Joint Activity and Attenuation Reconstruction From Multiple Energy Window Data With Photopeak Scatter Re-Estimation in Non-TOF 3-D PET

Ludovica Brusaferri, *Student Member, IEEE*, Alexandre Bousse, Elise C. Emond, *Student Member, IEEE*, Richard Brown, Yu-Jung Tsai, David Atkinson, Sébastien Ourselin, Charles C. Watson, Brian F. Hutton, *Senior Member, IEEE*, Simon Arridge, and Kris Thielemans, *Senior Member, IEEE*

***Abstract*—Estimation of attenuation from positron emission tomography (PET) data only is of interest for hybrid PET-MR and systems where CT is not available or recommended. However, when using data from a single energy window, emission-based non-time-of-flight (TOF) PET attenuation correction (AC) methods suffer from "cross-talk" artifacts. Based on earlier work, this article explores the hypothesis that cross-talk can be reduced by using more than one energy window. We propose an algorithm for the simultaneous estimation of both activity and attenuation images, as well as, the scatter component of the measured data from a PET acquisition, using multiple energy windows. The model for the measurements is 3-D and accounts for the finite energy resolution of PET detectors; it is restricted to single scatter. The proposed energy-based simultaneous maximum likelihood reconstruction of activity and attenuation with photopeak scatter re-estimation algorithm is compared with simultaneous estimation from a single energy window simultaneous maximum likelihood reconstruction of activity and attenuation with photopeak scatter re-estimation. The evaluation is based on simulations using the characteristics of the Siemens mMR scanner. Phantoms of different complexity were investigated. In particular, a 3-D XCAT torso phantom was used to assess the inpainting of attenuation values within the lung region. Results show that the cross-talk present in non-TOF maximum likelihood reconstruction of activity and attenuation reconstructions is significantly reduced when using multiple energy windows and indicate that the proposed approach warrants further investigation.**

***Index Terms*—Attenuation estimation, image reconstruction, iterative methods, optimization, positron emission tomography, quantification and estimation, scatter.**

Manuscript received December 1, 2019; revised January 25, 2020 and February 20, 2020; accepted March 2, 2020. Date of publication March 4, 2020; date of current version July 2, 2020. This work was supported in part by the Siemens Healthineers, in part by the UCL Impact Scheme, in part by the NIHR UCLH Biomedical Research Centre, in part by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging under Grant EP/L016478/1, and in part by the UCL Institute of Healthcare Engineering. The work of Elise C. Emond was supported by GlaxoSmithKline under Grant BIDS3000030921. *(Corresponding author: Ludovica Brusaferri.)*

Ludovica Brusaferri, Elise C. Emond, Richard Brown, Brian F. Hutton, and Kris Thielemans are with the Institute of Nuclear Medicine, University College London, London NW1 2BU, U.K. (e-mail: l.brusaferri.16@ucl.ac.uk).

Alexandre Bousse was with the Institute of Nuclear Medicine, University College London, London NW1 2BU, U.K. He is now with the Laboratoire de Traitement de l'Information Médicale, INSERM, Université de Bretagne Occidentale, U1101 Brest, France.

Yu-Jung Tsai was with the Institute of Nuclear Medicine, University College London, London NW1 2BU, U.K. He is now with the Yale PET Center, Yale University, New Haven, CT 06520 USA.

David Atkinson is with the Centre for Medical Imaging, University College London, London W1W 7TS, U.K.

Sébastien Ourselin is with the Department of Medical Physics and Biomedical Engineering, Kings College London, London WC1E 6BT, U.K.

Charles C. Watson is with Siemens Medical Solutions-Molecular Imaging, Knoxville, TN 37932 USA.

Simon Arridge is with the Department of Computer Science, University College London, London WC1E 6BT, U.K.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRPMS.2020.2978449

I. Introduction

IN QUANTITATIVE positron emission tomography (PET) imaging, photon attenuation, and Compton scatter must be taken into account. Errors in the attenuation image can significantly affect PET quantification, especially, in tissues such as the lung, where density values vary considerably among patients (up to a factor 2), and during the respiratory cycle [1], [2]. Although the problem of attenuation correction (AC) of PET images can be regarded as solved to a large extent for hybrid PET/CT scanners, it still represents an issue for PET/MR systems as mapping MR image intensities to PET 511 keV attenuation coefficients is extremely challenging in the thorax [3]. In current clinical practice, MR-based AC (MR-AC) consists of segmenting MR images into three or four tissue classes—namely, air, lung, and fat and soft-tissue—followed by the assignment of population-based density values to each tissue class. One of the main limitations of this approach is the neglect of the inter-patient/intrapatient heterogeneity of attenuation coefficients within each class [4]. These methods are, however, very successful in brain imaging, where intrapatient variation is relatively small. Other (MR-AC) methods include atlas/mapping techniques [5] and MR/CT learning [6]. These methods can benefit from different MR sequences, such as UTE [7] and ZTE [8] offering improved soft-tissue contrast and the possibility of an accurate bone segmentation. However, these advantages come at the cost of longer acquisition times compared to other MR sequences [9]. For thorax acquisitions, (MR-AC) methods are prone to errors, especially, in the lung,

due to high variability of attenuation values in the lung tissue both on a regional basis and from person to person [1]. More details can be found in a couple of recent review articles [10], [11], focusing, respectively, on brain and thorax imaging.

Attenuation estimation strategies from PET data seem particularly promising for overcoming the quantification errors induced by conventional MR-based approaches [12]. Methods that aim to estimate attenuation from the emission data can be essentially divided into analytic and iterative approaches. The first relies on the consistency conditions of the attenuated radon transform [13] and offers a direct mathematical solution from the known projections. This leads to relatively fast reconstruction techniques, but with the major drawback of not modeling the statistical variability of the emission data, as well as showing significant artifacts when the number of counts decreases. The second type of approach aims to find a solution by successive estimates to fit the measured PET data. In addition to adequately modeling the statistical nature of the data, the other advantages offered over the analytic algorithms include the possibility of modeling more complicated system geometries and physical processes such as Compton scattering.

The most popular method amongst the iterative algorithms is maximum likelihood reconstruction of activity and attenuation (MLAA) introduced in Nuyts *et al.* [14], a (penalized) maximum likelihood-based iterative algorithm that alternates between activity and attenuation estimation. However, in the absence of time-of-flight (TOF) information, the joint estimation problem is strongly ill-posed [15]; as a consequence, the activity and attenuation images estimated from non-TOF-MLAA suffer from cross-talk artifacts, where the features of the activity map propagate into the attenuation map and vice versa. On the other hand, TOF-MLAA has great potential for PET/MR applications [16], [17], but recent work has shown that it is necessary to re-estimate the photopeak scatter during TOF-MLAA iterations for best results [18].

Attenuation and scatter are intrinsically linked both on a physical level and when deriving the scatter and attenuation estimates. Scattered events are normally estimated by a 3-D model-based simulation [19], [20]. Quantitative errors in the attenuation image propagate in the scatter estimation, and, therefore, in the reconstructed activity distribution. This effect is non-negligible in the thorax, as scatter events can represent up to 40% of the total recorded coincidences [21].

This linking has led several authors to attempt to use the information contained in the scattered counts to estimate attenuation. As Compton scattering decreases the energy of the scattered photon, this could be achieved by using data acquired in several energy windows. Energy-based methods for attenuation estimation have first been investigated in single photon emission computed tomography (SPECT) [22]–[24] using an upper (photopeak) and a lower (scatter) energy window. However, all of these approaches assumed the possibility of distinguishing gamma rays exiting the patient that have not been scattered from those that are scattered.

The idea of deriving additional information from scatter has also been applied to PET [25], [26]. Although initial studies were restricted to simple 2-D phantoms and perfect energy resolution such that scattered events can be distinguished from those that are not scattered, the possibility of reconstructing a 3-D attenuation distribution from scattered data only (with known activity) was demonstrated with realistic energy measurement scenarios [27]. Recently, Berker *et al.* [28], [29] proposed a joint reconstruction algorithm from scatter and unscattered data. To be able to handle high attenuating (or large) objects, a "four-step algorithm" was proposed alternating between various activity and attenuation image reconstruction steps. However, the convergence of an alternating algorithm with each step optimizing a different objective function can be problematic [30]. In addition, the evaluation of [29] used 2-D phantoms and disregarded energy-measurement uncertainties. In practice, however, current PET scanners have an energy resolution in the order of 10% or worse, leading to the presence of scatter in the photopeak window and reduced information content in the scattered photons. Furthermore, PET scatter is inherently 3-D.

In this article, we investigate the feasibility of a more practically relevant method for the joint reconstruction activity and attenuation distributions from multiple energy window measurements by using a maximum likelihood framework. Particular interest was given to the inpainting of the attenuation values within the lung region; this was investigated with 3-D phantom simulations where the attenuation values outside the lung were assumed to be known.

Overall, the improvements on previous research studies include: 1) accounting for the uncertainty in the energy measurements to a large extent; 2) considering the presence of both scattered and unscattered events in the photopeak window; 3) simulating 3-D input and output according to existing scanner geometry and specifications; and 4) optimizing one unique objective function.

This article is organized as follows. We first cover the mathematical theory relevant to the framework, then give an overview of the proposed algorithm (with some details in the appendices). We finally present results from simulated data and provide a comparison of the proposed method against MLAA from a single energy window acquisition. Similarly to previous published work [28]–[30], the current study is restricted to single scatter only.

II. Theory

A. Objective Function

A reasonable statistical model for PET measurements is to describe the measured data $\boldsymbol{g}$ as independent Poisson variables. Assuming that the scanner allows energy discrimination of the detected photons

$$g_b \sim \text{Poisson}\{\bar{g}_b(\boldsymbol{\lambda}, \boldsymbol{\mu})\}, \quad b = 1, 2, \ldots, B \tag{1}$$

where B is the number of detection bins, characterized by their detector pair and energy window pair, $\boldsymbol{\lambda} \in \mathbb{R}^{n_v}$ and $\boldsymbol{\mu} \in \mathbb{R}^{n_v}$ are vectors that represent the activity and attenuation distributions of the object, respectively, and $\bar{g}_b(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is the expected value of the bth measurement. Taking the logarithm and ignoring the terms independent of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, the log-likelihood of the

measured data $\boldsymbol{g} \triangleq [g_1, \ldots, g_B]$ is given by

$$\mathcal{L}(\boldsymbol{g} \mid \bar{\boldsymbol{g}}(\boldsymbol{\lambda}, \boldsymbol{\mu})) = \sum_b g_b \log \bar{g}_b(\boldsymbol{\lambda}, \boldsymbol{\mu}) - \bar{g}_b(\boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{2}$$

The joint maximum-likelihood reconstruction of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ is traditionally achieved by solving

$$\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right) = \underset{\boldsymbol{\lambda}\geq\mathbf{0}, \boldsymbol{\mu}\geq\mathbf{0}}{\arg\max}\, \mathcal{L}(\boldsymbol{g} \mid \bar{\boldsymbol{g}}(\boldsymbol{\lambda}, \boldsymbol{\mu})). \tag{3}$$

B. Optimization

A common approach for maximizing a joint likelihood such as (3) consists of updating $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ in an alternating order

$$\begin{aligned} \boldsymbol{\lambda}^{k+1} &= \underset{\boldsymbol{\lambda}\geq\mathbf{0}}{\arg\max}\, \mathcal{L}\left(\boldsymbol{g} \mid \bar{\boldsymbol{g}}\left(\boldsymbol{\lambda}, \boldsymbol{\mu}^k\right)\right) \\ \boldsymbol{\mu}^{k+1} &= \underset{\boldsymbol{\mu}\geq\mathbf{0}}{\arg\max}\, \mathcal{L}\left(\boldsymbol{g} \mid \bar{\boldsymbol{g}}\left(\boldsymbol{\lambda}^{k+1}, \boldsymbol{\mu}\right)\right). \end{aligned} \tag{4}$$

This requires two inner subalgorithms to estimate $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ separately. Alternatively, both variables can be updated simultaneously—similar to [31]—to avoid complications related to the settings of inner loop parameters

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\geq\mathbf{0}}{\arg\max}\, \mathcal{L}(\boldsymbol{g} \mid \bar{\boldsymbol{g}}(\boldsymbol{\theta})) \tag{5}$$

where $\boldsymbol{\theta} = [\boldsymbol{\lambda}, \boldsymbol{\mu}] \in \mathbb{R}^{2n_v}$. The latter approach is used in this article.

C. Optimization of the Scatter Component

In PET, the expected counts $\bar{\boldsymbol{g}}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ are often expressed as

$$\bar{\boldsymbol{g}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{A}(\boldsymbol{\mu})\boldsymbol{\lambda} + \bar{\boldsymbol{g}}^{\text{sc}} + \bar{\boldsymbol{g}}^{r} \tag{6}$$

where $\mathbf{A}(\boldsymbol{\mu}) \in \mathbb{R}_+^{n_D \times n_v}$ is a matrix mapping from image space to data space, denoting the probability of detecting nonscattered coincidences, n_D is the number of detector pairs, n_v is the number of voxels in the image, and $\bar{\boldsymbol{g}}^{\text{sc}}$ and $\bar{\boldsymbol{g}}^{r}$ are the expected scatter and random sinograms, respectively. The scatter component is generally considered as a background term, here, we account for its dependency on the activity and attenuation distributions. This results in $\bar{\boldsymbol{g}}^{\text{sc}}$ being replaced by $\bar{\boldsymbol{g}}^{\text{sc}}(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

D. Multiple Energy Window Acquisition Model

In the proposed method, we assume that each photon of a photon pair is assigned to either the photopeak window (U) or to a lower energy window (L), resulting in the measurement of four different 3-D sinograms, one for each energy window combination ($\boldsymbol{g}_{\text{UU}}$, $\boldsymbol{g}_{\text{UL}}$, $\boldsymbol{g}_{\text{LU}}$, $\boldsymbol{g}_{\text{LL}}$). For all $(w, y) \in \{\text{U}, \text{L}\}^2$, the observed counts $\boldsymbol{g}_{wy}$ can be described as a Poisson process centered in $\bar{\boldsymbol{g}}_{wy}$, given by the sum of expected scattered $\bar{\boldsymbol{g}}^{\text{sc}}_{\text{wy}}$ and unscattered events $\bar{\boldsymbol{g}}^{\text{unsc}}_{\text{wy}}$. In this article, we assume that the lower energy window excludes any *unscattered* events. In addition, we restrict the study to single scatter coincidences. Therefore, we disregard $\boldsymbol{g}_{\text{LL}}$ as it is expected to contain few single scatter events [27]. The expected counts are therefore given by

$$\begin{aligned} \bar{\boldsymbol{g}}_{\text{UU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \bar{\boldsymbol{g}}^{\text{unsc}}_{\text{UU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \bar{\boldsymbol{g}}^{\text{sc}}_{\text{UU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \bar{\boldsymbol{g}}^{r}_{\text{UU}} \\ \bar{\boldsymbol{g}}_{\text{UL}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) &\approx \bar{\boldsymbol{g}}^{\text{sc}}_{\text{UL}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \bar{\boldsymbol{g}}^{r}_{\text{UL}} \\ \bar{\boldsymbol{g}}_{\text{LU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) &\approx \bar{\boldsymbol{g}}^{\text{sc}}_{\text{LU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \bar{\boldsymbol{g}}^{r}_{\text{LU}}. \end{aligned} \tag{7}$$

The following sections cover the forward scattered and unscattered model used in this article study.

1) Unscattered Events: In PET, the expected photopeak unscattered events $\bar{\boldsymbol{g}}^{\text{unsc}}_{\text{UU}}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ are often expressed as

$$\bar{\boldsymbol{g}}^{\text{unsc}}_{\text{UU}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{A}(\boldsymbol{\mu})\boldsymbol{\lambda} \tag{8}$$

where $\mathbf{A}(\boldsymbol{\mu})$ is the detection probability matrix, taking the attenuation into account.

2) Scattered Events: The forward model for the scatter is an extension of the single scatter simulation (SSS) model proposed in [19] to the case of a multiple energy window acquisition (see Appendix A for details). For computational efficiency, the scatter simulation is performed in low spatial resolution.

Let $\mathbf{P} \in \mathbb{R}_+^{n_D \times n_{D'}}$ be a prolongation operator that maps from low resolution (n'_D total number of detector pairs) to high resolution sinograms (n_D total number of detector pairs). Then, the scatter component is given by

$$\bar{\boldsymbol{g}}^{\text{sc}}_{wy}(\boldsymbol{\theta}) = \mathbf{P}\, \mathbf{S}_{wy}(\boldsymbol{\theta}) \tag{9}$$

with $\mathbf{S}_{wy}(\boldsymbol{\theta})$ indicating an operator that computes the expected scatter at each energy window pair (w, y), defined in Appendix A. For the results presented in this article, the prolongation operator $\mathbf{P}$ consists of a cubic B-spline interpolation.

III. Algorithms

In this section, we describe the algorithms used for the joint reconstruction of the activity and the attenuation images. Inputs for the reconstruction are the measured data $\boldsymbol{g}$ and an estimate of the μ-map, for instance on a PET-MR scanner obtained via MR-based attenuation correction (MR-AC).

A. Initialization via OSEM/SSS

Initial activity $\boldsymbol{\lambda}^{\text{init}}$ and photopeak scatter estimates $\hat{\boldsymbol{g}}^{\text{sc,init}}_{\text{UU}} \approx \bar{\boldsymbol{g}}^{\text{sc}}_{\text{UU}}(\boldsymbol{\theta}^{\text{init}})$, with $\boldsymbol{\theta}^{\text{init}} = [\boldsymbol{\lambda}^{\text{init}}, \boldsymbol{\mu}^{\text{init}}]$, are obtained from the photopeak data as follows: 1) set initial scatter estimate to zero; 2) reconstruct the activity image with OSEM (7 subsets, 70 subiterations); 3) estimate photopeak scatter with SSS. This process is repeated iteratively (see Algorithm 1).

B. MLAA-EB-S

Here, we describe the main energy-based simultaneous maximum likelihood reconstruction of activity and attenuation with photopeak scatter re-estimation (MLAA-EB-S), summarized in Algorithm 2. It can be seen as an evolution of MLAA-EB [30], improved on two main aspects: 1) the algorithm optimizes one unique objective function and 2) the activity and attenuation images are updated simultaneously. In particular, both unknown distributions $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are reconstructed from all the available data: $\boldsymbol{g}_{\text{UU}}$, $\boldsymbol{g}_{\text{UL}}$, and $\boldsymbol{g}_{\text{LU}}$.

Algorithm 1: Pseudo-Code for OSEM/SSS

Input: g_{UU}, $\boldsymbol{\mu}^{\mathrm{init}}$, $\hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,init}} = 0$
Output: Initial activity estimate $\boldsymbol{\lambda}^{\mathrm{init}}$

$\hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc},0} \leftarrow \hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,init}}$
for $i = 0, \ldots,$ MaxOSEMandSSSIter $- 1$ **do**
  $\boldsymbol{\lambda}^{i} \leftarrow$ **OSEM**$(\boldsymbol{g}_{\mathrm{UU}}, \boldsymbol{\mu}^{\mathrm{init}}, \hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,i}})$
  $\hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,i+1}} \leftarrow \mathbf{PS}_{\mathrm{UU}}(\boldsymbol{\lambda}^{\mathrm{i}}, \boldsymbol{\mu}^{\mathrm{init}})$
end
$\boldsymbol{\lambda}^{\mathrm{init}} \leftarrow \boldsymbol{\lambda}^{\mathrm{MaxOSEMandSSSIter}-1}$

Algorithm 2: Pseudo-Code for MLAA-EB-S

Input: $\boldsymbol{g}_{\mathrm{UU}}$, $\boldsymbol{g}'_{\mathrm{UL}}$, $\boldsymbol{g}'_{\mathrm{LU}}$, $\boldsymbol{\lambda}^{\mathrm{init}}$, $\boldsymbol{\mu}^{\mathrm{init}}$, $\hat{\boldsymbol{g}}_{\mathrm{UU}}^{sc,\mathrm{init}}$
Output: Estimated activity and attenuation images vector $\boldsymbol{\theta}^{\mathrm{est}}$

$\boldsymbol{\theta}_0^0 \leftarrow [\boldsymbol{\lambda}^{\mathrm{init}}, \boldsymbol{\mu}^{\mathrm{init}}]$
$\hat{\boldsymbol{g}}_{\mathrm{UU}}^{sc,0} \leftarrow \hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,init}}$
for $t = 0, \ldots,$ MaxOuterIter $- 1$ **do**
  for $k = 0, \ldots,$ MaxInnerIter $- 1$ **do**
    $\boldsymbol{\theta}_{k+1}^{t} \leftarrow$ **LBFGS-B**$(\boldsymbol{g}_{\mathrm{UU}}, \boldsymbol{g}'_{\mathrm{UL}}, \boldsymbol{g}'_{\mathrm{LU}}, \boldsymbol{\theta}_k^t, \hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,t}})$
  end
  $\boldsymbol{\theta}_0^{t+1} \leftarrow \boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{t}$
  $\hat{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc,t+1}} \leftarrow \mathbf{PS}_{\mathrm{UU}}(\boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{t})$
end
$\boldsymbol{\theta}^{\mathrm{est}} \leftarrow \boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{\mathrm{MaxOuterIter}-1}$

Special attention was given to reducing computational effort. The scatter gradient is computed during the reconstruction only for the UL and LU windows. The photopeak scatter estimate $\bar{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc}}$ is iteratively updated via a one-step-late approach. In addition, the input data in the low energy windows were downsampled to low resolution, given the presence of only scattered events in UL and LU (7).

Image Updates: The optimization strategy follows (5), leading to the simultaneous estimation of the two variables $[\boldsymbol{\lambda}, \boldsymbol{\mu}] = \boldsymbol{\theta}$, using a previous estimate of the scatter in the photopeak window. The objective function is given by the sum of each log-likelihood at a given energy window pair

$$\mathcal{L}^{\mathrm{tot}}(\boldsymbol{\theta}) \triangleq \mathcal{L}\big(\boldsymbol{g}_{\mathrm{UU}} \mid \bar{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{unsc}}(\boldsymbol{\theta}) + \bar{\boldsymbol{g}}_{\mathrm{UU}}^{\mathrm{sc}}\big(\boldsymbol{\theta}^{\mathrm{prev}}\big)\big) + \mathcal{L}\big(\boldsymbol{g}'_{\mathrm{UL}} \mid \gamma\, \mathbf{S}_{\mathrm{UL}}(\boldsymbol{\theta})\big) + \mathcal{L}\big(\boldsymbol{g}'_{\mathrm{LU}} \mid \gamma\, \mathbf{S}_{\mathrm{LU}}(\boldsymbol{\theta})\big) \quad (10)$$

with g'_{wy} and γ denoting, respectively, the downsampled sinogram and the down-sampling factor, and $\boldsymbol{\theta}^{\mathrm{prev}}$ indicating an estimate of $\boldsymbol{\theta}$ at previous iteration. Please note that $[(\sum_{n_D} g)/(\sum_{n'_D} g' = \gamma)]$.

The activity and the attenuation estimates are updated with bounded limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) [32]. Every update consists of a line-search step in a quasi-Newton direction

$$\boldsymbol{\theta}^{k} = \boldsymbol{\theta}^{k+1} - \alpha \mathbf{B}_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} \mathcal{L}^{\mathrm{tot}} \quad (11)$$

Algorithm 3: Pseudo-Code for MLAA-S

Input: $\boldsymbol{g}_{ww}$, $\boldsymbol{\lambda}^{\mathrm{init}}$, $\boldsymbol{\mu}^{\mathrm{init}}$, $\hat{\boldsymbol{g}}_{ww}^{sc,\mathrm{init}}$
Output: Estimated activity and attenuation images vector $\boldsymbol{\theta}^{\mathrm{est}}$

$\boldsymbol{\theta}_0^0 \leftarrow [\boldsymbol{\lambda}^{\mathrm{init}}, \boldsymbol{\mu}^{\mathrm{init}}]$
$\hat{\boldsymbol{g}}_{ww}^{sc,0} \leftarrow \hat{\boldsymbol{g}}_{ww}^{\mathrm{sc,init}}$
for $t = 0, \ldots,$ MaxOuterIter $- 1$ **do**
  for $k = 0, \ldots,$ MaxInnerIter $- 1$ **do**
    $\boldsymbol{\theta}_{k+1}^{t} \leftarrow$ **LBFGS-B**$(\boldsymbol{g}_{ww}, \boldsymbol{\theta}_k^t, \hat{\boldsymbol{g}}_{ww}^{\mathrm{sc,t}})$
  end
  $\boldsymbol{\theta}_0^{t+1} \leftarrow \boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{t}$
  $\hat{\boldsymbol{g}}_{ww}^{\mathrm{sc,t+1}} \leftarrow \mathbf{PS}_{ww}(\boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{t})$
end
$\boldsymbol{\theta}^{\mathrm{est}} \leftarrow \boldsymbol{\theta}_{\mathrm{MaxInnerIter}-1}^{\mathrm{MaxOuterIter}-1}$

where $\nabla_{\boldsymbol{\theta}} \mathcal{L}^{\mathrm{tot}}$ is the gradient of the objective function, $\mathbf{B}_{\boldsymbol{\theta}}$ is an approximation of the inverse Hessian matrix of $\mathcal{L}^{\mathrm{tot}}$ at $\boldsymbol{\theta}$, and α is the step-size found by a line-search.

C. MLAA-S

To fairly compare our approach with a single energy window acquisition, we implemented a simultaneous maximum likelihood reconstruction of activity and attenuation with photopeak scatter re-estimation (MLAA-S). The algorithm relies on the following strategy: 1) simultaneous optimization of both activity and attenuation maps with LBFGS-B (as for MLAA-EB-S); 2) photopeak scatter re-estimation (as for MLAA-EB-S); and 3) single energy window input data. The main difference between MLAA-S and MLAA-EB-S lies in the input data (one versus multiple energy window). This implies that no scatter gradient is computed during MLAA-S iterations, as the only scatter information comes from the photopeak window where the scatter is updated using a one-step-late approach. Pseudo code for MLAA-S is shown in Algorithm 3.

D. MLAA

In this article, a version of MLAA was also used: the framework follows the one of MLAA-S, without the photopeak scatter re-estimation. This method was first proposed in [31].

E. LBFGS-AC

An LBFGS emission reconstruction using the true attenuation map (LBFGS-AC) was also used as further comparison. The algorithm outputs an estimate of the activity image and inputs: 1) ground truth attenuation image $\boldsymbol{\mu}^{\mathrm{true}}$; 2) ground truth photopeak scatter $\boldsymbol{g}_{\mathrm{UU}}^{\mathrm{sc,true}}$; and 3) photopeak window projection data $\boldsymbol{g}_{\mathrm{UU}}$.

F. Implementation

The overall algorithm framework was written in MATLAB (The MathWorks, Natick, MA; version R2018a). The implementation of L-BFGS-B employed in this article is summarized in [32]. $\mathbf{B}_{\boldsymbol{\theta}}$ is constructed with a history length

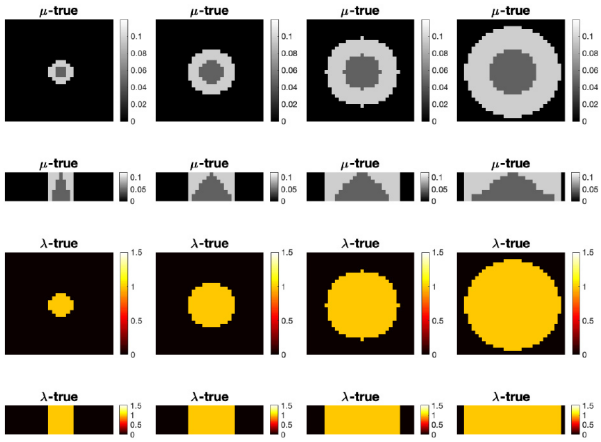

Fig. 1. Cylindrical phantoms of increasing diameters: 8 cm (first column), 16 cm (second column), 24 cm (third column), 32 cm (forth column). First and second rows: attenuation image axial and sagittal view. Third and fourth rows: activity image, axial, and sagittal view. The attenuation is expressed in cm^{-1}, the activity is in arbitrary units.

of 5 (inner) iterations. The objective and gradient functions for the unscattered model were implemented in MATLAB, whilst those related to the scatter model were written in C++ and implemented in open source software for PET and SPECT reconstruction (STIR) [33] (see Appendix C for further details). The simplified wrapper and interface generator (SWIG) [34] was used to call the STIR functions from MATLAB.

G. Stopping Criteria

The photopeak scatter is re-estimated every outer iteration. The reconstruction algorithms rely on three main stopping criteria: the normalized difference between two consecutive image estimates and norm of the projected gradient. Default values for the L-BFGSB implementation were used.

At first iteration, the line search step is initialized by

$$\alpha_0^{\text{init}} = \min\left(\frac{1}{\left\|\nabla \mathcal{L}^{\text{tot}}\left(\boldsymbol{\theta}^{\text{init}}\right)\right\|}, 1\right). \tag{12}$$

A maximum number of inner (MaxInnerIter) and outer iterations (MaxOuterIter) were set (see Section IV-C for details).

IV. Evaluation

The performance of MLAA-EB-S was evaluated with digital phantoms of differing complexity. Simulations were conducted in 3-D.

A. 3-D Phantoms

A first investigation was conducted on a cylindrical phantom with a conical insert (Fig. 1). The conical shape was chosen to simulate the lung. The image size was $30 \times 30 \times 8$ and the voxel dimensions were equal to $1.2 \times 1.2 \times 3.25$ cm^3.

The algorithm was also tested in a more realistic scenario. A 3-D volume from the XCAT torso phantom [35] was generated, cropped to a $60 \times 60 \times 8$ matrix with voxel size of

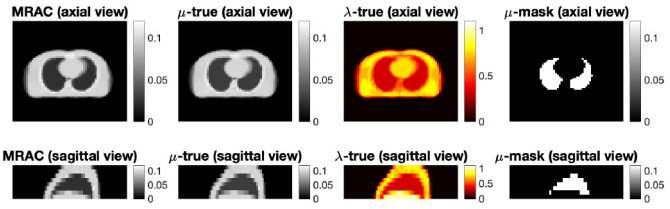

Fig. 2. XCAT phantom. First row: axial view. Second row: sagittal view. From left to right: MR-AC used as initialization $\boldsymbol{\mu}^{\text{init}}$, true attenuation $\boldsymbol{\mu}^{\text{true}}$, true activity $\boldsymbol{\lambda}^{\text{true}}$, lung mask $\boldsymbol{\mu}^{\text{mask}}$. The attenuation is expressed in cm^{-1}, the activity in arbitrary units.

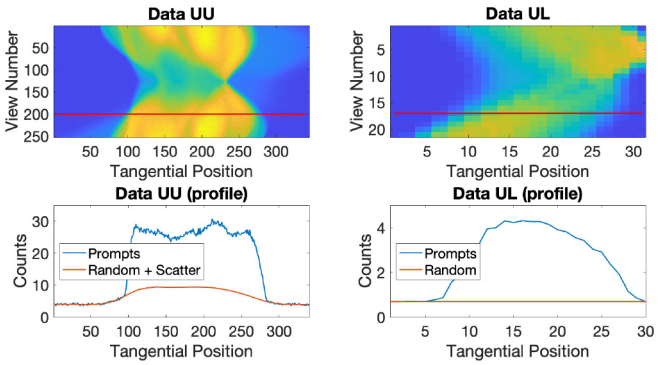

Fig. 3. XCAT phantom simulated data for MLAA-EB-S. UU data (first column) and UL data (second column). For display purpose: 2-D sinograms obtained by summing over the rings (first row) and relative profiles (second row).

TABLE I
Energy Window Thresholds [keV]

L	U_{std}	U	W
$350 - 460$	$430 - 610$	$460 - 570$	$350 - 570$

$0.8 \times 0.8 \times 3.25$ cm^3. Axial and sagittal views of the phantom are shown in Fig. 2. Please note that both cylindrical and XCAT phantoms have the same length in z-direction, covering the length of the scanner (26 cm) and the activity distribution is expressed in arbitrary units.

B. Projection Data

Unscattered data were simulated by forward projecting the ground truth activity image (taking attenuation into account) into sinograms, using the Siemens mMR geometry and specifications [8], [36]: 252 views and 344 tangential positions. The number of rings was downsampled to 8 to match the image voxel size.

The scatter component was computed in low resolution with 21 views, 31 tangential positions, and 8 rings from the analytical model given in (9). Simulations used in-plane detector pairs only. The energy resolution was set to 16%.

Experiments were conducted with both one and two energy windows. For the single window acquisition, we investigated the case of a standard window, as well as the case of a wide energy window WW, where $\boldsymbol{g}_{\text{WW}} = \boldsymbol{g}_{\text{UU}} + \boldsymbol{g}_{\text{UL}} + \boldsymbol{g}_{\text{LU}} + \boldsymbol{g}_{\text{LL}}$. Energy thresholds are shown in Table I. Please note that the energy window U_{std} was introduced to have a fair comparison with the standard energy window used in mMR.

TABLE II
RECONSTRUCTION PARAMETERS

Phantom	MaxOSEMandSSSIter	MaxInnerIter	MaxOuterIter
Cylinder	3	40 , 100	30
XCAT	3	100	15

When simulating data from XCAT, a uniform background was added in all the energy windows to simulate "random" coincidences equal to 39% of the total number of counts of the noise-free prompt data. Poisson noise was added to the XCAT simulated data. The total count level was chosen based on a 240-s PET/MR FDG thorax scan acquired in our institution.

C. Reconstruction Parameters

Both activity and attenuation updates within MLAA-EB-S use L-BFGS-B (Section III-B). In the current results, lung segmentation was incorporated in the algorithm by only updating the attenuation values within the inner cylinder/lung mask during iterations. This constraint is not used for the emission update, for which we only assumed the absence of activity outside the phantom (Fig. 2). Reconstruction parameters are shown in Table II. No further regularization was added at this stage, as it was not the object of this article.

D. Initial Conditions

An MR-based attenuation map was generated by decreasing the lung attenuation values by 20% with respect to the ground truth. In order to avoid dependency on initialization for the different algorithms, all the reconstructions were initialized with the same activity estimate, obtained by iterating between OSEM (3 subsets, 70 subiterations) and SSS [19] (see Table II).

E. Analysis

1) Cylindrical Phantom: Initially, we assessed the effects of the size of the phantom and the choice of energy windows with a two-variable analysis. In this exemplar problem, the algorithm only estimates two values: 1) the mean activity and 2) the attenuation in the insert. For each phantom, we computed the log-likelihood functions for each energy window pair and the sum of them. Furthermore, the relative condition number κ was obtained from the aspect ratio of the ellipse fitted to the contour plot. For this particular study, the energy resolution was set to 1%, so that it was easier to understand the nature of the joint problem under near-ideal conditions.

Then, the performance evaluations of MLAA-EB-S and MLAA-S were assessed for different iteration schemes (Table II). Analyses were conducted in terms of mean percentage error (MPE) in the lung of the estimated images over iterations.

2) XCAT Torso Volumes: For the XCAT reconstruction, 100 noise realizations were used to compute the (voxel-wise) MPE image in both $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ with respect to the ground truth images. The variance and covariance images, denoted VAR($\boldsymbol{\lambda}$), VAR($\boldsymbol{\mu}$), and COV($\boldsymbol{\lambda}, \boldsymbol{\mu}$) were also obtained. A numerical ROI analysis was also computed on $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\mu}}$ (mean estimate over all the noise realizations). The mean bias (MB) was

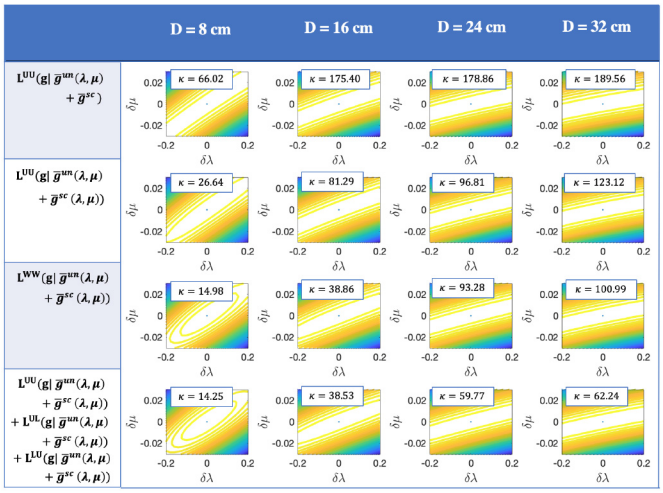

Fig. 4. Log-likelihood plots for cylindrical phantoms. From left to right: the diameter increases. From top to bottom: the energy window varies. κ indicates the condition number of each contour plot.

calculated as

$$\mathrm{MB}(\bar{\boldsymbol{\lambda}}) = \sum_{n=1}^{N} \bar{\lambda}_n^{\mathrm{est}} - \sum_{n=1}^{N} \bar{\lambda}_n^{\mathrm{true}} \tag{13}$$

with N being the number of voxels in the ROI. A similar definition applies to MB($\bar{\boldsymbol{\mu}}$). Mean variance and covariance were also calculated within the same ROI.

V. RESULTS

A. Cylindrical Phantoms

Exemplar Two-Variable Problem: Fig. 4 shows the log-likelihood contour plots for different energy windows for cylinders of diameters ranging between 8 and 32 cm. The first row shows the objective function contour plots for a single-energy window UU when the scatter component is considered as a known (and correct) background. In the second, third and fourth row instead, the dependency of scatter on the activity and attenuation was taken into account in the forward model. The increased curvature and a lower condition number κ demonstrate that the incorporation of the scatter information improves the conditioning of the problem, with larger benefit for lower energy thresholds (third row) and multiple energy windows (fourth row). With regard to the effect of the size of the phantom, this analysis showed that the contours rotate and elongate until becoming almost parallel lines as the diameter increases, leading to a larger condition numbers.

The change in orientation of the objective function gives an insight on the expected activity and attenuation cross-talk, i.e., in the extreme case where a valley is placed along one of the two axes, higher errors are expected in the image along which the valley lies. Ideally, a prior knowledge of the expected cross-talk would be useful for improving on the reconstruction output; in practice, it is not possible to draw contour plots for high dimensional problems.

1) Reconstruction Results—Noise Free Data: We tested the stability of the solution—under ideal conditions—on the

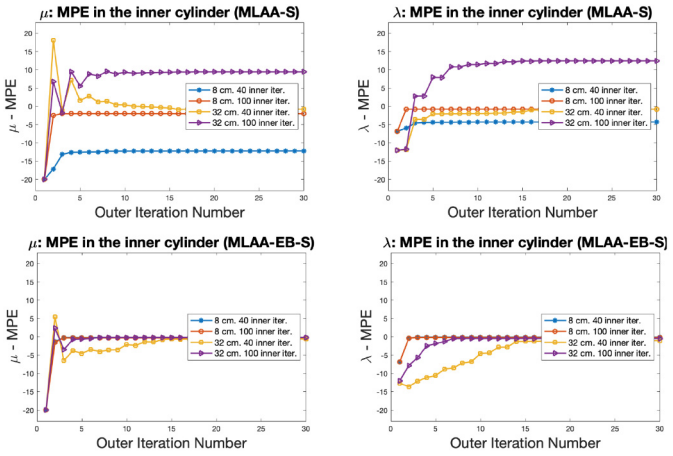

Fig. 5. MPE over outer iterations for the attenuation (left) and activity (right) estimations. First row: MLAA-S. Second row: MLAA-EB-S.

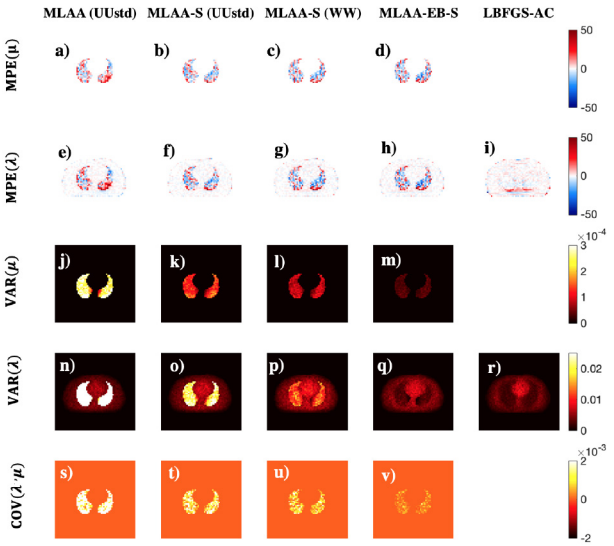

Fig. 6. Error metrics in the XCAT images for different reconstruction algorithms: MLAA (UU_{std}), first column; MLAA-S (UU_{std}), second column; MLAA-S (WW), third column; MLAA-EB-S, fourth column; LBFGS-AC, fifth column. From the top to the bottom: MPE images [%] (a)–(d) and variance (VAR) (j)–(m) for the attenuation from 100 noise realizations; MPE images [%] (e)–(i) and variance (VAR) (n)–(r) for the activity image. Covariance (COV) images (s)–(v).

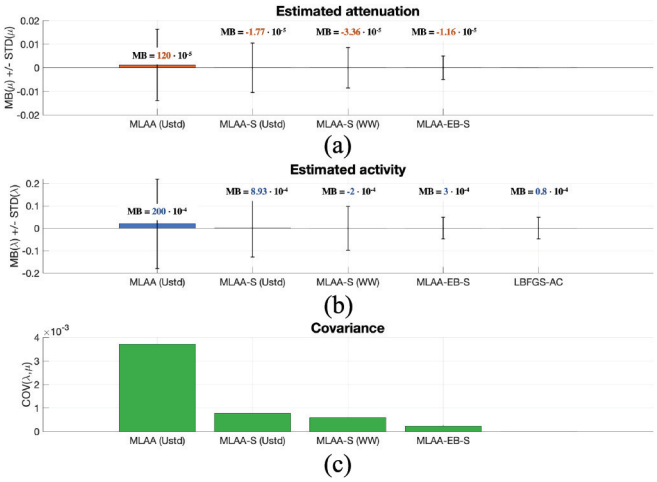

Fig. 7. MB and standard deviation (STD), indicated as error bars, in the estimated attenuation (a) and activity (b); covariance (c). Obtained with: MLAA (UU_{std}), MLAA-S (UU_{std}), MLAA-S (WW), MLAA-EB-S and LBFGS-AC. All mean values are computed over the lung ROI. $\tilde{\boldsymbol{\lambda}}^{\text{true}} = 0.3260$ and $\tilde{\boldsymbol{\mu}}^{\text{true}} = 0.02865$.

cylindrical phantoms in a noise-free scenario, when varying the number of iterations before recomputing the scatter. Fig. 5 shows the MPE in the volume of interest, the inner cylinder (lung), for every outer iteration of both MLAA-S and MLAA-EB-S. We only report curves pertaining to the 8 and 32 cm cylinders, since the intermediate diameters follow a similar trend. According to the two-variable problem analysis conducted in the previous section, reconstructing the larger phantoms is challenging. Nevertheless, MLAA-EB-S manages to find the correct and stable solution for different iteration schemes and for all the phantom sizes. By contrast, MLAA-S was not able to converge to the true solution, with results depending on the exact iteration scheme.

At the last iteration ($N = 30$), MLAA-EB-S achieved a maximum MPE in the attenuation image of 1.259% and 1.44% in the activity image for the 32 cm diameter. For the same phantom, MLAA-S showed a maximum MPE of 9.418% and 12.42% in the estimated attenuation and activity images, respectively (Fig. 5, purple curve).

B. XCAT Reconstruction—Noisy Data

Axial views of the mean error images from all the noise realizations for MLAA (from a standard energy window), MLAA-S (from both standard and wide energy windows), MLAA-EB-S and LBFGS-AC at the last iteration—where convergence is reached—are shown in Fig. 6. The ROI mean values in the lung region for relative bias, variance, and covariance are also reported in Fig. 7.

Results showed that MLAA and LBFGS-AC achieved the worst and best results, respectively, amongst all the four reconstruction methods. MLAA-S outperforms MLAA, thanks to the photopeak scatter re-estimation over iterations, with better results from wider energy window (WW). MLAA-EB-S further improves on MLAA-S in terms of stability of the solution. In particular, MLAA showed a higher bias in both the estimated attenuation [Fig. 6(a)] and activity [Fig. 6(e)] images, whilst both MLAA-EB-S and MLAA-S converged in mean to a similar solution [Fig. 6(b)–(d) and (f)–(h)], all showing a higher noise level in the lung region compared to the one from an LBFGS-AC [Fig. 6(i)]. However, MLAA-EB-S achieved a lower variance with respect to MLAA-S (WW) and MLAA-S (UU_{std}) and MLAA for both the attenuation [Fig. 6(j)–(m)] and activity distributions [Fig. 6(n)–(r)]. Furthermore, MLAA-EB-S was found to have the lowest covariance [Fig. 6(s)–(v)] between the four algorithms, demonstrating that the joint variability of the two unknown images is reduced.

Results from Fig. 7 show that MLAA-EB-S converged to a solution with mean relative bias and standard deviation

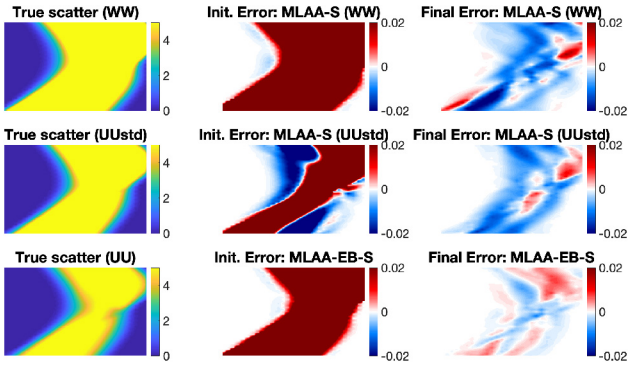

Fig. 8. First column: true scatter. Second column: initial error in the photopeak scatter estimate. Third column: final error in the photopeak scatter estimate. First row: MLAA-S (WW). Second row: MLAA-S (UU_{std}). Third row: MLAA-EB-S.

comparable to the one obtained with an LBFGS-AC reconstruction.

Finally, we compared the error in the photopeak scatter estimate for MLAA-S (WW), MLAA-S (UU_{std}), and MLAA-EB-S (Fig. 8). MLAA-EB-S shows the lowest error in the photopeak scatter estimate.

VI. Discussion

We have proposed a new method for the joint reconstruction of PET activity and attenuation, named MLAA-EB-S. The algorithm takes into account the mutual dependence of scatter, activity, and attenuation. The activity and attenuation distributions are updated simultaneously, whilst the photopeak scatter estimate uses a one-step-late approach.

Section V-A investigated the benefits of incorporating the scatter information into the system model with an exemplar two-variable-problem study where different energy windows and phantom sizes were used, and only two variables were estimated. The results indicate that accounting for the dependency of scatter on both unknown distributions changes the shape of the objective function compared to the "standard" MLAA problem (Fig. 4). Re-estimating the scatter could, therefore, guide an MLAA-like algorithm toward a more stable solution. In the case of one-energy-window acquisitions, larger benefits are present for lower energy thresholds. This is likely due to the larger amount of single scattered counts. This exemplar two-variable problem was also used to investigate the effect of the size of the object. Both the visual representation of the level sets and the condition number κ (Fig. 4) imply that reconstructing big objects is more challenging than small ones. This was also observed in [28] and [37] and further discussed in [29]. The reconstruction algorithms were then tested for the same cylindrical phantoms. MLAA-EB-S, which relies on multiple energy window acquisitions, was found to be stable for different iteration schemes, outperforming the single energy window optimization MLAA-S (see Fig. 5). Overall both MLAA-EB-S and MLAA-S benefit from the re-estimation of the photopeak scatter, as from the first outer iteration the output error was further reduced.

Simulations on 3-D XCAT volumes with different noise realizations showed that both MLAA-EB-S and MLAA-S improved on a standard MLAA reconstruction and converged on average to a similar solution, but appear noisier than the one from LBFGS-AC in the lung region [see Fig. 6(b)–(d) and (f)–(i)]; we believe this is because of additional uncertainty brought by the unknown attenuation values in the lung region. The ROI analysis in Fig. 7 shows that MLAA-EB-S converges in mean to the ground truth solution, with a variance only slightly higher than the one of LBFGS-AC. Reconstructions from MLAA-S exhibited a higher variance and covariance, with increasingly worse results for narrow windows (WW and UU_{std}), confirming the previous observations on the level-set plots (Fig. 4). As the co-variance can be interpreted as a measure of cross-talk, this result illustrates the ill-conditioning of non-TOF MLAA [Fig. 6(s)]. Overall, MLAA-EB-S outperformed MLAA-S in terms of stability of the ML solution.

This article has improved on previous publications in several ways: further understanding of the ill-posed nature of the joint problem, taking into account the finite energy resolution of PET detectors, reconstructing 3-D volumes with data simulated for a clinical PET/MR system in the presence of noise and investigating the cross-talk effects based on noise correlations.

The current study has, however, several limitations. In the current evaluation, the same forward model is used for the simulation of the projection data as for the reconstruction. While this allowed us to assess the accuracy of the reconstruction and the stability of the solution in the absence of model mismatch, the effects of errors in the forward model will need to be investigated.

An additional limitation is the reconstruction of low-resolution objects. Possibly, multiresolution methods [38] could be explored. However, the possibility of recovering high-frequency features will need further investigation. The large voxel size in the reconstruction leads to a relatively low amount of noise in the reconstructed images. The performance of the proposed algorithm at clinical voxel sizes with realistic noise level will need to be evaluated in future work.

An additional limitation is related to the absence of multiple scatter events. Potentially, multiple scatter estimation [39], [40] could be incorporated into our algorithm using the same strategy as for the single scatter estimation in the photopeak window (one-step-late approach). However, accurate estimation of multiple scatter is likely to be more difficult and computationally expensive for the low energy windows. This could impact the performance of the proposed algorithms. The fraction of multiple scatters will be larger in LL in the case of a multiple energy window acquisition. It might, therefore, be beneficial to ignore this data, indicating another potential advantage of MLAA-EB-S compared to using a single WW. However, further studies will be needed to confirm this hypothesis.

Our current model neglects detector scatter. In practice, the probability of scatter within the crystal itself will result in partial energy deposition of photopeak events. Consequently,

these photons can be wrongly assigned to the low energy windows. This effect could be taken into account via the modification of the detection efficiency model.

We expect that the accuracy of the detection efficiency model will affect the possibility of applying this approach in the case of realistic data sets. In addition to the necessity of accounting for detector scatter, normalization factors also differ for scattered and unscattered events [41], as scatter events show a heterogeneity of the angles of incidence and points of origination. This will inevitably impact the accuracy of the application of current normalization techniques to the case of lower energy windows. These considerations will need to be addressed to make a practical application of this type of methodology possible.

The activity and attenuation images used in this article were in similar intensity scales. The absolute scale of the images to be reconstructed influences both the line-search and the convergence rate. Referring to (12), the initial step size can be crucial and strongly dependent on the image intensity scale [31], [42], leading to suboptimal step-size at the first iteration for certain problems. Furthermore, different activity and attenuation gradient intensity scales would lead to slow convergence. Since the scale of the activity can vary with applications, a preconditioner could be introduced [42] into the reconstruction to address the aforementioned challenges.

The proposed methodology was tested on thorax acquisitions given the fact that, among the different tissue classes defined in standard MRAC methods, the lungs have the largest interpatient attenuation values variability [4]. However, assuming the knowledge of attenuation values outside the lung region represents a further restriction of our method. Nevertheless, this assumption is fairly common in PET/MR studies [1], [4].

Potentially, MLAA-EB-S could find additional application for metal hip implants, dental implants or cardiac pacemakers. Brain applications could also benefit from the proposed methodology. However, further investigation would be required to confirm this hypothesis.

MLAA-type algorithms are known to be limited by the possibility of estimating attenuation values in LORs within the support of the activity distribution [4]. However, by using additional information from scatter events it is possible that both MLAA-S and MLAA-EB-S could recover more of the attenuation image, as suggested in [25]. Out-of-FOV scattering could be accounted for by our method as the only constraint on the scatter locations is given by the boundaries of the attenuation map. The template image given to the algorithm could be extended beyond the FOV to place scatter locations outside the edge of the scanner. Similarly to how the arm regions were recovered in [27], there could be potential for the joint reconstruction case. However, this would require further assessment.

It has been shown that TOF information improves the conditioning of the joint estimation problem, although a remaining limitation is a global scaling factor in the estimated activity distribution [15]. Based on the non-TOF results in this article, it is likely that incorporating the scatter information would solve the scaling issues of TOF-MLAA, but this needs to be confirmed in future studies.

Finally, investigating the benefits of using more than two windows could also represent an interesting area of future research.

VII. Conclusion

PET activity and attenuation reconstruction is challenging due to the ill-posed nature of the inverse problem. The benefits of incorporating low energy window information were investigated. A new reconstruction algorithm, MLAA-EB-S was proposed. The method was tested on digital 3-D cylindrical phantoms and XCAT volumes and compared against the reconstruction from a single energy window (MLAA-S). Both MLAA-EB-S and MLAA-S re-estimate the photopeak scatter during the reconstruction. Quantitative results demonstrate that taking scatter into account reduces cross-talk between the activity and attenuation images, and that by using multiple energy windows, MLAA-EB-S outperforms MLAA-S.

This article provides the first evidence that the incorporation of scatter information is beneficial for joint activity and attenuation reconstruction in 3-D PET, even with finite energy resolution. It, therefore, warrants further investigation.

Appendix A
Single Scatter Forward Model

An extension of the SSS model proposed in [19] to the case of a multiple energy window acquisition is proposed in this article. Let $\mathbf{S}_{wy}(\boldsymbol{\theta}) \in \mathbb{R}_{+}^{n'_D \times n_y}$ be the forward operator computing the expected scatter at each energy window pair (w, y) for every pair of detectors $(i, j) \in \{1, \ldots, n'_D\}$

$$\left[\mathbf{S}_{wy}(\boldsymbol{\theta})\right]_{i,j} = \sum_{s=1}^{n_S} \bar{g}^{\text{sc}}_{wy,i,j,s}(\boldsymbol{\theta}) \tag{14}$$

where

$$\bar{g}^{\text{sc}}_{wy,i,j,s}(\boldsymbol{\theta}) = \epsilon_w(E)\epsilon_y(511)\mathrm{I}_{i,s,j} + \epsilon_w(511)\epsilon_y(E)\mathrm{I}_{j,s,i} \tag{15}$$

and E indicating the photon energy after (single) Compton scattering (in keV), as a function of the scatter angle φ for a scatter point at voxel s, $\epsilon_w(E)$ indicating the detector efficiency at a given energy window w and photon energy E; it follows a Gaussian distribution and it takes into account the finite energy resolution of PET detectors [39]. Finally, $I_{i,s,j}$ is defined as

$$\mathrm{I}_{i,s,j} = \left(\frac{\sigma_{i,s}\sigma_{j,s}}{R_{i,s}^2 R_{j,s}^2} \frac{d\sigma}{d\Omega}(\varphi)\ \mu_s \cdot \mathbf{K}_{i,s}\ \boldsymbol{\lambda}\ e^{-\mathbf{K}_{i,s}\boldsymbol{\mu}} e^{-\mathbf{K}_{j,s}\boldsymbol{\mu}_E} \right) \tag{16}$$

where $(d\sigma/d\Omega)(\varphi)$ is the differential cross section given by the Klein and Nishina [43], $\boldsymbol{\mu}_E$ and $\boldsymbol{\mu}$ indicate, respectively, the attenuation value at a given energy E and at 511 keV, $R_{i,s}$ is the distance between the scatter point s and detector i, $\sigma_{i,s}$ denotes the detector cross section presented to the ray i, s, and $\mathbf{K}_{i,s}$ indicates the line integral operator along the line i, s.

The dependency of the attenuation on the energy was addressed by assuming that the attenuation reconstruction is only dependent on Compton scatter and is, therefore, proportional to the total Compton scatter cross section σ_{tot} at a particular energy E

$$\boldsymbol{\mu}_E = \frac{\sigma_{\text{tot}}(E)}{\sigma_{\text{tot}}(511)}\boldsymbol{\mu} = f^*\boldsymbol{\mu}. \tag{17}$$

APPENDIX B
SSS JACOBIAN—ANALYTICAL DERIVATION

Here, we give an overview of the calculation of the Jacobian of the forward scatter model $\mathbf{J}^{\text{sc}}$ with respect to both the attenuation and activity images, needed to be able to compute the gradient of the log-likelihood $\nabla\mathcal{L}^{\text{tot}}_{\boldsymbol{\theta}}$ (10)

$$\nabla_{\boldsymbol{\theta}}\mathcal{L}^{\text{tot}} = \nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{UU}} + \nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{UL}} + \nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{LU}} \tag{18}$$

with

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{UU}} &= \left(\mathbf{J}^{\text{unsc}}_{\text{UU}}\right)^{\top}\left(\boldsymbol{g}_{\text{UU}} \oslash \bar{\boldsymbol{g}}_{\text{UU}} - \mathbf{1}\right)\\
\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{UL}} &= \left(\mathbf{J}^{\text{sc}}_{\text{UL}}\right)^{\top}\left(\boldsymbol{g}'_{\text{UL}} \oslash (\gamma\mathbf{S}_{\text{UL}}(\boldsymbol{\theta})) - \mathbf{1}\right)\\
\nabla_{\boldsymbol{\theta}}\mathcal{L}_{\text{LU}} &= \left(\mathbf{J}^{\text{sc}}_{\text{LU}}\right)^{\top}\left(\boldsymbol{g}'_{\text{LU}} \oslash (\gamma\mathbf{S}_{\text{LU}}(\boldsymbol{\theta})) - \mathbf{1}\right)
\end{aligned} \tag{19}$$

where $\oslash$ indicates the element-wise division and $\mathbf{1}$ is a vector of ones.

In this article, we use the following definition for the probability matrix (6):

$$\mathbf{A}(\boldsymbol{\mu}) = \mathbf{D}(\exp(-\boldsymbol{L}\boldsymbol{\mu}))\boldsymbol{L} \tag{20}$$

with $\mathbf{D}$ indicating a diagonal matrix and $\boldsymbol{L}$ computing the line integral operation.

For each energy window, $\mathbf{J}^{\text{unsc}} = [\mathbf{J}^{\text{unsc}}_{\lambda}, \mathbf{J}^{\text{unsc}}_{\mu}]$ and

$$\begin{aligned}
\mathbf{J}^{\text{unsc}}_{\lambda} &= \mathbf{A}(\boldsymbol{\mu})\\
\mathbf{J}^{\text{unsc}}_{\mu} &= -\mathbf{D}(\mathbf{A}(\boldsymbol{\mu})\,\boldsymbol{\lambda})\boldsymbol{L}.
\end{aligned} \tag{21}$$

In the following, we replaced the discretised images $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ by two functions $\lambda : \mathbb{R}^3 \to \mathbb{R}$ and $\mu : \mathbb{R}^3 \to \mathbb{R}$ for variational formulation. For simplicity, the same notation for the line integral operator $\mathbf{K}$ is used in the continuous case.

Attenuation

Here, we compute the (variational) Jacobian $\mathbf{J}^{\text{sc}}_{\mu}$ of the forward scatter model $\bar{\boldsymbol{g}}^{\text{sc}}$ with respect to μ (17). With regard to the scatter Jacobian $\mathbf{J}^{\text{sc}} = [\mathbf{J}^{\text{sc}}_{\lambda}, \mathbf{J}^{\text{sc}}_{\mu}]$, a simplified expression can be found in [28] and [29]. We derived it in our own notation, according to our forward model. We report the relevant calculations for a pair of detectors (i, j), an energy window pair (v, w), and a scatter point location r_S, where r indicates the Cartesian coordinates (x, y, z).

To compute the Jacobian, we rely on the definition of variational derivative [44]

$$\int \frac{\delta \bar{g}^{\text{sc}}_{wy,i,j,r_S}}{\delta\mu(r)}\phi(r)dr = \lim_{\epsilon\to 0}\frac{\bar{g}^{\text{sc}}_{wy,i,j,r_S}[\mu + \epsilon\phi] - \bar{g}^{\text{sc}}_{wy,i,j,r_S}[\mu]}{\epsilon} \tag{22}$$

where ϕ is an arbitrary function. For simplicity of notation, we omit the dependency of $\bar{g}^{\text{sc}}_{wy,i,j,r_S}$ on (λ, μ). In addition, we group in a constant $\text{C}_{i,r_S,j}$ all the terms that are independent of μ

$$\text{C}_{i,r_S,j} = \epsilon_i(E)\epsilon_j(511)\frac{\sigma_i\sigma_j}{R_i^2R_j^2}\,\frac{d\sigma}{d\Omega}(\varphi)\;\mathbf{K}_{i,r_S}\lambda. \tag{23}$$

Therefore, (15) becomes

$$\begin{aligned}
\bar{g}^{\text{sc}}_{wy,ij,r_S} = {}& \text{C}_{i,r_S,j}\;\mu(r_S)\;e^{-\mathbf{K}_{i,r_S}\,\mu}\;e^{-f^*\mathbf{K}_{j,r_S}\mu}\\
&+ \text{C}_{j,r_S,i}\;\mu(r_S)\;e^{-f^*\mathbf{K}_{i,r_S}\,\mu}\;e^{-\mathbf{K}_{jr_S}\,\mu}
\end{aligned} \tag{24}$$

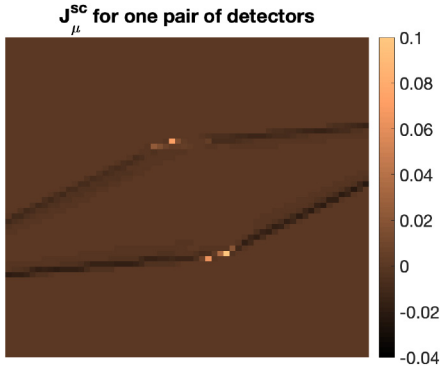

Fig. 9. Illustration of one row of $J^{\text{sc}}_{\boldsymbol{\mu}}$, for one given detector pair within the same ring. UL energy window. Cylindrical phantom ($D = 16$ cm).

where f^* was defined in (17). By taking the limit, the right-hand side of (22)

$$\begin{aligned}
&\tilde{\text{C}}_{i,r_S,j}\left[\phi(r_S) - \mu(r_S)\left(\mathbf{K}_{i,r_S}\,\phi + f^*\mathbf{K}_{j,r_S}\,\phi\right)\right]\\
&\quad + \tilde{\text{C}}_{j,r_S,i}\left[\phi(r_S) - \mu(r_S)\left(f^*\mathbf{K}_{i,r_S}\,\phi + \mathbf{K}_{j,r_S}\,\phi\right)\right]
\end{aligned} \tag{25}$$

where

$$\tilde{\text{C}}_{i,r_S,j} = \text{C}_{i,r_S,j}\left[e^{-\mathbf{K}_{i,r_S}\,\mu}\;e^{-f^*\mathbf{K}_{j,r_S}\,\mu}\right]. \tag{26}$$

To finally obtain $[(\delta\bar{g}^{\text{sc}}_{wy,i,j,r_S})/\delta\mu]$ from (22), an integral-expression for (25) is needed. This can be achieved by considering the integral

$$\phi(r_S) = \int \phi(r)\delta_{r_S}(r)dr \tag{27}$$

where δ_{r_S} is a Dirac function centered in the scatter point r_S and

$$\mathbf{K}_{i,r_S}\phi = \int_{L_{i,r_S}} \phi(r)dr = \int \phi(r)\delta_{L_{i,r_S}}(r)dr \tag{28}$$

where $\delta_{L_{i,r_S}}$ should be interpreted as a measure of integration along the line segment L_{i,r_S} between detector i and the scatter point location r_S.

Consequently, the first line of (25) can be rewritten as

$$\begin{aligned}
\tilde{\text{C}}_{i,r_S,j}\int dr\Big[&\phi(r)\delta_{r_S}(r) - \mu(r_S)\\
&\times\left(\phi(r)\delta_{L_{i,r_S}}(r) + f^*\phi(r)\delta_{L_{i,r_S}}(r)\right)\Big].
\end{aligned}$$

It follows that:

$$\begin{aligned}
\frac{\delta\bar{g}^{\text{sc}}_{wy,i,j,r_S}}{\delta\mu(r)} = {}& \tilde{\text{C}}_{i,r_S,j}\left[\delta_{r_S}(r) - \mu(r_S)\left(\delta_{L_{i,r_S}}(r) + f^*\delta_{L_{j,r_S}}(r)\right)\right]\\
&+ \tilde{\text{C}}_{j,r_S,i}\left[\delta_{r_S}(r) - \mu(r_S)\left(f^*\delta_{L_{i,r_S}}(r) + \delta_{L_{j,r_S}}(r)\right)\right].
\end{aligned} \tag{29}$$

Finally, the Jacobian $\mathbf{J}^{\text{sc}}_{\mu}$ is obtained by summing the contribution of all the scatter points r_S

$$\left[J^{\text{sc}}_{\mu}\right]_{i,j} = \sum_{r_S}\frac{\delta\bar{g}^{\text{sc}}_{wy,i,j,r_S}}{\delta\mu(r)}. \tag{30}$$

Fig. 9(d) provides an illustration for (29). Whilst the non-scatter Jacobian (21) corresponds to the well known set of

lines of response, the scatter Jacobian (30) can be interpreted as the set of "broken-paths" between each detector pair and the scatter point locations allowed for a given energy window.

Activity

The Jacobian $\mathbf{J}_{\lambda}^{\text{sc}}$ with respect to the activity can be obtained using a similar derivation to the one of the attenuation. We group all terms of (15) which are independent of λ

$$\mathbf{F}_{i,r_S,j} = \frac{\sigma_{i,r_S}\sigma_{j,r_S}}{R_{i,r_S}^2 R_{j,r_S}^2}\frac{d\sigma}{d\Omega}(\varphi)\mu(r_S)e^{-\mathbf{K}_{i,r_S}\,\mu}\; e^{-f^*\mathbf{K}_{j,r_S}\,\mu}. \quad (31)$$

This results in

$$\frac{\delta \bar{g}_{wy,i,j,r_S}^{\text{sc}}}{\delta\lambda(r)} = \mathbf{F}_{i,r_S,j}\delta_{L_{i,r_S}}(r) + \mathbf{F}_{j,r_S,i}\delta_{L_{j,r_S}}(r). \quad (32)$$

APPENDIX C
DISCRETISATION AND IMPLEMENTATION IN STIR

In order to be suitable for numerical computing, the variational derivatives shown in Appendix B need to be transformed to (ordinary) derivatives with respect to the discretised images $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$.

Discrete Formulation of SSS and SSS-Jacobian

The STIR SSS implementation approximates the line integrals in the forward model by line ray-tracing between the center of the detector and the scatter point. The ray-sums are obtained by computing the length of intersection $\mathbf{V}_{i,s,n}$ of a line from a scatter point in voxel s to detector i with the voxel n

$$\mathbf{K}_{i,r_s}h \approx \sum_{n=1}^{N} V_{i,s,n}h_n \quad (33)$$

where h_n is the value at voxel n of the discretised image $\boldsymbol{h}$. Please note that $\boldsymbol{h}$ can either be $\boldsymbol{\mu}$ or $\boldsymbol{\lambda}$. The attenuation and the activity image discretisation schemes are described in [45]. The set of scatter points is placed uniformly through the volume determined by the attenuation image. Each scatter point is placed at the center of one voxel of the image. The object's attenuation coefficient at each scatter point is checked and compared to a specified threshold; the scatter point is rejected if its attenuation value falls below the threshold to avoid wasting computational time on points that contribute very little to the total scatter estimate. For this article, the attenuation threshold was set to 0.01 cm^{-1}. The computational burden scales inversely as the product of the three mesh dimensions, whilst the accuracy of the scatter calculation is not highly sensitive to the grid size [19]. A mesh size of approximately 2–3 cm was found to be a good compromise between accuracy and speed of most studies [19]. The sum over all scatter points m gives the total contribution at each detector pair. After discretisation, (29) becomes

$$\frac{\partial \bar{g}_{wy,i,j,s}^{\text{sc}}}{\partial \mu_n} = \tilde{\mathbf{C}}_{i,s,j}\left[\delta_{m,n} - \mu_n\left(\mathbf{V}_{i,s,n} + f^*\,\mathbf{V}_{j,s,n}\right)\right] + \tilde{\mathbf{C}}_{j,s,i}\left[\delta_{m,n} - \mu_n\left(f^*\,\mathbf{V}_{i,s,n} + \mathbf{V}_{j,s,n}\right)\right] \quad (34)$$

where $\delta_{m,n}$ is the Kronecker delta. Similarly, (32) can be discretised as

$$\frac{\partial \bar{g}_{wy,i,j,s}^{\text{sc}}}{\partial \lambda_n} = \mathbf{F}_{i,s,j}\mathbf{V}_{i,s,n} + \mathbf{F}_{j,s,i}\mathbf{V}_{j,s,n}. \quad (35)$$

STIR Implementation Strategy for the SSS-Jacobian

The implementation of the Jacobian matrix into the STIR libraries follows the following strategy: 1) select a scatter point s from the set of sample scatter points and a pair of detectors (i, j); 2) compute the scatter angle φ; 3) compute the corresponding scattered energy $E(\varphi)$ and all the terms related to it (differential cross section, detection probabilities, etc.); and 4) compute the discretised Jacobian.

The Jacobian matrix is recomputed at each iteration. The current implementation uses caching and other techniques in order to significantly reduce the computational burden both in the likelihood and the gradient calculation.

Numerical Validation

The implementation of $\mathbf{J}_{\boldsymbol{\mu}}$ was tested by comparing it against a finite-difference approximation of the gradient, by using a small ϵ perturbation at each voxel n in the $\boldsymbol{\mu}$ image. The accuracy of the implementation was evaluated in terms of absolute error between the gradient and the finite differences, normalized with respect to the maximum absolute value of the gradient. For $\epsilon = 0.0005$, the order of magnitude of the mean and maximum relative error were found to be 10^{-4} and 10^{-3}, respectively.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. Schleyer (Siemens Healthineers), Dr. B. Jakoby (Siemens Healthineers), and R. Twyman (University College London) for helpful discussions.

REFERENCES

[1] Y. Berker, A. Salomon, F. Kiessling, and S. Volkmar, "Lung attenuation coefficient estimation using maximum likelihood reconstruction of attenuation and activity for PET/MR attenuation correction," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf. Rec. (NSS/MIC)*, 2012, pp. 2282–2284.

[2] B. F. Holman, V. Cuplov, B. F. Hutton, A. M. Groves, and K. Thielemans, "The effect of respiratory induced density variations on non-TOF PET quantitation in the lung," *Phys. Med. Biol.*, vol. 61, no. 8, pp. 3148–3163, 2016.

[3] G. Wagenknecht, H.-J. Kaiser, F. M. Mottaghy, and H. Herzog, "MRI for attenuation correction in PET: Methods and challenges," *MAGMA*, vol. 26, no. 1, pp. 99–113, 2013.

[4] A. Mehranian and H. Zaidi, "Emission-based estimation of lung attenuation coefficients for attenuation correction in time-of-flight PET/MR," *Phys. Med. Biol.*, vol. 60, no. 12, pp. 4813–4833, 2015.

[5] N. Burgos *et al.*, "Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2332–2341, Dec. 2014.

[6] F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, "Deep learning MR imaging-based attenuation correction for PET/MR imaging," *Radiology*, vol. 286, no. 2, pp. 676–684, 2018.

[7] L. B. Aasheim, "PET/MR brain imaging: Evaluation of clinical UTE-based attenuation correction," *Nucl. Med. Mol. Imaging*, vol. 42, no. 9, pp. 1439–1446, 2015.

[8] G. Delso *et al.*, "Performance measurements of the Siemens mMR integrated whole-body PET/MR scanner," *J. Nucl. Med.*, vol. 52, no. 12, pp. 1914–1922, 2019.

[9] G. Dournes *et al.*, "Quiet submillimeter MR imaging of the lung is feasible with a PETRA sequence at 1.5 T," *Radiology*, vol. 276, no. 1, pp. 258–265, 2015.

[10] Y. Chen, "Attenuation correction of PET/MR imaging," *Magn. Reson. Imaging Clinics North America*, vol. 25, no. 2, pp. 245–255, 2017.

[11] J. Lillington *et al.*, "PET/MRI attenuation estimation in the lung: A review of past, present, and potential techniques," *Med. Phys.*, vol. 60, no. 1, pp. 1649–1655, 2019.

[12] Y. Berker and Y. Li, "Attenuation correction in emission tomography using the emission data—A review," *Med. Phys.*, vol. 43, no. 2, pp. 807–832, 2016.

[13] A. V. Bronnikov, "Reconstruction of attenuation map using discrete consistency conditions," *Med. Phys.*, vol. 43, no. 2, pp. 807–832, 2016.

[14] J. Nuyts, P. Dupont, S. Stroobants, S. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 393–403, May 1999.

[15] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Phys. Med. Biol.*, vol. 57, no. 4, pp. 885–899, 2012.

[16] R. Boellaard, M. B. Hofman, O. S. Hoekstra, and A. A. Lammertsma, "Accurate PET/MR quantification using time of flight MLAA image reconstruction," *Mol. Imag. Biol.*, vol. 16, no. 4, pp. 469–477, 2014.

[17] J. Nuyts, A. Rezaei, and M. Defrise, "The validation problem of joint emission/transmission reconstruction from TOF-PET projections," *IEEE Trans. Rad. Plasma Med. Sci.*, vol. 2, no. 4, pp. 273–278, Jul. 2018.

[18] A. Rezaei, S. Schramm, S. M. A. Willekens, S. Delso, K. V. Laere, and J. Nuyts, "A quantitative evaluation of joint activity and attenuation reconstruction in TOF-PET/MR brain imaging," *J. Nucl. Med.*, vol. 60, no. 11, 2019.

[19] C. C. Watson, D. Newport, and M. E. Casey, "A single scatter simulation technique for scatter correction in 3D PET," in *Proc. 3-D Image Reconstr. Radiat. Nucl. Med.*, 1996, pp. 256–268.

[20] J. M. Ollinger, "Model-based scatter correction for fully 3D PET," *Phys. Med. Bio.*, vol. 41, no. 1, pp. 153–176, 1996.

[21] A. K. Sulka and U. Kumar, "Positron emission tomography: An overview," *Med. Phys.*, vol. 31, no. 1, pp. 13–21, 2006.

[22] S. C. Cade, S. Arridge, M. J. Evans, and B. F. Hutton, "Use of measured scatter data for the attenuation correction of single photon emission tomography without transmission scanning," *Med. Phys.*, vol. 40, no. 8, 2013, Art. no. 082506.

[23] M. Courdurier, F. Monard, A. Osses, and F. Romero, "Simultaneous source and attenuation reconstruction in SPECT using ballistic and single scattering data," *Inverse Problems*, vol. 31, no. 9, 2015, Art. no. 095002.

[24] A. Bousse, A. Sidlesky, N. Roth, A. Rashidnasab, K. Thielemans, and B. F. Hutton, "Joint activity/attenuation reconstruction in SPECT using photopeak and scatter sinograms," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Room Temp. Semicond. Detector Workshop (NSS/MIC/RTSD)*, 2016.

[25] Y. Berker, F. S. Kiessling, and V. Schulz, "Scattered PET data for attenuation-map reconstruction in PET/MRI," *Med. Phys.*, vol. 41, no. 10, 2014, Art. no. 102502.

[26] Y. Berker, J. S. Karp, and V. Schulz, "Numerical algorithms for scatter-to-attenuation reconstruction in PET: Empirical comparison of convergence, acceleration, and the effect of subsets," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 1, no. 15, pp. 426–434, Sep. 2017.

[27] L. Brusaferri *et al.*, "Potential benefits of incorporating energy information when estimating attenuation from PET data," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, 2017, pp. 1–4.

[28] Y. Berker, J. S. Karp, and V. Schulz, "Joint reconstruction of PET attenuation and activity from scattered and unscattered data," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, 2017.

[29] Y. Berker, V. Schulz, and J. S. Karp, "Algorithms for joint activity—Attenuation estimation from positron emission tomography scatter," *EJNMMI Phys.*, vol. 6, no. 18, p. 18, 2019.

[30] L. Brusaferri *et al.*, "Maximum-likelihood estimation of emission and attenuation images in 3D PET from multiple energy window measurements," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, 2018.

[31] N. Fuin *et al.*, "PET/MRI in the presence of metal implants: Completion of the attenuation map from PET emission data," *J. Nucl. Med.*, vol. 58, no. 5, pp. 840–845, 2017.

[32] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[33] K. Thielemans *et al.*, "STIR: Software for tomographic image reconstruction release 2," *Phys. Med. Biol.*, vol. 57, no. 4, pp. 867–883, 2012.

[34] D. M. Beazley, "SWIG: An easy to use tool for integrating scripting languages with C and C++," in *Proc. USENIX Tcl/TkC Conf.*, vol. 4, 1996, p. 15.

[35] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010.

[36] A. M. Karlberg, O. Sæther, L. Eikenes, and P. E. Goa, "Quantitative comparison of PET performance-Siemens biograph mCT and mMR," *EJNMMI Phys.*, vol. 3, no. 1, p. 5, 2016.

[37] L. Brusaferri *et al.*, "Effects of detected photon energies and phantom size on MLAA from scattered and unscattered data," in *Proc. 8th Conf. PET/MR SPECT/MR*, 2019.

[38] A. Raheja and A. P. Dhawan, "Wavelet based multiresolution expectation maximization image reconstruction algorithm for positron emission tomography," in *Proc. Comput. Med. Imag. Graphics*, 2000, pp. 359–376.

[39] C. Tsoumpas, P. Aguiar, D. Ros, N. Dikaios, and K. Thielemans, "Scatter simulation including double scatter," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf. Rec. (NSS/MIC)*, 2005.

[40] C. C. Watson, J. Hu, and C. Zhou, "Extension of the SSS PET scatter correction algorithm to include double scatter," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf. Rec. (NSS/MIC)*, 2018.

[41] J. M. Ollinger, "Detector efficiency and compton scatter in fully 3D PET," *IEEE Trans. Nucl. Sci.*, vol. 42, no. 4, pp. 1168–1173, Aug. 1995.

[42] Y.-J. Tsai *et al.*, "Fast Quasi-Newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 1000–1010, Apr. 2017.

[43] O. Klein and Y. Nishina, "Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von dirac," *Zeitschrift für Phys.*, vol. 52, nos. 11–12, pp. 853–859, 1929.

[44] M. Giaquinta and S. Hildebrandt, "Calculus of variations 1. The Lagrangian formalism," *Grundlehren der Mathematischen Wissenschaften*, 1996.

[45] C. Tsoumpas, P. Aguiar, K. S. Nikita, D. Ros, and K. Thielemans, "Evaluation of the single scatter simulation algorithm implemented in the STIR library," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf. Rec. (NSS/MIC)*, 2004.