Z$	-0.01	0.19	0.19	-0.58	0.89	1.06
DECT ETD	-0.33	0.20	0.38	-0.14	1.05	1.06

C. Estimation of uncertainties in CT-number-to-SPR conversion

A detailed uncertainty budget is performed based on the statistical behavior of the data and estimations of experimental uncertainties and uncertainties from range degradation due to lateral inhomogeneities. In tables VI and VII, we summarize the uncertainties involved in the WER estimation arising from CT calibration and I-value. We establish an uncertainty budget for soft tissues and bones separately, due to differences in u_{deg} for the two different tissue types. The uncertainties arising from range degradation as well as I-value occur in both imaging modalities (i.e., SECT and DECT). The differences in the observed uncertainties u_{WER} for SECT and DECT must therefore arise from the CT calibration and conversion technique. Once the uncertainty on CT calibration and conversion only is isolated, in soft tissues we find an overall difference of -0.34% (68th percentile) between u_{comb} of DECT and SECT. In bones, we find an overall difference of -0.31% (68th percentile) on u_{comb} for DECT over SECT, which are mostly dominated by the bias observed with the SECT method. The results are summarized in Tab. VIII.

The voxelwise prediction of the SPR using SECT and DECT is subject to CT imaging artifacts and errors from the CT-number-to-SPR conversion. These errors and resulting range uncertainties were investigated within a theoretical scope in our previous work [24]. For noise levels comparable to this study (7 HU for SECT, 8 HU and 12 HU for DECT), we reported a theoretical error on range (95th percentile) in soft tissues of 0.68% for SECT, and 0.30% for DECT. These values are taken from figure 4a) in [24]. Based on the methods applied in the previous study, we estimate the theoretical range uncertainties to compare to the findings of the present study. The theoretical range uncertainty estimation for soft tissues is done for 16 cm of water equivalent range to match the thickness of the samples. For bones, we quadratically combine the uncertainty from 8 cm of bones and 8 cm of soft tissues. In soft tissues we find a difference of -0.34% (68th percentile) between u_{CT} of DECT and SECT. In bones, we find a difference of -1.14% (68th percentile) on u_{CT} for DECT over SECT, which is mostly dominated by the bias observed in the SECT method. The results are summarized in Tab. VIII.

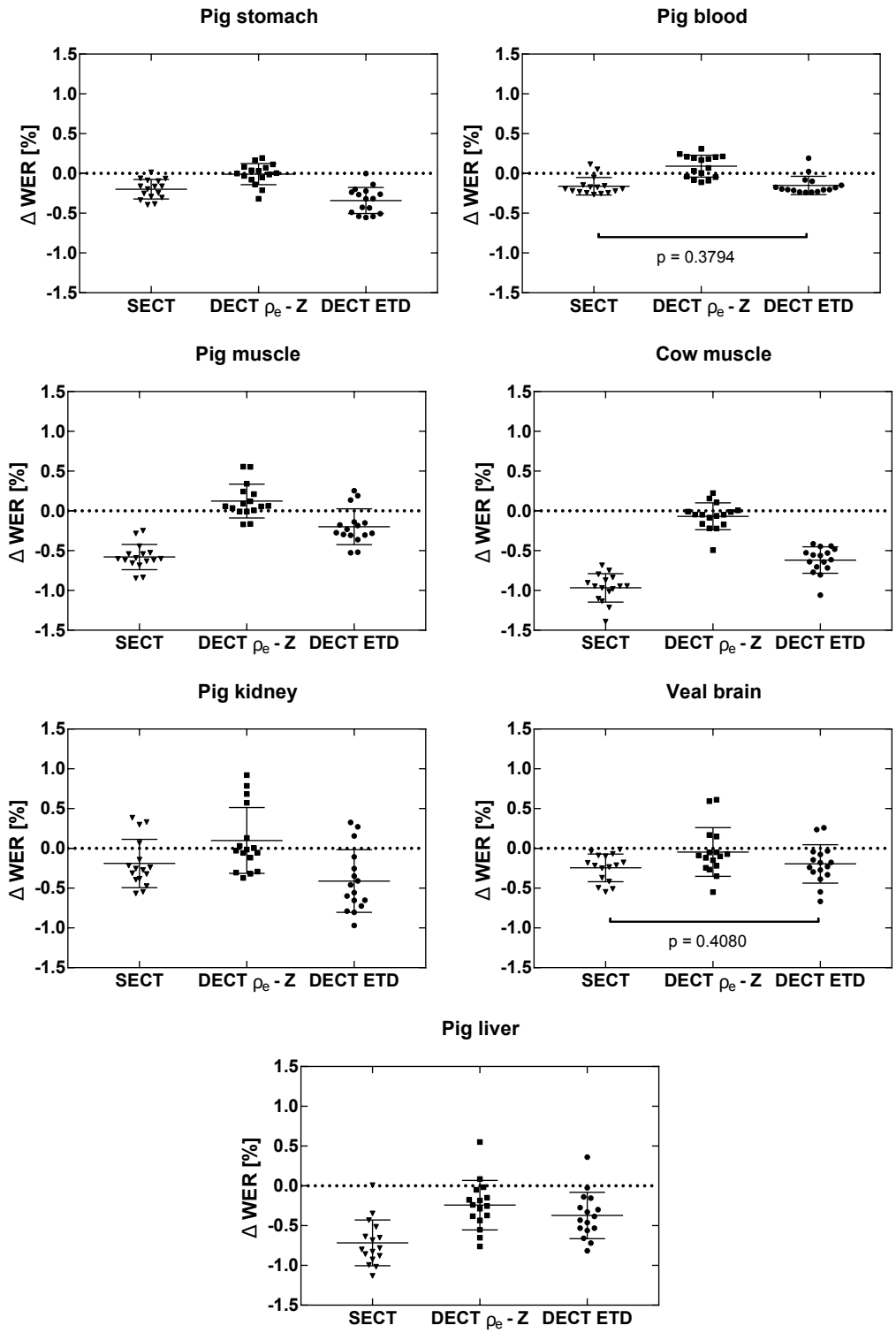

FIG. 6: Soft tissue samples: The tissue-specific percentage difference between the WER determined with dose extinction and CT imaging is shown here. For every tissue type, the WER differences, mean error and standard deviation are shown for the stoichiometric SECT calibration (Schneider *et al.*), the stoichiometric DECT calibration (Bourque *et al.*) and the DECT ETD method (Lalonde *et al.*). The observed WER distributions are all significantly different with two exceptions (SECT vs. DECT ETD for pig blood and veal brain).

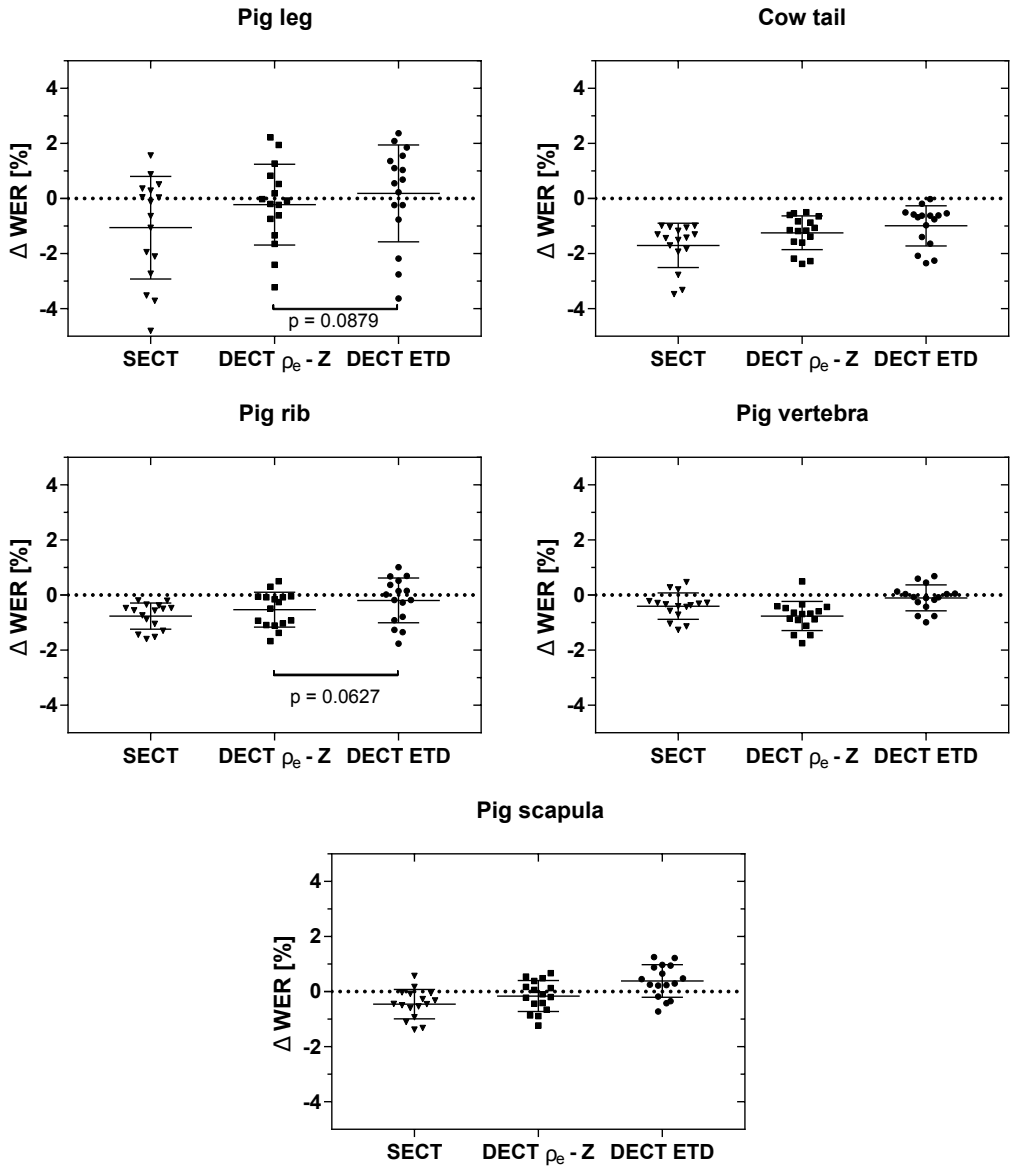

FIG. 7: Bone samples: The tissue-specific percentage difference between the WER determined with dose extinction and CT imaging is shown here.For every tissue type, the WER differences, mean error and standard deviation are shown for the stoichiometric SECT calibration (Schneider *et al.*), the stoichiometric DECT calibration (Bourque *et al.*) and the DECT ETD method (Lalonde *et al.*).

IV. DISCUSSION

A. Performance of DECT methods in determining SPR

In this study, we measure the WER of biological samples containing several soft and bony tissues from three different types of animals. We pack isolated organs (when possible) in plastic boxes, filled up with saline water. We measure the WER of a proton beam through the samples using the dose extinction method with the Matrixx detector. We scan the samples with SECT and DECT and determine the SPR voxelwise from three different calibration methods. We then predict the WER of the proton beam through the sample by performing Monte Carlo

TABLE VI: Uncertainties of estimated WER in % over all soft tissues. The uncertainty due to beam degradation from multiple Coulomb scattering is assumed negligible for soft tissues. The combined uncertainty on CT calibration and conversion and I-value u_{comb} is found by quadratic subtraction (Eq. 10).

Soft tissues	u_{WER}	u_{deg}	u_{comb}
SECT Schneider *et al.*	0.53	negligible	0.53
DECT $\rho_{\mathrm{e}} - Z$	0.19	negligible	0.19
DECT ETD	0.38	negligible	0.38

TABLE VII: Uncertainties of estimated WER in % over all bones. The uncertainty due to beam degradation from multiple Coulomb scattering is estimated from Paganetti (2012) [6]. The combined uncertainty on CT calibration and conversion and I-value u_{comb} is found by quadratic subtraction (Eq. 10).

Bones	u_{WER}	u_{deg}	u_{comb}
SECT Schneider *et al.*	1.37	0.14	1.36
DECT $\rho_{\mathrm{e}} - Z$	1.06	0.14	1.05
DECT ETD	1.06	0.14	1.05

simulations on the CT-predicted SPR maps. Our measurement method is validated against experimental reference SPR values of tissue-equivalent materials. The SPRs determined with the dose extinction method show excellent agreement with their respective reference SPRs. The reference SPRs determined experimentally are then used to validate the CT-based SPR values. SPRs of tissue-equivalent materials predicted from DECT show a much better agreement with reference values than SPRs predicted by SECT.

Our results with animal samples show that overall DECT performs well in determining SPRs of biological samples. The errors on WER prediction using SECT reported herein agree partially with earlier studies [5, 27]. The WER of soft tissues can be predicted well within 1% uncertainty, but higher uncertainties for bones are observed, mostly due to the bias observed with the SECT technique. Range errors in bones are lower than the results reported by Schaffner and Pedroni [5]. They predicted an overall error in range, caused by CT artifacts and CT calibration, of 1.1% in soft tissues and 1.8% in bones. It should be mentioned here that the range uncertainty for bones in their study was based on a single piece of bone scanned in air, for which they adapt a special calibration curve. Since their bone samples were scanned in air, they also scanned their calibration materials in air and adjusted the calibration curve accordingly. This procedure establishes similar beam hardening conditions for calibration and scanning and therefore reduces the bias in bone tissues. In the present study, we use a single calibration curve for all tissues to reproduce clinical conditions. Additionally, we used five different kinds of bones from two different animals, and take 16 measurements per piece of bone.

The stoichiometric DECT calibration by Bourque *et al.* improves range predictions by 0.34% in soft tissue and 0.31% in bones. Figure 6 shows that the SPR prediction from Bourque *et al.* reduces the bias in range prediction for each of the investigated soft tissues. For the bony tissues, we observe a reduced bias for 4 out of 5 investigated tissue types, with the exception of pig vertebra. Even though there is no improvement observed for this particular sample between SECT and DECT, the bias for both techniques is very low.

As a third CT calibration method, we apply the eigentissue decomposition method as proposed by Lalonde *et al.*.

TABLE VIII: Comparison of the uncertainty from imaging (u_{comb}) of the method by Bourque *et al.* with theoretical data. Theoretical data are estimated using the method of Bär *et al.*[24]. Theoretical data are not affected by the uncertainty on I-values ($u_I = 0$).

	Experimental			Theoretical		
	SECT	DECT	Difference	SECT	DECT	Difference
Soft tissue	0.53	0.19	-0.34	0.55	0.21	-0.34
Bones	1.36	1.05	-0.31	1.76	0.62	-1.14

We observe improvement in range predictions by 0.15% for soft tissues and 0.31% for bones. From figure 6 we observe that for three soft tissue types, the ETD method outperforms SECT based range predictions. For two tissue types, we observe a performance similar to SECT. For two tissue types, we observe that SECT predicted ranges are slightly better than ETD predicted ranges. It is worth noting that the ETD method is outperformed by the $\rho_{\text{e}} - Z$ formalisms in soft tissues. One hypothesis is that the important presence of salt in the saline solution affected the accuracy of the ETD method, designed specifically for human tissue compositions. This behavior is expected due to the penalizing of unlikely solutions in the Bayesian approach of the ETD, which is used in this study. The prior function is constructed from human tissue composition, which makes the method potentially less applicable to tissues different from human tissues. It is however possible to adjust the severity of the regularization function. In this study, we used a regularization value of $\alpha = 0.4$, as suggested in the original publication of Lalonde *et al.*. However, it would be possible to optimize this parameter to reduce the severity of regularization, which can increase suitability of the method to tissues with compositions different from those of human tissues. It should be mentioned here that the main advantage of this method, contrarily to the two others, is that the ETD method is also suitable to estimate any other physical quantity relevant for proton therapy because it estimates elemental compositions prior to calculate the SPR. One possible application would be to use the estimated compositions to define the materials used for Monte Carlo simulations. In this study, Monte Carlo input data are obtained from SPR maps using a look-up table. This is one way of defining input materials, although not necessarily optimal. Future work can focus on implementing the direct use of DECT predicted elemental compositions, as suggested in earlier studies [20, 21, 25]. Finally, one other advantage of the ETD formalism is its generality, that makes it suitable for any number of CT energy. Using more sophisticated detectors, such as photon counting detectors (PCD), the method could potentially yield a lower bias on range prediction.

B. Improvements from SECT to DECT

This study is the first to validate SECT and DECT predicted water equivalent ranges in a variety of animal soft tissues and bones using Monte Carlo to simulate the proton path. One major finding of this study is the validity of the SECT calibration in the investigated soft tissues. SECT and DECT both predict the experimentally measured ranges with a high accuracy level of 1% or better, and neither method is largely biased. This finding validates that an SECT stoichiometric calibration curve is well suited for the use in soft tissues. Future work could focus on the investigation of adipose tissue, which is not considered in this study.

In samples containing bones, we observe an overall improvement of 0.37% with DECT. These results indicate that one major gain in the use of DECT over SECT lies in the bony tissues. Before concluding on the validity of a DECT method and recommend it for clinical use, it's benefit in bones needs to be validated, especially since SECT performs very well in soft tissues and no major advantages are to be expected. Furthermore, intensive studies need to be performed taking lung tissues into account.

We attempt to provide a detailed uncertainty budget to isolate the uncertainty coming from tissue characterization from SECT and DECT imaging only. Besides the uncertainties of the WER measurement, we estimate the sources of uncertainties from beam range degradation due to lateral heterogeneities, the I-value, as well as the CT imaging and calibration methods (i.e., conversion of CT data to SPR). Our soft tissue samples are homogeneous, hence one expects fewer partial volumes artifacts and negligible effects from range degradation due to Coulomb and nuclear scattering, which is more relevant in bones. Here, we use a constant uncertainty value of 0.14% to quantify the remaining uncertainties from lateral inhomogeneities. The uncertainty on SPR arising from the I-value is expected to be smaller in soft tissues than in bones due to the significant water content. Uncertainties reported in Tab. VI and VII are based on approximate estimations of uncertainty contributions from range degradation, however we report a combined uncertainty for CT imaging, CT calibration and I-value.

For water, an I-value of 75.3 eV is used. This value is derived using the Bragg additivity rule, elemental I-values for hydrogen and oxygen are taken from the ICRU report 37. We choose to comply with the Bragg additivity rule for consistency reasons, since the DECT method by Lalonde and Bouchard derives elemental compositions and uses the Bragg additivity rule to calculate composite I-values. A different I-value for water can potentially result in a systematic shift of the CT-predicted WER values. However, the impact of a different I_{water} on the SPR and therefore the WER uncertainty depend on the tissue type, since correlations in soft tissues need to be taken into account.

The uncertainties we quote here are the combined uncertainties from CT imaging and calibration and I-values. The uncertainty on the I-value was determined in earlier studies by Bichsel and Hiraoka [32] and Kumazaki *et al.* [31], quoting a contribution on range uncertainty of 1.5% [6]. Our results however suggest that this number is overestimated, especially in soft tissues. As mentioned before, it can be shown mathematically that the contribution of I-value uncertainties on range uncertainty depends in the tissue type. As De Smet *et al.* [35] show, the uncertainty on SPR originating from I-value uncertainties can be as low as 0.1%, depending on the tissue's water content. Another study by Yang *et al.* reports similar results, quoting SPR uncertainties of 0.23% (soft tissues) and 0.65% (bones) from uncertainties in I-values [36]. Hence, range uncertainties in soft tissues are much lower than range uncertainties in bones, which is in agreement with what we observe in this study.
In this study we quote the improvement from SECT to DECT as a difference in the combined uncertainty. We expect that the uncertainty arising from the I-value (u_I) is larger than the uncertainty arising from CT imaging

and calibration (u_{CT}). Consequently, it can be shown mathematically that the uncertainty difference arising from CT imaging and calibration only must be higher than the combined uncertainty difference. Hence our quoted values represent the lowest possible improvement in (u_{CT}) that can be achieved using DECT over SECT, when I-value uncertainties are taken into account. This is in agreement with our theoretical study. Results in Tab. VIII show that the experimental results for soft tissues are in good agreement with theoretical estimations of range uncertainties, suggesting overall improvements on SPR estimations with DECT. For bones, we observe smaller improvements experimentally than theoretically. However, the improvements in the theoretical study are only quoting improvements from CT imaging and calibration and are hence expected to be larger than the combined uncertainty difference. This suggests that further work is necessary to achieve accurate information on the uncertainty of the I-value and evaluate the effects of using the Bragg additivity rule to determine I-value on the proton beam range errors.

Additional sources of uncertainties could also arise from our setup. Although our experimental setup is reproducible since we use organs tightly packed in plastic containers, slight movement of tissues within the containers during the transport between CT and treatment room are difficult to avoid. Furthermore, the representation of human tissues using animal tissues has some limitations. While we observe very similar ED and EAN in soft tissues, animal bones can be denser than human bones. For this reason, we deliberately avoided the use of very dense animal bones such as a cow's leg.

V. CONCLUSION

This study aims at validating two DECT formalisms to extract SPR values using heterogeneous animal tissue samples. SECT and DECT scans of homogeneous tissue-equivalent plastic materials are used to estimate their SPRs and the same technique is applied on the biological samples to predict their WER in Monte Carlo simulations. Using WER measurements with the dose extinction method, we isolate an uncertainty estimation of combined CT imaging artifacts, CT-number-to-SPR conversion and I-value on the range of a proton beam. We observe clear improvements in determining SPRs with DECT for homogeneous tissue substitutes as well as improvements in determining WER in animal tissues since range errors are dominated by the bias produced by the SECT method. We conclude benefits of 0.34% in soft tissues and 0.31% in WER prediction in the presence of bony tissues using DECT over SECT. Future MC-based treatment planning systems however might further benefit from improved input data derived with the ETD method.

VI. ACKNOWLEDGEMENTS

This work is funded by EPSRC UK (case studentship No. 20873) and the National Physical Laboratory (UK).

[1] JE Munzenrider and NJ Liebsch. Proton therapy for tumors of the skull base. *Strahlentherapie und Onkologie*, 175(2):57–63, 1999.

[2] U Mock, D Georg, J Bogner, T Auberger, and R Pötter. Treatment planning comparison of conventional, 3D conformal, and intensity-modulated photon (IMRT) and proton therapy for paranasal sinus carcinoma. *International Journal of Radiation Oncology* Biology* Physics*, 58(1):147–154, 2004.

[3] U Schneider, E Pedroni, and A Lomax. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Physics in medicine and biology*, 41(1):111, 1996.

[4] HQ Woodard and DR White. The composition of body tissues. *The british journal of Radiology*, 59(708):1209–1218, 1986.

[5] B Schaffner and E Pedroni. The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between ct-hu and proton stopping power. *Physics in medicine and biology*, 43(6):1579, 1998.

[6] H Paganetti. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in medicine and biology*, 57(11):R99, 2012.

[7] A-C Knopf and A Lomax. In vivo proton range verification: a review. *Physics in medicine and biology*, 58(15):R131, 2013.

[8] M Yang, G Virshup, J Clayton, XR Zhu, R Mohan, and L Dong. Theoretical variance analysis of single-and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues. *Physics in medicine and biology*, 55(5):1343, 2010.

[9] BJ Heismann, J Leppert, and K Stierstorfer. Density and atomic number measurements with spectral x-ray attenuation method. *Journal of applied physics*, 94(3):2073–2079, 2003.

[10] M Bazalova, J-F Carrier, L Beaulieu, and F Verhaegen. Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations. *Physics in medicine and biology*, 53(9):2439, 2008.

[11] N Hünemohr, B Krauss, C Tremmel, B Ackermann, O Jäkel, and S Greilich. Experimental verification of ion stopping power prediction from dual energy CT data in tissue surrogates. *Physics in medicine and biology*, 59(1):83, 2013.

[12] G Landry, J Seco, M Gaudreault, and F Verhaegen. Deriving effective atomic numbers from DECT based on a parameterization of the ratio of high and low linear attenuation coefficients. *Physics in medicine and biology*, 58(19):6851, 2013.

[13] AE Bourque, J-F Carrier, and H Bouchard. A stoichiometric calibration method for dual energy computed tomography. *Physics in medicine and biology*, 59(8):2059, 2014.

[14] M Saito. Potential of dual-energy subtraction for converting ct numbers to electron density based on a single linear relationship. *Medical physics*, 39(4):2021–2030, 2012.

[15] M Saito and S Sagara. A simple formulation for deriving effective atomic numbers via electron density calibration from dual-energy CT data in the human body. *Medical Physics*, pages –, 2017.

[16] C Möhler, P Wohlfahrt, C Richter, and S Greilich. Range prediction for tissue mixtures based on dual-energy CT. *Physics in medicine and biology*, 61(11):N268–, 2016.

[17] D Han, JV Siebers, and JF Williamson. A linear, separable two-parameter model for dual energy CT imaging of proton stopping power computation. *Medical physics*, 43(1):600–612, 2016.

[18] V T Taasti, J BB Petersen, L P Muren, J Thygesen, and D C Hansen. A robust empirical parametrization of proton stopping power using dual energy CT. *Medical Physics*, 43(10):5547–5560, 2016.

[19] P Wohlfahrt, C Möhler, V Hietschold, S Menkel, S Greilich, M Krause, M Baumann, W Enghardt, and C Richter. Clinical implementation of dual-energy ct for proton treatment planning on pseudo-monoenergetic ct scans. *International Journal of Radiation Oncology* Biology* Physics*, 97(2):427–434, 2017.

[20] G Landry, K Parodi, JE Wildberger, and F Verhaegen. Deriving concentrations of oxygen and carbon in human tissues using single-and dual-energy CT for ion therapy applications. *Physics in medicine and biology*, 58(15):5029, 2013.

[21] N Hünemohr, H Paganetti, S Greilich, O Jäkel, and J Seco. Tissue decomposition from dual energy CT data for MC based dose calculation in particle therapy. *Medical physics*, 41(6):061714, 2014.

[22] A Lalonde and H Bouchard. A general method to derive tissue parameters for Monte Carlo dose calculation with multi-energy CT. *Physics in Medicine and Biology*, 61(22):8044, 2016.

[23] Arthur Lalonde, Esther Bär, and Hugo Bouchard. A Bayesian approach to solve proton stopping powers from noisy multi-energy CT data. *Medical Physics*, pages –, 2017.

[24] E Bär, A Lalonde, G Royle, and H Bouchard. The potential of dual-energy CT to reduce proton beam range uncertainties. *Med. Phys.*, 2017. Under Review.

[25] Nace Hudobivnik, Florian Schwarz, Thorsten Johnson, Linda Agolli, George Dedes, Thomas Tessonnier, Frank Verhaegen, Christian Thieke, Claus Belka, and Wieland H Sommer. Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy ct images. *Medical physics*, 43(1):495–504, 2016.

[26] Jiahua Zhu and Scott N Penfold. Dosimetric comparison of stopping power calibration with dual energy CT and single energy CT in proton therapy treatment planning. *Medical physics*, 43(6):2845–2854, 2016.

[27] Rongxiao Zhang, Esther Baer, Kyung-Wook Jee, Gregory C Sharp, Jay Flanz, and Hsiao-Ming Lu. Investigation of real tissue water equivalent path lengths using an efficient dose extinction method. *Physics in medicine and biology*, 62(14):5640–5651, 2017.

[28] ICRU. Stopping powers for electrons and positrons. *ICRU report 37*, 1984.

[29] W Schneider, T Bortfeld, and W Schlegel. Correlation between ct numbers and tissue parameters needed for monte carlo simulations of clinical dose distributions. *Physics in medicine and biology*, 45(2):459, 2000.

[30] Martin J Berger. Estar, pstar and astar: Computer programs for calculating stopping powers and ranges for electrons, protons and helium ions. pages –, 1995.

[31] Y Kumazaki, T Akagi, T Yanou, D Suga, Y Hishikawa, and T Teshima. Determination of the mean excitation energy of water from proton beam ranges. *Radiation Measurements*, 42(10):1683–1691, 2007.

[32] Hans Bischel and Takeshi Hiraoka. Energy loss of 70 mev protons in elements. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 66(3):345–351, 1992.

[33] ICRU. Stopping powers and ranges for protons and alpha particles. *ICRU report 49*, 1993.

[34] N Huenemohr, B Krauss, J Dinkel, C Gillmann, B Ackermann, O Jäkel, and S Greilich. Ion range estimation by using dual energy computed tomography. *Zeitschrift für Medizinische Physik*, 23(4):300–313, 2013.

[35] V De Smet, F Labarbe R and, Vander Stappen, B Macq, and E Sterpin. A Water-Content Based Formalism to Mitigate Stopping Power Ratio Uncertainties Caused by Mean Excitation Energy Uncertainties in Hadron Therapy. *Medical Physics*, 2017.

[36] M Yang, XR Zhu, PC Park, U Titt, R Mohan, G Virshup, JE Clayton, and L Dong. Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration. *Physics in medicine and biology*, 57(13):4095, 2012.