The potential of dual-energy CT to reduce proton beam range uncertainties

Esther Bär,[1,2] Arthur Lalonde,[3] Gary Royle,[2] Hsiao-Ming Lu,[4] and Hugo Bouchard[1,3,5]

[1] *Acoustics and Ionising Radiation Team, National Physical Laboratory, Teddington TW11 0LW, UK*

[2] *Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, UK*[*]

[3] *Département de physique, Université de Montréal, Montréal, QC H3T 1J4, Canada*

[4] *Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA*

[5] *Centre de recherche du Centre hospitalier de l'Université de Montréal, Montréal, QC H2X 0A9, Canada*[†]

Purpose: Dual-energy CT (DECT) promises improvements in estimating stopping power ratios (SPRs) for proton therapy treatment planning. Although several comparable mathematical formalisms have been proposed in literature, the optimal techniques to characterize human tissue SPRs with DECT in a clinical environment are not fully established. The aim of this work is to compare the most robust DECT methods against conventional single-energy CT (SECT) in conditions reproducing a clinical environment, where CT artifacts and noise play a major role on the accuracy of these techniques.

Methods: Available DECT tissue characterization methods are investigated and their ability to predict SPRs is compared in three contexts: 1) a theoretical environment using XCOM cross sections database; 2) experimental data using a dual-source CT scanner on a calibration phantom; 3) simulations of a virtual humanoid phantom with the *ImaSim* software. The latter comparison accounts for uncertainties caused by CT artifacts and noise, but leaves aside other sources of uncertainties such as CT grid size and the I-values. To evaluate the clinical impact, a beam range calculation model is used to predict errors from the probability distribution functions determined with *ImaSim* simulations. Range errors cause by SPR errors in soft tissues and bones are investigated.

Results: Range error estimations demonstrate that DECT has the potential of reducing proton beam range uncertainties by 0.4% in soft tissues using low noise levels of 12 and 8 HU in DECT, corresponding to 7 HU in SECT. For range uncertainties caused by the transport of protons through bones, the reduction in range uncertainties for the same levels of noise is found to be up to 0.6 to 1.1 mm for bone thicknesses of ranging from 1 to 5 cm, respectively. We also show that for double the amount noise, i.e., 14 HU in SECT and 24 and 16 HU for DECT, the advantages of DECT in soft tissues are lost over SECT. However in bones, the reduction in range uncertainties is found to be between 0.5 and 0.9 mm for bone thicknesses ranging from 1 to 5 cm, respectively.

Conclusion: DECT has a clear potential to improve proton beam range predictions over SECT in proton therapy. However, in the current state high levels of noise remain problematic for DECT characterization methods and do not allow getting the full benefits of this technology. Future work should focus on adapting DECT methods to noise and investigate methods based on raw-data to reduce CT artifacts.

Keywords: Proton therapy, Range uncertainties, Dual-Energy CT, Tissue characterization, Proton stopping power

I. INTRODUCTION

The benefit of proton therapy lies in the favorable energy deposition properties of its particles. Protons deposit most of their energy at the end of their tracks due to the low scattering power of most human tissues, allowing for highly conformal dose distributions and a high degree of normal tissue sparing distal to the target volume. Conventionally, radiotherapy planning is based on computed tomography (CT) images. For proton therapy dose calculation, CT numbers need to be converted into tissue stopping power ratios relative to water (SPRs), which are used to calculate the beam range in the patient and the energy deposited along the penetration path. To exploit the full benefits of protons and to avoid errors in dose delivery at the distal fall-off, accurate conversion from CT numbers to SPR is essential. To further improve clinical outcomes of proton therapy, one must aim at a higher precision, which allows us to reduce safety margins and thus irradiate less healthy tissue, while maintaining conformal target dose.

In clinical practice, human tissue characterization for treatment planning is achieved by acquiring a CT scan on the patient and then converting the data into SPRs. Conventionally, the CT scan is acquired using one single energy spectrum, e.g., single-energy CT (SECT), and one clinically reliable method to obtain SPR from CT numbers is the calibration method proposed by Schneider *et al.* (1996) [1]. In this procedure, a relation between calculated SPRs of human reference tissues [2, 3], and CT numbers (in Hounsfield units) are determined using a plastic phantom with radiological properties equivalent to that of human tissues.

The calibration of Schneider *et al.* (1996), referred to as the SECT stoichiometric calibration method throughout this paper, is fairly accurate in predicting human tissue SPR [4]. Schaffner and Pedroni (1998) verified the SECT stoichiometric calibration by measuring pairs of CT numbers and SPR using animal tissue samples. They found a precision in SPRs of ± 1.1% for soft tissues and ± 1.8% for bones, which translates into range uncertainties of up to 3 mm for therapeutic energies. In more recent work, combined uncertainty in proton range estimation coming from CT calibration was reported to be 2.7-3.5% + 1.0-1.2 mm (1.5 standard deviation), excluding biological effects [5, 6]. While a large uncertainty is associated to the knowledge of the mean excitation energy (I-value) [6], another limitation in the accuracy of proton beam treatment planning is from CT data. In SECT, data is limited to a single dimension per voxel and this is problematic since HU-SPR calibration curves are not one-to-one relations (i.e., bijections) for human tissues. While both HU and SPR values are dominated by the electron density (ED), these quantities depend on other properties of the tissues, such as the effective atomic number (EAN) or the I-value [5, 6]. In turn, these properties depend on the elemental composition. Small patient-to-patient variations in density and elemental compositions were shown to introduce significant changes in CT numbers [7]. These variations are not necessarily resolved by the SECT stoichiometric calibration since the HU-to-SPR conversion approach cannot explicitly decouple the dependence of CT

*Electronic address: esther.baer.11@ucl.ac.uk
†Electronic address: h.bouchard@umontreal.ca

numbers to elemental compositions and mass density, therefore limiting the precision to which tissue characteristics can be resolved.

Dual-energy CT (DECT) has the potential to improve the conversion of CT data to SPR. Over the last decade, several papers were published on DECT to either show potential benefits for radiotherapy or to propose a mathematical formalism to extract tissue parameters relevant to dose calculation. Recent publications propose the extraction of ED and EAN (or alternatively, the I-value), from DECT images [8–16]. These methods rely on post-reconstruction data analysis, conversely to sinogram-based methods (e.g., Refs. [17, 18]) which are yet to be fully explored. Studies on DECT for proton therapy typically report errors on stopping power determination between 0.5% and 1.5% [8–16]. Although there exists no direct relation between X-ray attenuation and stopping powers, it was shown that DECT has the potential to substantially improve proton radiotherapy planning as it is widely clinically available.

The present paper aims at evaluating the potential of DECT to reduce proton beam range uncertainties in a clinical context, with focus on CT artefacts and noise, and leaving aside uncertainties related to the I-value and the CT grid size. The performance of different mathematical techniques to predict proton stopping powers are compared theoretically, experimentally and with simulated CT data. Since a consistent basis is needed for such comparison, all methods are compared under the same conditions using the same calibration phantom, CT images and statistical quantities. The resulting distributions of SPR errors are used to estimate the impact of proton beam range uncertainties, in this way allowing estimating the gain in precision provided by DECT in a clinical environment. A comparison against the SECT stoichiometric method is achieved in order to predict the potential clinical impact of DECT in proton therapy dose calculation.

II. METHODS

A. An overview of investigated DECT tissue characterization methods

In literature, several techniques to extract proton stopping powers from DECT images exist. Commonly, these methods extract the density ρ, or alternatively, the electron density ρ_{e}, plus the effective atomic number Z_{eff}[19] or Z_{med}[13] to derive the I-value via a parametric relationship converting Z to I for human tissues[7, 13] . Some published methods [15, 16] do not require the concept of effective atomic number to determine tissue parameters. The key elements of all the formalisms studied are summarized in tables I and II.

While they are reported in chronological order, there are two types of techniques compared. The first type is based on parameter extraction, i.e., either ρ_{e}-Z or ρ_{e}-I. With this type, proton SPRs can be calculated with Bethe's equation:

$$S = \rho_{\mathrm{e}} \frac{k_0}{\beta^2} \left[\ln \left(\frac{2 m_e c^2 \beta^2}{I(1-\beta^2)} \right) - \beta^2 \right], \tag{1}$$

by taking the ratio of the resulting stopping power S for a given ρ_e, I-value and reference energy. Note that in this paper, $I_w = 73.924$ eV is used. For the techniques extracting Z, I is calculated depending on which definition of effective atomic number is applied. The conversion Z into I proposed either by Yang *et al.* [7] or Bourque *et al.* [13] is used for Z_{eff} or Z_{med}, respectively.

The second type of technique is meant to extract elemental weights fractions and mass or electron density. There exists three methods compared herein predicting elemental weight fractions and density from DECT. From the predictions of these methods, the I-value of each pixel is calculated using the Bragg additivity rule[20]:

$$\ln I = \sum_i \lambda_i \ln I_i. \tag{2}$$

where λ_i are the elemental electronic fractions. The SPR is then obtained with equation 1 using the electron density, either obtained directly or calculated from the mass density and the elemental composition allowing estimating Z/A. It is worth noting that methods predicting elemental compositions and density are suitable with Monte Carlo radiation transport algorithms, which are known to improve the accuracy of range predictions in heterogeneous media [6]. However, since most clinical dose calculation engines require SPRs, the present focus is on the ability to predict these ratios and further evaluate the impact on beam range predictions using an analytic model.

1. Bazalova et al. 2008

A tissue characterization method for monoenergetic photons was proposed by Torikoshi *et al.* [21], but first adapted by Bazalova *et al.* [9] for the use in commercial CT scanners. In this parametrization, the photoelectric attenuation and Compton scattering are expressed as quadratic functions $F(E, Z)$ and $G(E, Z)$. $F(E, Z)$ and $G(E, Z)$ are obtained by fitting of quadratic functions to elemental cross sections (i.e., the XCOM database[22]). For the use of this parametrization in a spectrum of energies, spectral weights and integration over the energy must be taken into account. Z_{eff} is found via numerical solution from two energies, ρ_{e} is obtained by substitution of Z_{eff}. In Bazalova *et al.*'s method presented here, the numerical solution for Z_{eff} is obtained using the MATLAB (The MathWorks, Inc., Natick, MA, USA) build in numerical solver *fzero*. Additionally, spectral attenuation in the examined object must be taken into account. Hence, the output spectrum of the X-ray tube is not used for tissue parameter extraction, but a tissue filtered spectrum. This tissue filtered spectrum is calculated using an analytical absorption model, which employs the attenuation law. As Bazalova *et al.* evaluated in their paper, it is valid to assume a filtering of 16 cm of water to describe every position within the round-shaped phantom.

2. Landry et al. 2013

To extract the effective atomic number with DECT, Landry *et al.* [11] developed a method combining previously proposed techniques. The approach was inspired by the SECT stoichiometric calibration by Schneider *et al.* (1996) [1]. The parametrization of Rutherford *et al.* [23] was utilized. This parametrization, in contrast to the parametrization by Alvarez and Macovski, comprises a term to take coherent scatter into account. In their method, Landry *et al.* proposed using the ratio of attenuation coefficients measured with the CT scanner at low and high energy in a two-step calibration procedure. In a first step, the attenuation coefficients of a calibration phantom are measured at two energy spectra. The measured values are used to find the stoichiometric parameters $k_{1\mathrm{kVp}}$ and $k_{2\mathrm{kVp}}$ as proposed by Schneider *et al.* (2000) [24] per energy. These parameters are then used to calculate attenuation coefficients of a set of human reference tissues [3]. The ratio of the calculated attenuation coefficients of human tissues serves as a basis data set to find the fit parameters $A_{\mathrm{l,h}}$, $B_{\mathrm{l,h}}$ and $C_{\mathrm{l,h}}$. These parameters correspond to A, B and C in in table I, with l for the low and h for the high energy spectrum. Z_{eff} is obtained by solving the parametrization for Z. To determine the electron density, Landry *et al.* recommended that the method by Saito is used to obtain ρ_{e}. Saito [10] developed a method to only extract electron densities from DECT. This approach employs a ΔHU, which is obtained as a linear combination of $\mathrm{HU_l}$ and $\mathrm{HU_h}$, with a single weighting factor. This factor is scanner specific and must be found in a calibration process, employing a calibration phantom.

As an extension of their method, Landry *et al.* [25] proposed a segmentation method to extract a full elemental composition from any Z_{eff} and ρ_{e} couple. First, these two parameters are calculated for a dataset of reference human tissues. Then, the tissue assigned in each voxel is the one showing the shortest generalized distance with the measured data in the ρ_{e}-Z_{eff} space. The segmentation technique allows assigning a tissue to each voxel and a generic elemental composition to the tissue. In the present paper, the determination of electron density and effective atomic number with the method of Landry *et al.* is referred to as Landry *et al.* # 1, while the one extracting elemental weights is Landry *et al.* # 2.

3. Hünemohr et al. 2014

The first of the existing DECT tissue parameter extraction methods based for clinical use was published in 2003 by Heismann *et al.* [8]. They employed the attenuation cross section (μ) parametrization from Alvarez and Macovski [26] and developed their formalism on post-reconstruction data. In the model, one first term describes the attenuation due to photoelectric effect, while the other term describes Compton scattering. Each physical effect has an associated coefficient (α and β) which quantifies the magnitude of the effect. The coefficients are energy-specific and can be found in a calibration process employing a DECT scan of materials with known compositions. The energy dependence of the system is furthermore taken into account by introducing parameters (g_{L} and g_{H}) integrating the over the energy spectrum using spectral weights $w_{\mathrm{L,H}}$. Hünemohr *et al.* [12] adapted the approach by Heismann *et al.* and propose a

calibration using a tissue characterization phantom instead of the integration over the spectral energies. Furthermore, the authors employ the mathematical methodologies of the ρ-Z projection of Heismann *et al.*, but substitute the mass density ρ by the electron density ρ_e. In this work, we chose to implement the version of Hünemohr *et al.*, employing ρ_e instead of ρ.

To take into account potential elemental composition variation for a given tissue within a population, Hünemohr *et al.* [27] proposed to parametrize elemental weights as a function of Z_eff and ρ_e. Thus, for each of the 13 elements (H, C, N, O, Na, Mg, P, S, Cl, K, Ca, Fe and I), a reference dataset of tissues is used to create a linear fit describing the weight of each element as a combination of ρ_e, Z_eff, and $\rho_e Z_\mathrm{eff}$, as recommended in their publication. In the present work, the determination of electron density and effective atomic number with the method of Hünemohr *et al.* is referred to as Hünemohr *et al.* # 1, while the one allowing to obtain elemental weights is Hünemohr *et al.* # 2.

4. Bourque et al. 2014

In the method by Bourque *et al.* [13], the attenuation coefficient relative to water is parametrized as a polynomial of the order M-1 with coefficients b_m. The parameters b_m are obtained from a least square fit to measured μ/μ_w from a CT scan of the calibration phantom. A specific definition of effective atomic number is used, Z_med, and their values for the phantom materials has previously been calculated and averaged for both energy spectra. The fit procedure to obtain coefficients b_m must be performed for both energies of the DECT scan separately. In analogy to the attenuation coefficient, Bourque *et al.* define a parametrization for the estimation of the effective atomic number, as listed in table I. It uses the dual-energy ratio Γ (defined as the attenuation coefficient of the low-energy scan relative to the high-energy scan) for its independence on electron density. To find the model parameters c_k, Γ is measured for the inserts of the calibration phantom and a least square fit of order $K-1$ is performed. For a dual energy CT scan of unknown tissues, Z_med and ρ_e are found by measurement of $(\mu/\mu_\mathrm{w})_\mathrm{L}$ and $(\mu/\mu_\mathrm{w})_\mathrm{H}$.

5. Van Abbema et al. 2015

Van Abbema *et al.* [14] developed a method that is not based on calibration, but requires spectral knowledge. They use the electron cross section parametrization ${}_\mathrm{e}\sigma^\mathrm{tot}(E,\widehat{Z})$ of Jackson and Hawkes [28], extended with fit functions to yield a dependency on E and Z. Knowledge of the spectral weighting function $w(E)$ at every energy increment $\mathrm{d}E$ is necessary. Z_eff is found by solving the ratio of attenuation coefficients at low and high energy numerically for Z and ρ_e is obtained by substitution of Z_eff. As this method makes use of spectral knowledge, the attenuation of the examined object must be taken into account, similarly to the method proposed by Bazalova *et al.*. To account for spectral hardening, van Abbema *et al.* propose to apply a $w(E)$ local weighting function (LWF), which is obtained iteratively from spectral weights $w(E)$ and the measured attenuation coefficients in the corresponding voxel.

6. Han et al. 2016

A recent paper by Han *et al.* [15] proposed a two-parameter model. They assume that the attenuation coefficient of an unknown material in a given voxel can be described as a linear combination of the attenuation coefficient of two basis materials μ_1 and μ_2. The basis materials are chosen as water and polystyrene for soft tissues, and water and an aqueous $CaCl_2$ solution (23%) for bony tissues. The parameters c_1 and c_2 are material specific, found by measuring the attenuation coefficients of the basis materials as well as the unknown material at two different energies. The integration over all energies of the spectrum is approximated in this model by using the mean energy of spectrum. ρ_e and I of unknown tissues are then found using the determined parameters c_1 and c_2, according to table I.

7. Lalonde and Bouchard 2016

Lalonde and Bouchard [16] introduced a representation of human tissues based on principal component analysis (PCA). An optimal basis of virtual materials (principal components, PC) is defined from a reference dataset of tissues, each of the described by a mass density and array of elemental compositions (H, C, N, O, Na, Mg, P, S, Cl, K, Ca, Fe and I). The partial electronic density y_k of each PC is retrieved by performing a material decomposition from DECT data. Once the y_k are solved, their sum equals the electronic density and the elemental composition is unfold from the PC content. To estimate the electronic cross section of each PC (i.e., f_k in table I), a calibration method similar to Bourque *et al.* [13] is proposed, but without the need for defining the effective atomic number. In this way, the attenuation coefficient relative to water is parametrized using a series of power specific average atomic numbers, i.e., $\overline{Z}, \overline{Z^2}, \overline{Z^3}, \ldots$, referred to as Z-space. The fit parameters are obtained for each energy and scanning protocol from a least square fit on measured μ/μ_w from a CT scan of a calibration phantom. It should be noted that only the formalism of Lalonde and Bouchard gives directly a complete set of elemental weights and mass density without intermediate step. However, two other methods (Landry *et al.* # 2 and Hünemohr *et al.* # 2) can be adapted to convert measured ρ_e and Z to suitable Monte Carlo inputs. These methods are investigated in this study and compared to the PCA approach of Lalonde and Bouchard.

B. Comparison of DECT tissue characterization methods

This section describes how the performance of the different DECT methods is compared. Firstly, a theoretical comparison with the XCOM photon cross sections database is performed in order to evaluate the theoretical robustness of the method. Secondly, methods are compared with respect to experimental measurements in order to eliminate the ones that are not practical for a clinical environment. Thirdly, methods are compared in an imaging simulation environment in order to reproduce the context of noise and imaging artefacts while allowing a comparison with ground truth values.

TABLE I: Summary of the theoretical foundation of different DECT formalisms.

	μ parametrization	Z definition	Requires CT calibration
Bazalova *et al.*	$\mu = \rho_e \sum_i w_i \left(Z^4 F(E_i, Z) + G(E_i, Z)\right)$	Mayneord ($m = 3.5$)	No
Landry *et al.* #1 and #2	$\mu = \rho_\mathrm{e} \left(A + BZ^m + CZ^n\right)$	Mayneord ($m = 3.3$)	Yes
Hünemohr *et al.* #1 and #2	$\mu = \rho_\mathrm{e} \left(\alpha \frac{Z^m}{E^l} + \beta\right)$	Mayneord ($m = 3.1$)	Yes
Bourque *et al.*	$\mu/\mu_\mathrm{w} = \rho_\mathrm{e} \sum_{m=1}^{M} b_m Z^{m-1}$	Behavior of electronic cross sections for elements	Yes
Van Abbema *et al.*	$\mu = \int_0^\infty w(E)\, {}_\mathrm{e}\sigma^\mathrm{tot}(E, \widehat{Z})\, \mathrm{d}E$	Behavior of $\frac{\mu_\mathrm{L}}{\mu_\mathrm{H}}$ for mixtures	No
Han *et al.*	$\mu = c_1\mu_1 + c_2\mu_2$	None	Yes
Lalonde and Bouchard	$\mu/\mu_\mathrm{w} = \overline{y}_0 f_0 + \sum_{k=1}^{K} y_k f_k$	None	Yes

TABLE II: Summary of different formalisms to predict tissue parameters with DECT.

	EAN	I-value	ED
Bazalova *et al.*	solve $\frac{u_\mathrm{L}}{u_\mathrm{H}}$ numerically	Yang *et al.*	substitute $\widehat{Z}$
Landry *et al.* #1 and #2	solve $\frac{u_\mathrm{L}}{u_\mathrm{H}}$ for Z	Yang *et al.* Bragg additivity rule	$\widehat{\rho}_\mathrm{e} = \frac{\Delta\mathrm{HU}}{1000} + 1$
Hünemohr *et al.* #1 and #2	substitute $\widehat{\rho}_\mathrm{e}$	Yang *et al.* Bragg additivity rule	$\widehat{\rho}_\mathrm{e} = \frac{1}{\beta} \frac{g_\mathrm{L}\mu_\mathrm{H} - g_\mathrm{H}\mu_\mathrm{L}}{g_\mathrm{L} - g_\mathrm{H}}$
Bourque *et al.*	$\widehat{Z}_\mathrm{eff} = \sum_{k=1}^{K} c_k \Gamma^{k-1}$	5^th-order fit with Z_med	$\widehat{\rho}_\mathrm{e,L/H} = \frac{u_\mathrm{L/H}}{\sum_{m=1}^{M} b_{m,\mathrm{L/H}} Z_\mathrm{eff}^{m-1}}$
Van Abbema *et al.*	solve $\frac{\mu_\mathrm{L}}{\mu_\mathrm{H}}$ numerically	Yang *et al.*	substitute $\widehat{Z}$
Han *et al.*	None	$\widehat{I}_x = f_I \left(\frac{c_1}{c_1+c_2}\right) \exp\left(\frac{c_1\rho_\mathrm{e1}\ln(I_1) + c_2\rho_\mathrm{e2}\ln(I_2)}{c_1\rho_\mathrm{e1} + c_2\rho_\mathrm{e2}}\right)$	$\widehat{\rho}_{ex} = c_1\rho_\mathrm{e1} + c_2\rho_\mathrm{e2}$
Lalonde and Bouchard	None	Bragg additivity rule	$\widehat{\rho}_\mathrm{e} = \overline{y}_0 + \sum_{k=0}^{K} y_k$

1. XCOM photon cross sections

A theoretical comparison of tissue characterization methods is performed using a set of 34 ICRU reference tissues [29]. The reference tissues with corresponding electron density are listed in table 2 of Bourque *et al.* [13] (see also corrigendum). For methods that require calibration, theoretical CT numbers of the tissue characterization phantom Gammex 467 (Sun Nuclear, Melbourne, FL, USA) are calculated and used for calibration (Hünemohr *et al.* #1 and #2, Landry *et al.* #1 and #2, Bourque *et al.*, Lalonde and Bouchard). For Han *et al.*, the calibration is done with water, polystyrene and a $CaCl_2$ acqueous solution (23%). The spectra used are from a dual source dual energy

CT scanner, kindly provided by the manufacturer (Somatom Definition Flash, Siemens Sector Healthcare, Forcheim, Germany), for energies of 100 kVp and 140 kVp/Sn (Siemens custom tin filtration). Values of ρ_e and Z_{eff} (or Z_{med}) are derived for the complete set of reference tissues using the listed tissue characterization methods. Theoretical SPR values are calculated using the given electron densities and atomic compositions of the 34 human tissues. The theoretical I-values of the tissues as given from ICRP 23 [30] are calculated using the Bragg additivity rule. Although there are uncertainties in the knowledge of the I-value, such calculated theoretical SPR values provide a comparison reference to our best nowadays knowledge and form the ground truth for our study. All methods are implemented using MATLAB.

2. Experimental comparison with calibration phantom

A comparison based on experimental data is performed. The Gammex 467 phantom is scanned in a Siemens Somatom Definition Flash DECT scanner. The tube voltages are 100 kV and 140 kV/Sn with tube currents 300 mAs and 232 mAs respectively. CT numbers of the tissue equivalent inserts are measured using a circular region of interest (ROI) readout (17.3 cm^3) over all slices of the phantom. The measured CT numbers are used to calibrate the methods that require calibration. Spectral knowledge is required for the spectral-based methods. The spectra of the Somatom scanner were kindly provided by the manufacturer. ρ_e and Z_{eff} are determined from the CT numbers measured in the ROIs, using each of the tissue characterization methods. A list of tissue equivalent inserts and their nominal electron densities (as specified by the phantom manufacturer) can be found in table 2 of Bourque *et al.* [13]. Again, theoretical reference values of SPRs are calculated using elemental I-values from ICRP 23 as well as the Bragg additivity rule shown in equation 2.

3. Simulated CT images

To evaluate the performance of DECT tissue characterization methods for proton therapy, it is not sufficient to test the accuracy of the methods on plastic phantoms only. Phantoms are often regular-shaped and made of similar chemical compositions, which do not entirely reproduce chemical compositions in patients. Hence, the methods need to be tested on an object resembling a patient anatomy and chemical composition of tissues, while being in a controlled environment with known reference values (referred here as ground truth).

To simulate CT images, the software *ImaSim*, developed by Landry *et al.* [31], is used. In their previous study comparing *ImaSim* against DECT phantom images, the authors concluded that the tool is suitable to explore applications of DECT imaging in radiotherapy [32]. However, they found differences of up to 15% when comparing simulated against experimentally measured relative attenuation coefficients μ/μ_w. While discrepancies are to be expected due to the complexity of reproducing realistic CT scanners (i.e., spectra non-uniformity, reconstruction algorithms, artifact corrections, etc.), some of the features in *ImaSim* are simplified compared to the clinical reality, which could partially

explain the magnitude of these differences. For the purpose of the present study, we need to assure that *ImaSim* can reproduce most imaging artifacts encountered in clinical conditions but also CT numbers with accuracy comparable to commercial CT scanners . Therefore, a validation of the software in its ability to predict μ/μ_{w} values is necessary to assure the performance of the basic reconstruction technique and the beam hardening correction algorithm. Furthermore, since we found that the ability to reproduce realistic noise with tube current settings is questionable, a model to account for image noise is used independently from *ImaSim*.

For the image simulations, four geometries are designed. To simulate the calibration procedures, the geometry of a Gammex 467 phantom is defined, reproducing the dimensions and materials of its homogeneous disk (i.e., a diameter of 32 cm) using specifications provided by the manufacturer. A second calibration phantom is defined specifically for the method of Han *et al.*[15]. It has the same dimensions and base material as the Gammex 467 phantom, but it has only 3 inserts: water, polystyrene and $CaCl_2$ aqueous solution (23%). A third calibration phantom meant to validate *ImaSim* is defined. It has the same 13 inserts but its cylinder base is replaced by an oval-shaped cylinder (i.e., an elliptic cylinder) of 32 cm width by 24 cm height. This allows us to evaluate the accuracy of the beam hardening correction in heterogeneous phantoms of irregular shapes. The fourth phantom designed has that same oval-shaped geometry and is a virtual patient phantom resembling a slice through human abdomen. The virtual patient consists of various structures filled with the elemental compositions and mass densities of 15 human tissues described by Woodard and White [2, 3]. The phantom is illustrated in figure 1, and the list of tissues used is found in table III.

All phantoms scans are simulated with 3 spectra available by default in the software: 100 kVp, 120 kVp and 140 kVp/Sn. For the SECT tissue characterization techniques, the 120 kVp spectrum is used, while for DECT the 100 kVp and 140 kVp/Sn spectra are used. Image simulations are performed with infinite tube current (mAs) to disregard noise. Reconstructions are performed with the Filtered Back Projection method using a Shepp Logan filter. For all simulations, the CT grid size is set to $0.9 \times 0.9 \times 1.0$ mm^3 voxels.

To study the impact of noise, Gaussian noise is added to simulated HU values obtained with *ImaSim*. For a consistent comparison between SECT and DECT, an equivalent amount of noise in the SECT image in terms of photon dose in water is calculated with the following relation:

$$\frac{\overline{\mu}_{\mathrm{w,SECT}}}{\Delta \mathrm{HU}^2_{\mathrm{SECT}}} = \frac{\overline{\mu}_{\mathrm{w,L}}}{\Delta \mathrm{HU}^2_{\mathrm{L}}} + \frac{\overline{\mu}_{\mathrm{w,H}}}{\Delta \mathrm{HU}^2_{\mathrm{H}}}. \tag{3}$$

with $\Delta \mathrm{HU}_{\mathrm{SECT}}$, $\Delta \mathrm{HU}_{\mathrm{L}}$ and $\Delta \mathrm{HU}_{\mathrm{H}}$ the noise levels in SECT, DECT low kVp and DECT high kVp, respectively. The average attenuation coefficients in water $\overline{\mu}_{\mathrm{w,SECT}}$, $\overline{\mu}_{\mathrm{w,L}}$ and $\overline{\mu}_{\mathrm{w,H}}$ are calculated using the 120 kVp, 100 kVp and 140 kVp/Sn, respectively. This relation is derived using Poisson's distribution for shot noise assuming an equal dose of photons used to generate the SECT image and the DECT image pair (i.e., $D_{\mathrm{SECT}} = D_{\mathrm{L}} + D_{\mathrm{H}}$). Note that because the dose is approximately proportional to the number of photons times the mass absorption coefficient in water $\mu_{\mathrm{ab,w}}/\rho$, neglecting electron transport (hence approximating that $\overline{\mu}_{\mathrm{ab,w}}/\rho \approx \overline{\mu}_{\mathrm{w}}$) and assuming shot noise to dominate ΔHU

yields equation 3. We study two levels of noise: 1) the low level, corresponding to SECT noise of $\Delta\mathrm{HU}_{\mathrm{SECT}} = 7$ and DECT noises of $\Delta\mathrm{HU}_{\mathrm{L}} = 12$ and $\Delta\mathrm{HU}_{\mathrm{H}} = 8$, and 2) the high level, corresponding to SECT noise of $\Delta\mathrm{HU}_{\mathrm{SECT}} = 14$ and DECT noises of $\Delta\mathrm{HU}_{\mathrm{L}} = 24$ and $\Delta\mathrm{HU}_{\mathrm{H}} = 16$.

A thorough validation of *ImaSim* is performed to assure the software to be reliable for this study. The data is validated against XCOM photon cross sections taken at the effective energies corresponding to each photon spectrum. This choice of using effective energies instead of full spectra is based on the nature of the filtered back projection reconstruction that is being used in ImaSim. The attenuation coefficients depend on the energy in each voxel. Since a spectrum is used for simulation, the energy changes along the line of response due to beam hardening. Thus, in filtered back projection, the existence of an effective attenuation coefficient is assumed and by definition different from the average attenuation coefficient over the energy spectrum. The relative attenuation coefficients of the 13 inserts are determined and averaged over circular ROIs. For each spectrum, the effective energy E_{eff} is defined at which the residual differences between simulated and theoretically calculated relative attenuation coefficients (XCOM) are zero on average. The consistency of HU is also evaluated by comparing the simulated data as a function of the phantom shape. The averaged HU over circular ROIs of the 13 inserts are compared between the cylindrical and oval-shaped Gammex 467 calibration phantoms. Differences in HU are used to compare the accuracy of *ImaSim* to clinical tolerances.

To calculate ground truth maps of SPRs in the humanoid phantom, electron densities and tissue compositions of the Woodard and White tissue database[2, 3] are used and equations 1 and 2 are applied pixelwise. For each method, tissue-specific probability distribution functions (PDFs) of SPR errors are determined by comparing predicted SPR values to ground truth pixelwise. The PDFs are then grouped into two types of tissues: 1) soft tissues and 2) bones. This further allows determining the DECT method accuracies to predict SPRs and evaluate the effect on range uncertainties. PDFs in the absence of noise are first used to establish which DECT method is well conditioned for further comparison against SECT. The robustness to noise of the chosen DECT method is evaluated and adapted in order to determine the potential benefit of DECT over SECT in clinical conditions.

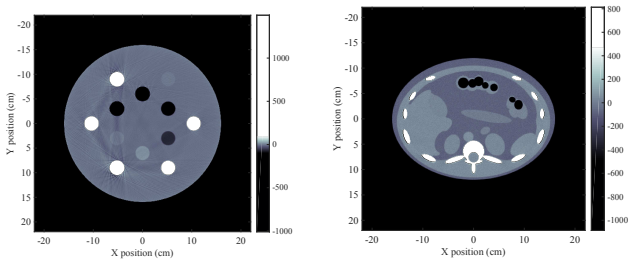
It is worth noting that the ground truth SPR map of the virtual humanoid phantom is not affected by noise or imaging artifacts. However, limitations caused by the CT grid size are left aside by avoiding analyzing data adjacent to interfaces, this way assuring voxels to be homogeneous. Also, because the accuracy of reference values is limited by the Bragg additivity rule, the present study leaves aside uncertainties related to the I-value by (directly or indirectly) using the same rule to predict SPR. This way, the present work focuses mainly on the effect of CT noise and artifacts, leaving the effects of CT grid size and I-value aside.

C. Evaluation of range uncertainties

The impact of the DECT methods on proton beam range uncertainty is evaluated using numerical models. To evaluate the impact in soft tissues, a WEPL-based method [33] is used in combination to SPR error sampling at

TABLE III: List of the 15 human tissues specified by Woodard and White[2, 3] used to simulate CT images and calculate ground truth SPRs. The I-values are calculated using the Bragg additivity rule from the elemental composition using equation 2 and I-values recommended by ICRP[30]

Tissue number	Tissue name	Electron density relative to water	I-value (eV)
1	Adipose tissue	0.951	64.780
2	Adrenal gland	1.025	70.835
3	Aorta	1.038	75.160
4	Blood, whole	1.050	75.203
5	Gallbladder bile	1.026	75.245
6	Kidney	1.040	74.286
7	Liver	1.041	74.355
8	Mammary gland	1.014	70.294
9	Muscle, skeletal	1.040	74.621
10	Ribs 6th and 2nd	1.347	90.722
11	Small intestine wall	1.024	74.285
12	Spleen	1.051	74.980
13	Stomach	1.042	74.194
14	Vertebral column C4	1.355	91.218
15	White matter	1.034	73.126



(a) (b)

FIG. 1: Geometries used for the *ImaSim* simulation: (a) a simulated 140 kVp/Sn CT image of the calibration phantom (resembling the Gammex RMI 467) with added noise ($1\sigma = 16$ HU), showing artifacts reproduced by *ImaSim*; (b) a simulated 100 kVp/Sn CT image of the virtual humanoid phantom geometry.

depth increments of 1 mm, to be consistent with the largest dimension of CT voxels used in the *ImaSim* simulations. For each tissue characterization technique (SECT or DECT), beam range errors are sampled repeatedly by individually sampling SPR errors at each depth increment of 1 mm with PDFs determined from results of the simulated CT images in soft tissues. This way, the performance of the method in extracting SPR from simulated CT images determines the probability distribution of SPR errors. Each statistical sample of range error is calculated analytically from a random array of SPR errors through which the beam is transported. For a given beam energy, depth-dose curves of pristine proton beams are calculated by remapping the depth-dose curve in water, initially calculated with the PSTAR lookup table[34], to the array of WEPL values associated to the random array of SPR error values set in each 1 mm depth increment. That is, to one range error sample corresponds one array of SPR errors set in each depth increment. The calculated range is then compared to the expected range in water (i.e., without SPR errors) to estimate the range error for that random array of SPR errors. In the dose falloff, the final depth increment is reduced to the proton track-end in order for the result not to be influenced by the size of the CT grid. The statistical distributions of beam

range errors in soft tissues are used to estimate the 95% confidence intervals for each tissue characterization technique. The same rationale is used to evaluate the impact of SPR errors in bones and its effect on the range uncertainty. The error in range caused specifically by transport in bones is attributed to its uncertainty in energy loss through them. For a given bone thickness, a number of depth increments is defined (and again set to 1 mm) and a random array of SPR errors is sampled with the PDFs determined in bones from the simulated CT images. Energy loss errors in bones are estimated with Bethe's formula (equation 1). The calculated errors on energy loss are translated into range shift by using the PSTAR energy-range lookup table in water[34] as a function of the beam energy.

III. RESULTS

A. Theoretical comparison of tissue characterization methods

All methods are applied on theoretical attenuation coefficients to predict the SPR of 34 human reference tissues. The residual analysis between predicted and theoretical SPR values is found in table IV. All methods are capable of predicting the SPR of human tissues within 1% under ideal conditions. The methods by Bazalova *et al.* and Bourque *et al.* appear to give the most accurate SPR predictions within a theoretical setup, this considering the negligible bias (i.e., the mean error) and the smallest root mean square error, although Bourque *et al.* contains its errors within the smallest unbiased interval (i.e., less than $\pm 0.4\%$). The method by van Abbema *et al.* introduces a bias to tissues with a high effective atomic number. In their publication, van Abbema *et al.* discovered that effective atomic numbers determined from their method suffer a systematic deviation. Therefore, the authors suggest that their method should only be used for electron density determination.

TABLE IV: Statistics of residual errors of theoretically determined SPRs for 34 human reference tissues using the investigated formalisms.

Method	Min (%)	Max (%)	Mean (%)	RMS (%)
Bazalova *et al.*	-0.47	0.26	-0.02	0.16
Landry *et al.* #1	-0.46	0.33	-0.06	0.20
Landry *et al.* #2	-0.72	0.34	0.17	0.17
Hünemohr *et al.* #1	-0.46	0.33	0.03	0.19
Hünemohr *et al.* #2	-0.43	0.29	0.16	0.16
Bourque *et al.*	-0.38	0.38	0.04	0.16
Van Abbema *et al.*	-0.84	-0.04	-0.30	0.41
Han *et al.*	-0.55	0.60	0.01	0.23
Lalonde and Bouchard	-0.48	0.54	-0.01	0.19

B. Experimental comparison of tissue characterization methods

All methods are used with scanned images to predict the SPR of the Gammex 467 phantom. The results, displayed in table V, are compared to theoretically calculated SPR values for the 13 inserts. The spectral based methods

(Bazalova *et al.*, van Abbema *et al.*) suffer from a systematic bias in the region of higher effective atomic number. This problem was addressed by both authors. Bazalova *et al.* suggested a semi-empirical correction to the subset of data points that are affected by this bias. Van Abbema *et al.* suggest a LWF for every pixel in the image. Although this LWF is applied here, we still observe a bias for higher-Z materials, which was discussed in the paper by van Abbema *et al.* and is addressed above. During our study we found that the calculation of the LWF and the process numerically solving $\mu_{\mathrm{L}}/\mu_{\mathrm{H}}$ requires high computational effort and time.

Calibration-based methods show a good overall performance in a phantom setup. The methods by Landry *et al.* #1 and #2, Hünemohr *et al.* #1 and Bourque *et al.* describe SPRs of phantom materials within ±2%. This residual analysis compares both approaches (spectral- and calibration-based) and is intended to show that spectral-based methods need further consideration to reach the accuracy of calibration based-methods. Despite that both approaches can reach similar theoretical performances (see table IV), the calibration-based methods yield more accurate residuals with experimental data (see table V), since the spectral information is likely not to be representative of the actual spectrum. Also, due to beam hardening effects, the spectrum is not unique in space for all projections. Therefore, one could assume the existence of an effective spectrum giving optimal experimental results. Fitting the spectrum to the experiments would improve the model, but would end up being considered as a calibration-based method. The observed discrepancies between theoretically calculated SPRs (i.e., based on electron densities and compositions provided by the vendor) and those found using the calibrations have three major uncertainty components: 1) experimental uncertainties, 2) uncertainties in the phantom composition and 3) uncertainties in the models themselves. With the residual analysis performed herein, we compare the uncertainties of the models consistently without changing the other first two sources of uncertainties, therefore consistently comparing the models under the same conditions. It is worth noting that the method of Lalonde and Bouchard is designed only to describe human tissues only, as the principal components used in the material decomposition are not applicable to the Gammex phantom materials. This might explain some of the large differences reported in table V, although the method is overall unbiased with a negligible mean error. Also, note that the method of Han *et al.* and is not included in the experimental comparison as the technique requires the use of solutions which was not considered in the present study.

TABLE V: Statistics of residual errors of experimentally determined SPRs of the Gammex 467 calibration phantom using the investigated formalisms.

Method	Min (%)	Max (%)	Mean (%)	RMS (%)
Bazalova *et al.*	-1.49	4.29	0.57	1.67
Landry *et al.* #1	-1.61	1.78	-0.11	0.80
Landry *et al.* #2	-1.52	1.20	-0.12	0.70
Hünemohr *et al.* #1	-1.73	1.25	-0.23	0.81
Hünemohr *et al.* #2	-2.22	1.93	-0.10	1.21
Bourque *et al.*	-1.57	1.12	-0.25	0.68
Van Abbema *et al.*	-2.04	8.55	1.12	3.19
Han *et al.*	-	-	-	-
Lalonde and Bouchard	-2.52	2.82	0.06	1.66

C. Comparison of tissue characterization methods based on simulated CT images

1. Validation of ImaSim

The ability of *ImaSim* to reproduce attenuation coefficients is evaluated on the results obtained with the cylindrical and oval-shaped calibration phantoms. In comparing results of the cylindrical phantom with XCOM cross sections data, the worse case scenario is found for the 100 kVp spectrum ($E_{\text{eff}} = 69.3\,\text{keV}$) with errors ranging from -0.9% to 1.1%, and a root mean square error of 0.7%. The same analysis with experimental data of the Gammex 467 phantom scanned with a Siemens Somatom Flash Definition dual-source CT yields mean absolute errors of range from -1.7% to 1.9%, and a root mean square error of 1.0%, for the 100 kVp spectrum ($E_{\text{eff}} = 71.6\,\text{keV}$). Because experimental data are expected to be higher than numerical simulations due to additional sources of uncertainties, this shows that *ImaSim* is reliable for cylindrical geometries. In its performance with the oval-shaped calibration phantom, the worst discrepancies on average HU values between the cylindrical and oval-shaped phantoms are found to be for the 100 kVp spectrum and range between -2.2 and 0.5 HU as well as 7.1 and 37.5 HU for the plastics equivalent to soft tissues and bones, respectively. However, because only two bones are defined in the virtual humanoid phantom, i.e., vertebral column and ribs, two of the materials in the calibration phantom are out of range in terms of density. Removing these in the analysis yields a maximum discrepancy of 10.4 HU. These results show that the beam hardening correction is acceptable for soft tissues, compared to typical vendor recommendation of ± 4 HU for water. However, errors are slightly higher in bones than expected. But when comparing the oval-shaped results against XCOM cross sections with the same effective energy as found for the cylindrical phantom ($E_{\text{eff}} = 69.3\,\text{keV}$), leaving the two high-density inserts aside (i.e., SB3 and CB2 - 50%) yields errors ranging from -1.7% to 1.3% with a root mean square error of 1.0%. This is comparable to experimental results obtained with the cylindrical Gammex 467 phantom. Therefore, we conclude that *ImaSim* is an acceptable tool for the present study.

2. Estimated probability distribution functions of SPR errors

To reproduce clinical use, only calibration-based methods are used to predict SPRs from simulated DECT images pixelwise. The differences between predicted SPR maps and ground truth SPR values are analyzed. The SECT method proposed by Schneider *et al.* (1996) serves as a gold standard for ρ_e-Z formalisms. PDFs of SPR errors in the absence of noise are displayed in figure 2. The statistics of the methods is summarized in table VI. For soft tissues, all investigated DECT methods predict SPRs with a smaller mean error than the SECT method of Schneider *et al.* (1996), therefore introducing a smaller bias and decreased errors on proton range. Among our implementations, the method by Bourque *et al.* is found to have the smallest mean error, thus introducing a quasi-null bias on proton range prediction, as well as the smallest standard deviation on SPR. For bones, not all DECT methods have a smaller mean error than SECT methods. Three DECT methods introduce a higher bias than the method of Schneider *et al.* (1996), and four introduce a higher bias than the SECT method of Schneider *et al.* (2000), which is shown to

improve the characterization of bones compared to the gold standard SECT. The method of Lalonde and Bouchard was found to have a quasi-null bias and the smallest standard deviation. To determine if the population means of the probability density functions are statistically different, we performed pairwise Welsh's t-tests. In soft tissues, for each pair of PDFs, we found p-values smaller than 10^{-5}, indicating that all distributions are significantly different from each other ($p < 0.05$), with one exception. The distributions derived from Hünemohr *et al.* #1 ($\mu = 0.1068$) and Han *et al.* ($\mu = 0.1148$) are statistically similar ($p = 0.293$). For bones, we found that all distributions are significantly different from each other ($p < 0.05$), with the exception of Hüemohr *et al.* #1 and Landry *et al.* #1 ($p = 0.064$).

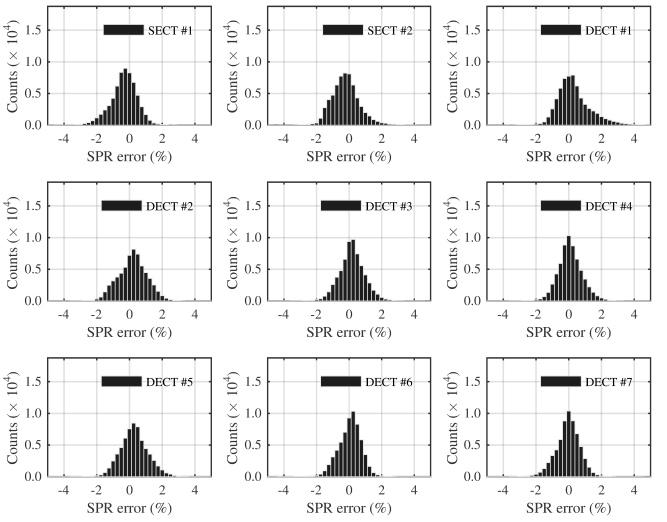
It is worth noting that the methods suitable to predict Monte Carlo inputs (i.e., Landry *et al.* #2, Hünemohr *et al.* #2 and Lalonde and Bouchard) do not perform better in soft tissues than the $\rho_e - Z$ decomposition method of Bourque *et al.*. These results lead to believe that the intermediate step of assigning elemental weight fractions before calculating SPR might not be optimal as it can reduce the accuracy of the estimation. However, the potential improvement on dose calculation using Monte Carlo simulation over analytic tools used commercially are not shown explicitly in these results. Therefore the DECT techniques suitable for Monte Carlo should not be literally compared with the ones suitable with analytic methods.

TABLE VI: Statistics of the PDFs of SPR errors of all investigated tissue characterization methods in the absence of noise: the mean ($\hat{\mu}$) and the standard deviation ($\hat{\sigma}$).

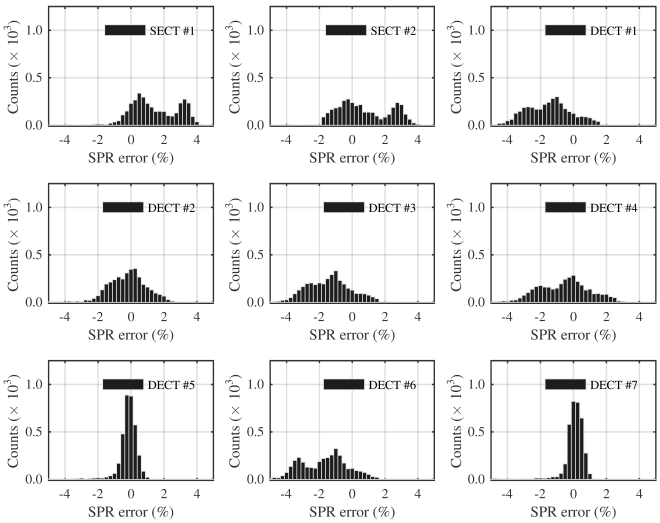
Method	Soft tissues		Bones	
	$\hat{\mu}$ (%)	$\hat{\sigma}$ (%)	$\hat{\mu}$ (%)	$\hat{\sigma}$ (%)
Schneider *et al.* (1996)	-0.43	1.42	1.34	1.61
Schneider *et al.* (2000)	-0.29	1.49	0.65	1.72
Landry *et al.* #1	0.27	1.40	-1.72	1.87
Han *et al.*	0.11	1.34	-0.41	1.71
Hünemohr *et al.* #1	0.11	1.28	-1.64	1.83
Bourque *et al.*	0.02	1.25	-0.77	1.95
Hünemohr *et al.* #2	0.23	1.26	-0.34	1.22
Landry *et al.* #2	-0.04	1.27	-1.92	1.89
Lalonde and Bouchard	-0.13	1.27	-0.08	1.14

D. Proton beam range error estimations

Results are calculated for each noise level separately, i.e., none, low and high. It is worth noting that for accurate estimations of range error confidence intervals, a sufficiently large number of samples is required to get a smooth behaviour of the results as a function of the beam energy and/or bone thickness. The number of samples per method and per beam energy is set to N=2200, totalling 415 800 range error sampling for soft tissues for all 3 levels of noise. For bones, the number of samples is 2 079 000 since five bone thicknesses are investigated, totalling about 2.5 millions of range error samples.



(a)



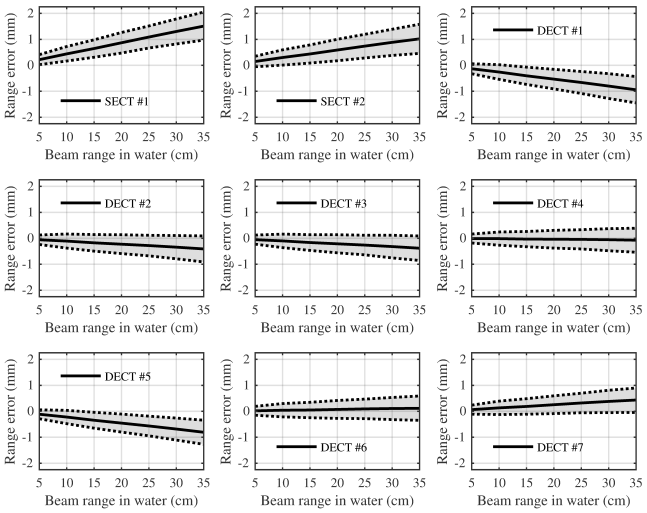
(b)

FIG. 2: PDFs or SPR errors generated with each method applied on the simulated images in the absence of noise: (a) soft tissues and (b) bones. The SECT methods #1 and #2 are Schneider *et al.* (1996) and Schneider *et al.* (2000), respectively. The DECT methods from #1 to #7 are Landry *et al.* #1, Han *et al.*, Hünemohr *et al.* #1, Bourque *et al.*, Hünemohr *et al.* #2, Landry *et al.* #2 and Lalonde and Bouchard, respectively. The display of errors is reduced to within ±5%, although larger errors occur.

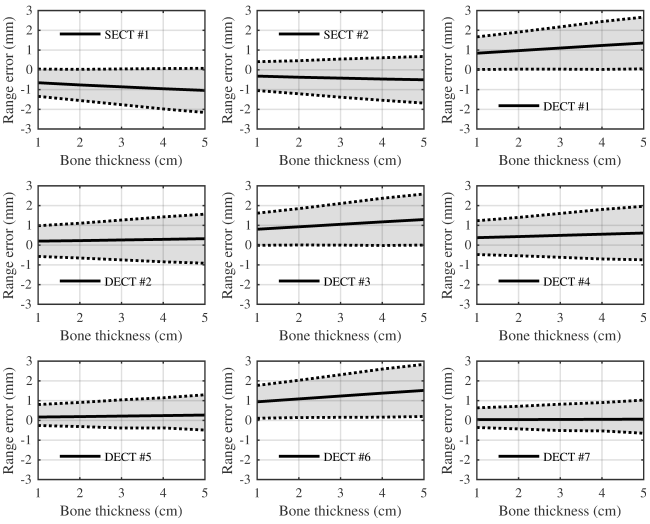
1. Comparison of DECT methods in the absence of noise

Two independent sources of range uncertainties are evaluated from PDFs. The first effect is the range error limited by the precision of SPR predictions in soft tissues. The second effect in the range error caused by proton beam transport through bones before being aimed at a tumour (located in soft tissue). Resulting effects on range errors are shown in figure 3. The effects are consistent with the statistics of the PDFs reported in table VI. In soft tissues, both SECT methods are systematically biased, while our implementations of the DECT methods show smaller bias and 95% range error distribution, with five out of seven methods having low bias: Han *et al.*, Hünemohr *et al.* #1,

Bourque *et al.*, Landry *et al.* #2 and Lalonde and Bouchard. The smallest range errors were found in the method by Bourque *et al.*, with maximal beam range errors within -0.54 mm and 0.39 mm, with a probability of 95%, for beam energies corresponding to ranges in water of up to 35 cm. For the impact of transporting proton beams through bones, both SECT methods are systematically biased, while in our implementations four of out seven DECT methods yield low bias: Han *et al.*, Bourque *et al.*, Hünemohr *et al.* #2 and Lalonde and Bouchard. The smallest range errors were found in the method by Lalonde and Bouchard. It shows maximal beam range errors within -0.91 mm and 1.05 mm (with a probability of 95%) for bone thicknesses up to 5 cm and for beam energies corresponding to a ranges in water of up to 35 cm.



(a)



(b)

FIG. 3: Comparison of estimated range errors in soft tissues from CT data excluding noise for: a) soft tissues, and b) bones. The plain line shows the mean error values and the dotted lines show the boundaries of the 95% confidence intervals of range errors. The energy used the effect in bones is 196 MeV, corresponding to a range in water of 25 cm. The method's numbering is the same as in figure 2.

2. Range uncertainties in clinical conditions: the impact of noise

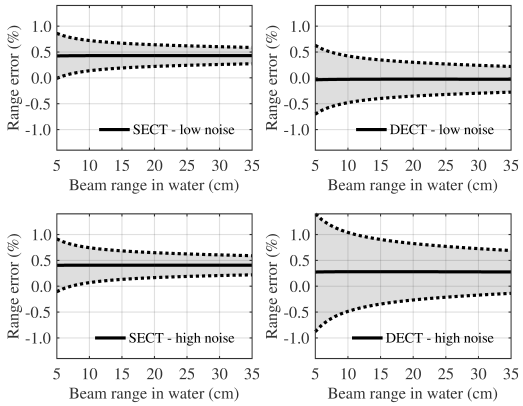
The impact of noise on beam range uncertainties is evaluated by applying the range error estimator models on PDFs calculated with two levels of noise. For soft tissues, the SECT method used is the gold standard method of Schneider *et al.* 1996 and the DECT method is the one of Bourque *et al.*, but adapting its fit parameter of the dual-energy index versus Z to lower order to make it more robust to noise (i.e., $K = 3$ instead of $K = 5$). When noise is present in the image, the values for the dual-energy index can fall out of the calibration domain. By choosing a lower fit order, we are able to control the behavior of the calibration curve outside the calibration domain. An alternative approach would be to use the high fit order ($K = 5$) for values within the calibration domain, and additionally describe values outside the calibration domain using a linear extrapolation. For bones, the SECT method used is Schneider *et al.* 1996 and the DECT used is the method of Lalonde and Bouchard without any modification.

Results are shown in figure 4. The mean errors and boundary values of the 95% confidence interval of range errors in SECT and DECT are compared. For soft tissues, results are displayed as a function of the beam energy, reported in terms of range in water. The comparison shows that for the low level of CT noise, range errors with DECT methods are unbiased compared to SECT, with interval boundary values closer to zero. For the high level of CT noise, the DECT interval is slightly biased and the boundary values are much higher than for SECT, which sensitivity to noise is small. In the absence of noise, maximum range absolute errors with DECT are decreased by about 0.5% relative to the beam range in water, while for the low level of CT noise they are reduced by up to 0.4% relative to the beam range in water. However, for the high level of noise SECT had smaller range uncertainties than DECT, despite its bias in predicting the range.

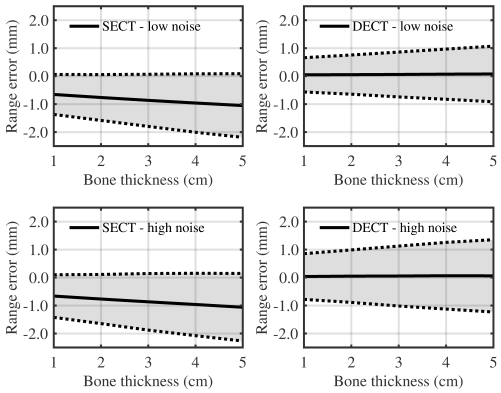
For bones, results are displayed as a function of bone thicknesses through which a 196 MeV beam is transported. The comparison shows that for the low and high levels of noise, DECT errors are unbiased compared the SECT, with interval boundary values closer to zero. Between 1 and 5 cm bone thickness, maximum range absolute errors are reduced by values of up to about 0.6 to 1.1 mm with DECT. For the high level of noise, the same calculations (not shown here) lead maximum range absolute errors reductions between 0.5 and 0.9 mm for bone thicknesses between 1 and 5 cm, respectively.

IV. DISCUSSION AND CONCLUSION

In the present study, the potential of DECT is evaluated over SECT in the context of proton beam range prediction. Nine different techniques are compared in their ability to predict proton SPRs. The methods are implemented and evaluated in three different contexts to evaluate their theoretical foundation (i.e., with XCOM cross sections data), their practicality in a clinical environment (i.e., with experimental measurements) and their performance with a patient-like geometry under constraints of CT artifacts and noise (i.e., *ImaSim* simulations and Gaussian noise model). The first two contexts allow reducing the number of suitable methods to seven. The performance of the



(a)



(b)

FIG. 4: Comparison of range error statistics between SECT and DECT for low and high CT noise in a) soft tissues, and b) bones. The dotted lines represent the boundary values of the 95% confidence interval and the plain line is the average error.

DECT methods with a humanoid phantom is first estimated in the absence of noise to allow choosing techniques being the most robust to CT artefacts, i.e., Bourque *et al.* for soft tissues and Lalonde and Bouchard for bones. It is worth noting that all methods are implemented with the best of our knowledge, based on the publications available in literature. We use the theoretical comparison based on XCOM data, as well as the experimental data, as an indicator to assure that the methods are implemented properly. Our results reproduce values that were quoted by the authors of each method, leading to the conclusion that all methods should be correctly implemented.

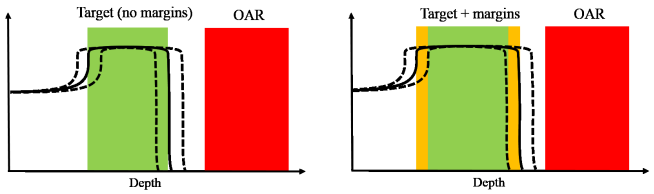
The most clinically-relevant results of this study are the ones where CT artifacts and noise are present. Range error estimations clearly demonstrate the advantages of DECT over SECT in the presence of low CT noise, since SECT is generally more robust to noise due to the mathematical nature of its techniques (i.e., linear models). Overall, one could expect DECT to reduce range uncertainties (to the 95% confidence level) by about 0.4% in soft tissues, and up to about 1 mm for beams of therapeutic energies transported through bones. For high levels of CT noise, the benefits of DECT can be lost over the robustness of SECT in soft tissues. While this is expected due to the mathematical complexity of DECT techniques, it is yet to be demonstrated that some techniques could be further adapted for high

CT noise. For instance, Bourque *et al.* is used in soft tissues with minimal adaptation (i.e, just changing $K = 5$ to $K = 3$ in the dual-energy index conversion to Z), and this could explain why it is only robust to low noise levels. As for the method of Lalonde and Bouchard, it is surprising that despite no adaptation it stills outperforms the SECT gold standard for low or high noise levels. This could suggest that an effort in adapting the method for the presence of noise could yield even better results. The results suggest that DECT-predicted SPR can benefit from an increase in mAs defined in the scanning protocol. Therefore we recommend to investigate SPR uncertainties before establishing a clinical DECT protocol for radiotherapy planning. We would like to emphasize that errors arising from spectral differences between these calibration and patient scan are not taken into account here. Therefore we recommend to perform the calibration for each scanner model and scanning protocol individually.

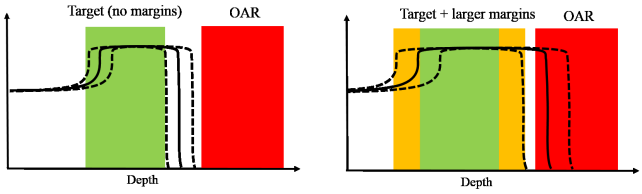
While the benefits of DECT over SECT are expected to be improved by refined robustness to noise, one could also seek for more sensible values in range uncertainties to be obtained with a more realistic dose calculation model, such as Monte Carlo simulations. However, performing such a study with Monte Carlo transport simulations is rather difficult, yet impossible, as a high number of range error samples is required (i.e., nearly 2.5 millions in this study), which in the context of cross sections become multidimensional rather than simply the SPR error, requiring to redefine a set of materials and a full calculation (with millions of histories) for each sample. Nonetheless, it is quite conceivable that the numbers estimated in the present study are realistic due to the consistency of the methods. The simulation of CT images using *ImaSim* has the advantage of allowing SPR estimation with various techniques in a controlled and consistent environment, with focus on CT artifacts and noise, leaving the effects of CT grid size, uncertainties in I-values and other sources aside. Finally, while the WEPL-based model is not entirely accurate, it is still used consistently and therefore should yield correct estimations of errors.

A simplified interpretation of the results presented in the present study allows comparison with the topical review by Paganetti[6]. In that publication, uncertainties in CT conversion to tissue as well as CT imaging and calibration each contribute to 0.5% of the range uncertainty (1.5σ), and the overall uncertainty recommended for proton beam range is 2.7% + 1.2 mm. While adding the two uncertainty sources in quadrature yields about 0.9% for a significance level of 95%, this value corresponds to the maximum error found in the present study at the highest noise level for SECT. From this, we could conclude that the recommended uncertainty with DECT should be reduced to 2.4% + 1.2 mm (i.e., reporting the 95% level of confidence to a statistical significance of 1.5σ). But a closer look at the present results suggest a deeper analysis, which is addressed in figure 5. Here we illustrate the main advantage unbiased range errors, as it allows reducing the size of the margins. Indeed, DECT have the advantage of reducing uncertainties as only of the interval boundary needs to be considered as an uncertainty for each direction with respect to the beam, conversely to using the maximum absolute error in SECT.

The method proposed in the present study provides a more detailed estimation of range uncertainties than more simplistic rules used in the clinic (i.e., set to 3.5% of the range in water for all energies). An interesting result in figure 4a shows that range uncertainties relative to the beam range in water are larger for smaller energies. This can be explained by the fact that the smaller the energy, the smaller the amount of voxels contribute to the average



(a)



(b)

FIG. 5: Illustration of the effect of range uncertainties on the definition of margins adjacent to still targets in two situations: a) unbiased range uncertainties, and b) biased range uncertainties. In each figure, the left graph shows the spread-out Bragg peak (SOBP) adapted for the target, while in the second the SOBP it is adapted to the target plus margins, accounting for range uncertainties. In this example, the systematic bias of the error doubles the size of the margins and compromises OAR sparing.

SPR. This way, the uncertainty on the average SPR is inversely proportional to the square root of the number of voxels traversed. And because the range relative to that of water equals the inverse of the average SPR, with a few manipulations we show that the relative range uncertainty is given by [13]

$$\frac{\Delta R}{R} = \frac{\Delta \mathrm{SPR}_{\mathrm{ave}}}{\mathrm{SPR}_{\mathrm{ave}}} = \sqrt{\frac{\Delta x}{R}}\frac{\Delta \mathrm{SPR}}{\mathrm{SPR}_{\mathrm{ave}}}, \tag{4}$$

with Δx the size of the voxels in which SPR values are assumed homogeneous and ΔSPR the uncertainty on SPR in each voxel. This relation predicts that for a fixed CT grid size and uncertainty on SPR the relative range uncertainty in soft tissue (where $\mathrm{SPR}_{\mathrm{ave}}$ is approximately constant) is inversely proportional to the square root of the range, which is consistent with results shown in figure 4a.

Finally, although it could be possible to improve SECT methods, notably by using Schneider *et al.* 2000 or attempting to correct for the bias, the present study suggest that DECT can go beyond the capabilities of SECT in the context of proton therapy. However, noise remains a major limiting factor and needs to be carefully addressed if the patient imaging dose is to be kept to the same level as in conventional radiotherapy treatment planning. We conclude that DECT has substantial potential for reducing range uncertainties in proton therapy and that further developments of DECT methods should focus on their robustness to noise, since mathematical formalisms might have found their full maturity at the present time. Also, it is expected that DECT methods based on raw-data should enable the reduction of CT artifacts, and therefore range uncertainties. Moreover, improvements in CT grid size (i.e., such in future developments in spectral CT) could help improving the precision of proton therapy planning.

Acknowledgements

We would like to acknowledge Danis Blais and Guy Charron from CHUM for their help with the experimental data. We also like to thank Dr Stéphane Bedwani from CRCHUM for his ideas on the beam hardening correction. This work was funded by EPSRC UK (case studentship No. 20873) and the National Physical Laboratory (UK).

References

[1] 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.

[2] DR White, HQ Woodard, and SM Hammond. Average soft-tissue and bone models for use in radiation dosimetry. *The British journal of radiology*, 60(717):907–913, 1987.

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

[4] 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.

[5] 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.

[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] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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, 2014.

[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] JK van Abbema, M-J van Goethem, MJW Greuter, A van der Schaaf, S Brandenburg, and ER van der Graaf. Relative electron density determination using a physics based parameterization of photon interactions in medical dect. *Physics in medicine and biology*, 60(9):3825, 2015.

[15] 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.

[16] 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.

[17] J D Evans, B R Whiting, J A O'Sullivan, D G Politte, P H Klahr, Y Yu, and J F Williamson. Prospects for in vivo estimation of photon linear attenuation coefficients using postprocessing dual-energy CT imaging on a commercial scanner: Comparison of analytic and polyenergetic statistical reconstruction algorithms. *Medical physics*, 40(12):121914, 2013.

[18] J-É Tremblay, S Bedwani, and H Bouchard. A theoretical comparison of tissue parameter extraction methods for dual energy computed tomography. *Medical physics*, 41(8):081905, 2014.

[19] WV Mayneord. The significance of the roentgen. *Acta of the International Union against Cancer*, 2:271, 1937.

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

[21] M Torikoshi, T Tsunoo, M Sasaki, M Endo, Y Noda, Y Ohno, T Kohno, K Hyodo, K Uesugi, and N Yagi. Electron density measurement with dual-energy x-ray ct using synchrotron radiation. *Physics in medicine and biology*, 48(5):673, 2003.

[22] M.J. Berger et al. Xcom: Photon cross section database (version 1.5). Online, 2010. National Institute of Standards and Technology, Gaithersburg, MD.

[23] RA Rutherford, BR Pullan, and I Isherwood. Measurement of effective atomic number and electron density using an emi scanner. *Neuroradiology*, 11(1):15–21, 1976.

[24] 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.

[25] 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.

[26] RE Alvarez and A Macovski. Energy-selective reconstructions in x-ray computerised tomography. *Physics in medicine and biology*, 21(5):733, 1976.

[27] 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.

[28] DF Jackson and DJ Hawkes. X-ray attenuation coefficients of elements and mixtures. *Physics Reports*, 70(3):169–233, 1981.

[29] ICRU. Tissue substitutes in radiation dosimetry and measurement. *ICRU report 44*, 1989.

[30] International Commission on Radiological Protection. Task Group and Walter Stephen Snyder. *Report of the task group on reference man*, volume 23. Pergamon Oxford, 1975.

[31] G Landry, F DeBlois, M Bazalova, and F Verhaegen. Mo-d-303a-06: Imasim, an animated tool for teaching imaging. *Medical Physics*, 36(6):2696–2697, 2009.

[32] G Landry, B Reniers, PV Granton, B van Rooijen, L Beaulieu, JE Wildberger, and F Verhaegen. Extracting atomic numbers

575 and electron densities from a dual source dual energy ct scanner: experiments and a simulation model. *Radiotherapy and*
576 *Oncology*, 100(3):375–379, 2011.
577 [33] Barbara Schaffner. *Range precision of therapeutic proton beams*. PhD thesis, Diss. Naturwiss. ETH Zürich, Nr. 12474,
578 1997. Ref.: Peter Niederer; Korref.: Peter Rüegsegger; Korref.: Eros Pedroni, 1997.
579 [34] Martin J Berger. Estar, pstar, and astar: Computer programs for calculating stopping-power and range tables for electrons,
580 protons, and helium ions. *Unknown*, 1992.