Fast Positron Range Calculation in Heterogeneous Media for 3D PET Reconstruction

László Szirmay-Kalos, Milán Magdics, Balázs Tóth, Balázs Csébfalvi, Tamás Umenhoffer, Judit Lantos, and Gergely Patay

***Abstract*—This paper presents a fast GPU-based solution to compensate positron range effects in heterogeneous media for iterative PET reconstruction. We assume a factorized approach, where projections are decomposed to phases according to the main physical effects. Positron range is the first effect in this chain, which causes a spatially varying blurring according to local material properties. In high-resolution small animal PET systems, the average free path length of positrons may be many times longer than the linear size of voxels. This means that positron range significantly compromises the reconstruction quality if it is not compensated, and also that the material dependent blurring should have a very large support so its voxel space calculation would take prohibitively long. Frequency domain filtering does not have such computational complexity problems, but its direct form is ruled out by the fact that we need a spatially variant filtering in heterogeneous media. To handle heterogeneous media, we execute Fast Fourier Transforms for each material type and for appropriately modulated tracer densities and merge these partial results into a density that describes the composed, heterogeneous medium. Fast Fourier Transform requires the filter kernels on the same resolution as the tracer density is defined, so we also present efficient methods for re-sampling the probability densities of positron range for different resolutions and basis functions. The algorithm is implemented on the GPU, built into the TeraTomo™ system, and requires just a few seconds on high resolution voxel arrays.**

I. Introduction

In Positron Emission Tomography (PET), isotopes decay with positron emission. As a positron travels through the tissue it gives up its kinetic energy by Coulomb interactions with electrons, which results in a chaotic path terminated by positron–electron annihilation. The statistical properties of these paths heavily depend on the initial kinetic energy, i.e. the type of the isotope and on the electron density of the particular tissue, i.e. the type of the material (e.g. bone, flesh, air, etc.). The positron range, i.e. the translation between the isotope decay and the positron annihilation results in positional inaccuracies in tomography reconstruction, which is one of the most important limiting factors of the resolution in small animal PETs [11]. This problem can be attacked by the inclusion of the positron range calculation [4], [3] in ML-EM reconstruction, which consists of the iteration of particle transport calculations, called forward projection, and corrective back projection [11]. The inputs of the reconstruction are the measured values in *Lines of Responses* or *LORs*: $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{\text{LOR}}})$, and the *material map* of the examined object, which is typically obtained by a CT scan.

The output of the reconstruction method, which is simultaneously another input of a single iteration step, is the *tracer density* function $x(\vec{v})$, which is approximated in a *finite function series* form:

$$x(\vec{v}) = \sum_{V=1}^{N_{\text{voxel}}} x_V b_V(\vec{v}), \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\text{voxel}}})$ are the coefficients and $b_V(\vec{v})$ $(V = 1, \ldots, N_{\text{voxel}})$ are pre-defined *basis functions* [1], which are typically defined on a *voxel grid*. The correspondence between the coefficients of the tracer density function (voxel values) and the LOR hits is established by the *system sensitivity* $T(\vec{v} \to L)$ defining the probability that a radioactive decay happened in $\vec{v}$ is detected by LOR L.

Forward projection computes the expectation value of the number of hits in each LOR L:

$$\tilde{y}_L = \int_{\mathcal{V}} x(\vec{v}) T(\vec{v} \to L) \mathrm{d}v = \sum_{V=1}^{N_{\text{voxel}}} \mathbf{A}_{LV} x_V \tag{2}$$

where $\mathcal{V}$ is the domain of the reconstruction, i.e. the field of view of the tomograph, and $\mathbf{A}_{LV}$ is the *system matrix*:

$$\mathbf{A}_{LV} = \int_{\mathcal{V}} b_V(\vec{v}) T(\vec{v} \to L) \mathrm{d}v. \tag{3}$$

Taking into account that the measured hits follow a Poisson distribution, after each forward projection, the ML-EM scheme executes a *back projection* correcting the voxel estimates based on the ratios of measured and computed LOR values [12]:

$$x'_V = x_V \cdot \frac{\sum_L \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L}}{\sum_L \mathbf{A}_{LV}}. \tag{4}$$

The iteration will converge to a fix point where the voxel scaling factors of this formula are equal to one, thus the ML-EM scheme solves the following equation:

$$\sum_L \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L} = \sum_L \mathbf{A}_{LV}.$$

This equation can also written in matrix form

$$\mathbf{A}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}} = \mathbf{A}^T \cdot \mathbf{1}, \tag{5}$$

This work has been supported by the TeraTomo project of the National Office for Research and Technology, TÁMOP-4.2.2.B-10/1-2010-0009, OTKA K-101527, OTKA K-104476, and OTKA K-719922.

L. Szirmay-Kalos, M. Magdics, B. Tóth, and T. Umenhoffer are with Budapest University of Technology and Economics (e-mail: szirmay@iit.bme.hu).

J. Lantos and G. Patay are with Mediso Ltd. (e-mail: Judit.Lantos@mediso.hu).

where ^{T} represents transpose, $\mathbf{y}$ is the vector of measured LOR values, vector division is defined as element-wise division, and $\mathbf{1}$ is a vector containing 1 for every voxel.

System sensitivity can be decomposed according to the main physical effects, like positron range, geometric projection and attenuation, scattering in the tissues, scattering in the detector crystals, and finally the process of detection. In this respect, positron range is modeled by conditional probability density $P(\vec{v}_p \to \vec{v}_a)$ of positron annihilation in $\vec{v}_a$ provided that a positron was born in point $\vec{v}_p$. The annihilation density $x^a(\vec{v}_a)$ is obtained from the tracer density $x(\vec{v}_p)$ applying the blurring caused by the positron range:

$$x^a(\vec{v}_a) = \int_{\mathcal{V}} x(\vec{v}_p) P(\vec{v}_p \to \vec{v}_a) \mathrm{d}v_p.$$

Substituting the finite element approximation of the tracer density, this convolution is expressed by a discrete filtering operation:

$$x^a(\vec{v}_a) = \sum_{V=1}^{N_{\mathrm{voxel}}} x_V \int_{\mathcal{V}} b_V(\vec{v}_p) P(\vec{v}_p \to \vec{v}_a) \mathrm{d}v_p.$$

The finite element coefficient of the annihilation density is computed by multiplying both sides with the *adjoint basis function* $\tilde{b}'_{V'}$, that are orthonormal to the original basis functions, i.e.

$$\int_{\mathcal{V}} b_V \tilde{b}_{V'} \mathrm{d}v = 1 \quad \text{if } V = V' \text{ and zero otherwise.}$$

We use two options, piece-wise constant basis functions when the adjoints are also piece-wise constant basis functions, and tri-linear basis functions when the adjoints are Dirac-delta functions selecting the voxel corners. The result of the scalar product is

$$x^a_{V'} = \int_{\mathcal{V}} \tilde{b}_{V'}(\vec{v}_a) x^a(\vec{v}_a) \mathrm{d}v_a =$$

$$\sum_{V=1}^{N_{\mathrm{voxel}}} x_V \int_{\mathcal{V}} \int_{\mathcal{V}} \tilde{b}_{V'}(\vec{v}_a) b_V(\vec{v}_p) P(\vec{v}_p \to \vec{v}_a) \mathrm{d}v_p \mathrm{d}v_a =$$

$$\sum_{V=1}^{N_{\mathrm{voxel}}} \mathbf{P}_{V',V} x_V,$$

where the discrete filter kernel is

$$\mathbf{P}_{V',V} = \int_{\mathcal{V}} \int_{\mathcal{V}} \tilde{b}_{V'}(\vec{v}_a) b_V(\vec{v}_p) P(\vec{v}_p \to \vec{v}_a) \mathrm{d}v_p \mathrm{d}v_a.$$

If piece-wise constant basis functions are used, then matrix element $\mathbf{P}_{V',V}$ is the probability that a positron is annihilated in voxel V' provided that it was born in voxel V.

System matrix $\mathbf{A}$ is also often factored according to the main physical effects and is approximated by a product of matrices, each representing a physical phenomenon:

$$\mathbf{A} = \mathbf{D} \cdot (\mathbf{G} + \mathbf{S}) \cdot \mathbf{P},$$

where $\mathbf{D}$ is the matrix representing detector effects, $\mathbf{G}$ is the non-square matrix of *geometric projection* and *phantom attenuation*, $\mathbf{S}$ is the non-square matrix of *scattered projection*, and $\mathbf{P}$ is an $N_{\mathrm{voxel}} \times N_{\mathrm{voxel}}$ size matrix of the *positron range*.

Unfortunately, matrix $\mathbf{P}$ is huge and depends both on the radiotracer isotope, e.g. ^{18}F, ^{15}O or ^{82}Rb, and on the tissue type of different voxels. For isotope and material types where the positron range is small, like ^{18}F in bones, the matrix is sparse because the probability that a positron gets far is approximately zero. However, less dense materials like water and air, or isotopes emitting high kinetic energy positrons, like ^{82}Rb, correspond matrices of much fewer zero elements.

II. Methods

In order to simulate positron range, we need effective techniques to implement the filtering operator. We consider two cases, a special one when the material is homogeneous and the filtering becomes spatially invariant, and a general case when the material is inhomogeneous, and the filtering is spatially varying.

A. Homogeneous material

If the tissue is homogeneous, we could exploit the translational symmetry of the positron range, i.e. its probability matrix depends just on the distance of positron generation and annihilation:

$$P(\vec{v}_p \to \vec{v}_a) = P(\vec{v}_a - \vec{v}_p)$$

which makes positron range calculation equivalent to a *convolution*

$$x^a(\vec{v}_a) = \int_{\mathcal{V}} x(\vec{v}_p) P(\vec{v}_a - \vec{v}_p) \mathrm{d}v_p.$$

The computation of the discrete version of the convolution kernel can also benefit from the spatial invariance. Matrix element

$$\mathbf{P}_{V',V} = \int_{\mathcal{V}} \int_{\mathcal{V}} \tilde{b}_{V'}(\vec{v}_a) b_V(\vec{v}_p) P(\vec{v}_a - \vec{v}_p) \mathrm{d}v_p \mathrm{d}v_a = \mathbf{P}_{O,V_r} \tag{6}$$

depends just on the relative location V_r of voxel V' with respect to voxel V, and thus it is enough to compute it for a reference voxel O and all voxels V_r.

A convolution can be evaluated both in spatial domain, i.e. voxel space, and in frequency space having applied Fourier transformation. As the computational complexity of filtering in spatial domain is proportional to the product of the voxel numbers in the positron density volume and the filter kernel, spatial filtering gets prohibitively expensive for large kernels. Note that the linear voxel size of small animal PETs may be about 0.1–0.2 mm, while the FWHM of the positron range effect in water is about 1–4 mm depending on the isotope [2], thus the required size of the 3D filter kernel is greater than 10^3–10^4 voxels. Approximating the filter kernel by a separable approximation like the Gaussian filter can speed up the process, but the Gaussian would be a rather poor approximation of the positron range phenomenon [6]. Fortunately, the convolution can also be evaluated in frequency domain having applied 3D Fast Fourier Transforms $\mathcal{F}$, and

the computational complexity of frequency domain filtering is independent of the kernel size:

$$x^a(\vec{v}) = \mathcal{F}^{-1}\left[\mathcal{F}[x(\vec{v})] \cdot \mathcal{F}[P(\vec{v})]\right].$$

The actual form of kernel $P(\vec{v}_a - \vec{v}_p) = P(\vec{v})$ depends on the material–isotope pair. We calculate $P(\vec{v})$ on high resolution off-line with GATE [5] simulations. The noise of Monte Carlo simulation is filtered out by fitting the simulation data on functions of form

$$P(\vec{v}) = \frac{a\alpha e^{-\alpha|\vec{v}|} + b\beta e^{-\beta|\vec{v}|}}{2\pi|\vec{v}|},$$

which is based on [4], [10], [6] stating that the probability density of positron range projected onto Cartesian axis X can be well approximated by $p_X(X) = ae^{-\alpha X} + be^{-\beta X}$ where parameters a, α, b, β depends on the material–isotope pair. During fitting, we also impose the requirement that p_X is a probability density, thus it integrates to 1.

Matrix elements $\mathbf{P}_{V',V}$ also depend on the discretization, i.e. the basis functions and on the size of voxels. So, having the probability density of the positron range in a continuous analytical form or defined as histograms of measured data, the positron range effect compensation would require the re-sampling of these functions according to the resolution and the basis functions of the voxel grid (equation 6), and then a convolution operation with the currently estimated tracer density. As in pre-clinical research PETs, the voxel size is also a user controllable parameter, the re-sampling is executed on-the-fly applying numerical quadrature when the input data are loaded.

B. Inhomogeneous material

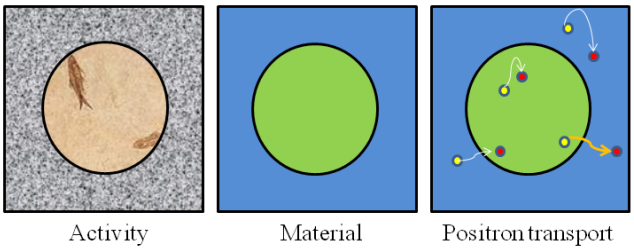

Fig. 1. Positron range is a spatially variant blurring operator. The path of a positron depends on the material where it is born, on the material where it is annihilated, and also on the material that is visited between the generation and the annihilation.

In inhomogeneous objects, blurring kernel $P(\vec{v}_p \to \vec{v}_a)$ also depends on the material (e.g. bone, air, soft tissue) distribution of the measured object, i.e. the material of every voxel (Fig. 1). The precise treatment of this phenomenon would require the consideration of all possible positron paths, which would lead to a high-dimensional integral for every point pair, and would pose prohibitive computational requirements in PET systems. However, assuming that the material is homogeneous, i.e. the blurring kernel is independent of the material type in points $\vec{v}_a$, $\vec{v}_p$ and elsewhere in the object, would be the other extreme approach that would ignore the significantly different probability densities associated with different materials.

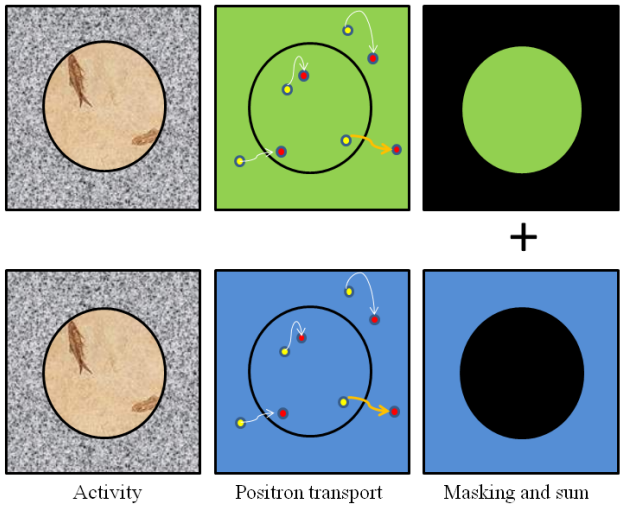

Fig. 2. Intuitive explanation of the simplified method when the material of the positron generation is used. Blurring each voxel with the filter kernel associated with the material in this voxel means the replacement of a single spatial variant filtering by one spatial invariant filtering for each material and a summation.

This paper proposes a practical compromise that is sufficiently accurate for PET reconstruction and can be computed in reasonable time with the support of highly parallel GPU hardware. The basic idea is that instead of considering the material in all points, we take into account the material type only at one end of the positron path. This means that we blur each voxel with the filter kernel associated with the material in this voxel and ignore the fact that there might be a material boundary nearby. This simplification replaces a spatially variant filtering by several spatially invariant convolutions and a summation.

Selecting the material of the positron generation location (Fig. 2) and denoting the index of the material at point $\vec{v}_p$ by $m(\vec{v}_p)$, we obtain:

$$x^a(\vec{v}_a) \approx \sum_m \int_{\mathcal{V}} x(\vec{v}_p)\xi_m(\vec{v}_p)P_m(\vec{v}_a - \vec{v}_p)\mathrm{d}v_p \qquad (7)$$

where $\xi_m(\vec{v}_p)$ is an indicator function that is 1 if there is material of index m in point $\vec{v}_p$ and zero otherwise, and $P_m(\vec{v})$ is the probability density of the positron translation between its generation and annihilation in homogeneous material of index m.

Computing Fourier transform $\mathcal{F}$ for each term and then a single inverse Fourier transform, we get:

$$x^a(\vec{v}) \approx \mathcal{F}^{-1}\left[\sum_m \mathcal{F}[x(\vec{v})\xi_m(\vec{v})] \cdot \mathcal{F}[P_m(\vec{v})]\right].$$

Note that this computation requires the Fourier transforms of the blurring functions computed during pre-processing for each material, the Fourier transformation of the positron density once for each material type (usually two or three), and a single inverse Fourier transformation.

Instead of using the kernel associated with the material of the positron generation location, we can also apply the kernel

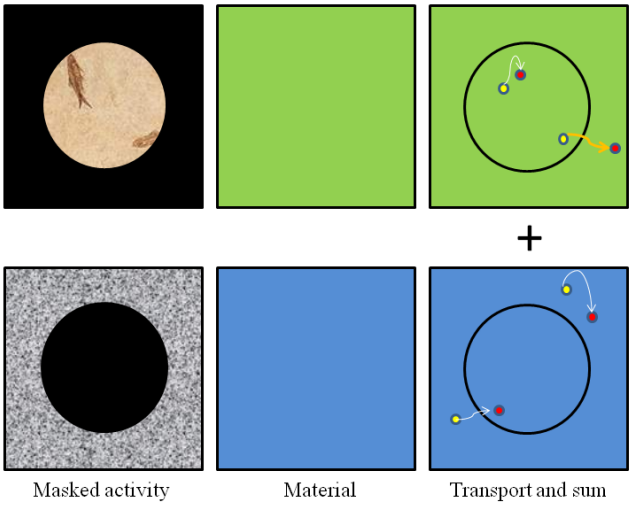

Fig. 3. Intuitive explanation of the simplified method when the material of the positron annihilation is used. Blurring each voxel with the filter kernel associated with the material in this voxel means the execution of a spatial invariant filtering for each material and a summation after masking according to the material.

of the material at the position of the annihilation (Fig. 3), which leads to the following formula:

$$x^a(\vec{v}_a) = \int_{\mathcal{V}} x(\vec{v}_p) P_{m(\vec{v}_p)}(\vec{v}_a - \vec{v}_p) \mathrm{d}v_p = \sum_m \xi_m(\vec{v}_a) \int_{\mathcal{V}} x(\vec{v}_p) P_m(\vec{v}_a - \vec{v}_p) \mathrm{d}v_p.$$

The convolutions in the sum can also be computed via Fourier transformations:

$$x^a(\vec{v}) \approx \sum_m \xi_m(\vec{v}) \mathcal{F}^{-1}\left[\mathcal{F}[x(\vec{v})] \cdot \mathcal{F}[P_m(\vec{v})]\right].$$

This second option is more expensive computationally since here both the number of Fourier transforms and the number of inverse Fourier transforms are equal to the number of materials. The accuracy of the two techniques depends on whether or not the material including most of the radioisotopes occupies a larger part of the object. For typical materials and isotopes, the difference of the reconstructed volumes is negligible.

III. Positron Range in Back Projection

The ML-EM back projector considers the ratios of measured and estimated LOR values and executes the steps of forward projection backwards in reverse order to update the voxel estimates. The positron range operator can be reversed, the only difference is that if we define kernels according to the material at the location of positron generation in the forward projection, then kernels should correspond to the material at the position of annihilation in the back projection, or vice versa.

The geometric back projection is essential in the iterative reconstruction algorithm but positron range and even the detector response may be skipped in this phase, making forward and back projections *unmatched* [13]. The application of unmatched back projectors can reduce the time needed by a single iteration and can improve convergence, but generally makes the system less stable and accurate for low-dose and therefore low-statistics measurements. Ignoring positron range in back projection, however, does not reduce accuracy in the limiting case, i.e. this modification does not distort the fixed point of the iteration. To show why this is true, let us consider a matched EM iteration using system matrix $\mathbf{A}$ both in the forward and back projections, and an unmatched one where the system matrix is modified to $\mathbf{B}$ in back projection, and suppose that the missing factor is the voxel space blurring matrix $\mathbf{P}$ that is responsible for positron range, i.e. $\mathbf{A} = \mathbf{B} \cdot \mathbf{P}$. The matched iteration will converge to voxel values $\mathbf{x}^m$ that satisfy

$$\mathbf{A}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}^m} = \mathbf{A}^T \cdot \mathbf{1}. \tag{8}$$

Suppose that this is a well defined reconstruction problem, i.e. it has a unique solution, which requires $\mathbf{A}^T \cdot \mathbf{A}$ to be non-singular, which means that neither $\mathbf{P}$ can be singular.

The limiting case of the unmatched projection satisfies

$$\mathbf{B}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}^u} = \mathbf{B}^T \cdot \mathbf{1}.$$

As multiplying both sides by $\mathbf{P}^T$, this equation becomes similar to equation 8 and since $\mathbf{P}$ is an invertible square matrix, the two equations are equivalent, thus their solutions are equal: $\mathbf{x}^m = \mathbf{x}^u$.

IV. Results

The proposed methods are implemented in CUDA [9], integrated into the TeraTomo™ system [7], and run on NVIDIA GeForce GTX 480 GPUs. 3D Fourier transformations are computed with the NVIDIA cuFFT library. In this system the positron range calculation for 3 materials at 128^3 and 256^3 resolutions take 0.6 sec and 2 sec, respectively. We note that the numerical accuracy of cuFFT is improved if the voxel array is padded to have resolutions that are powers of two. During padding, we use the voxel values at the borders similarly to the "clamp" texturing mode of GPUs.

To demonstrate the potential of the proposed algorithms, we used Mediso's *nanoScan PET/CT* [8], which has 12 detector modules consisting of 81×39 crystals of size $1.12^2 \times 13$ mm^3.

In the first set of experiments, we considered simulation data obtained by GATE. We reconstructed the *ring phantom* of homogeneous activity put into water and bone materials with and without positron range compensation (Fig. 4). Note that the homogeneous activity inside the ring could be well reconstructed for ^{18}F and ^{15}O isotopes. The ring geometry also shows up nicely for ^{82}Rb, but the homogeneous activity is compromised at material boundaries. The reason of this artifact is our approximate model which uses the material at the position of the annihilation and assumes that the positron was also born in the same material.

We also examined the *Micro Derenzo phantom* with rod diameters $1.0, 1.1, \ldots, 1.5$ mm in different segments, which was reconstructed at $174^2 \times 146$ resolution (0.15^3 mm^3 voxels)

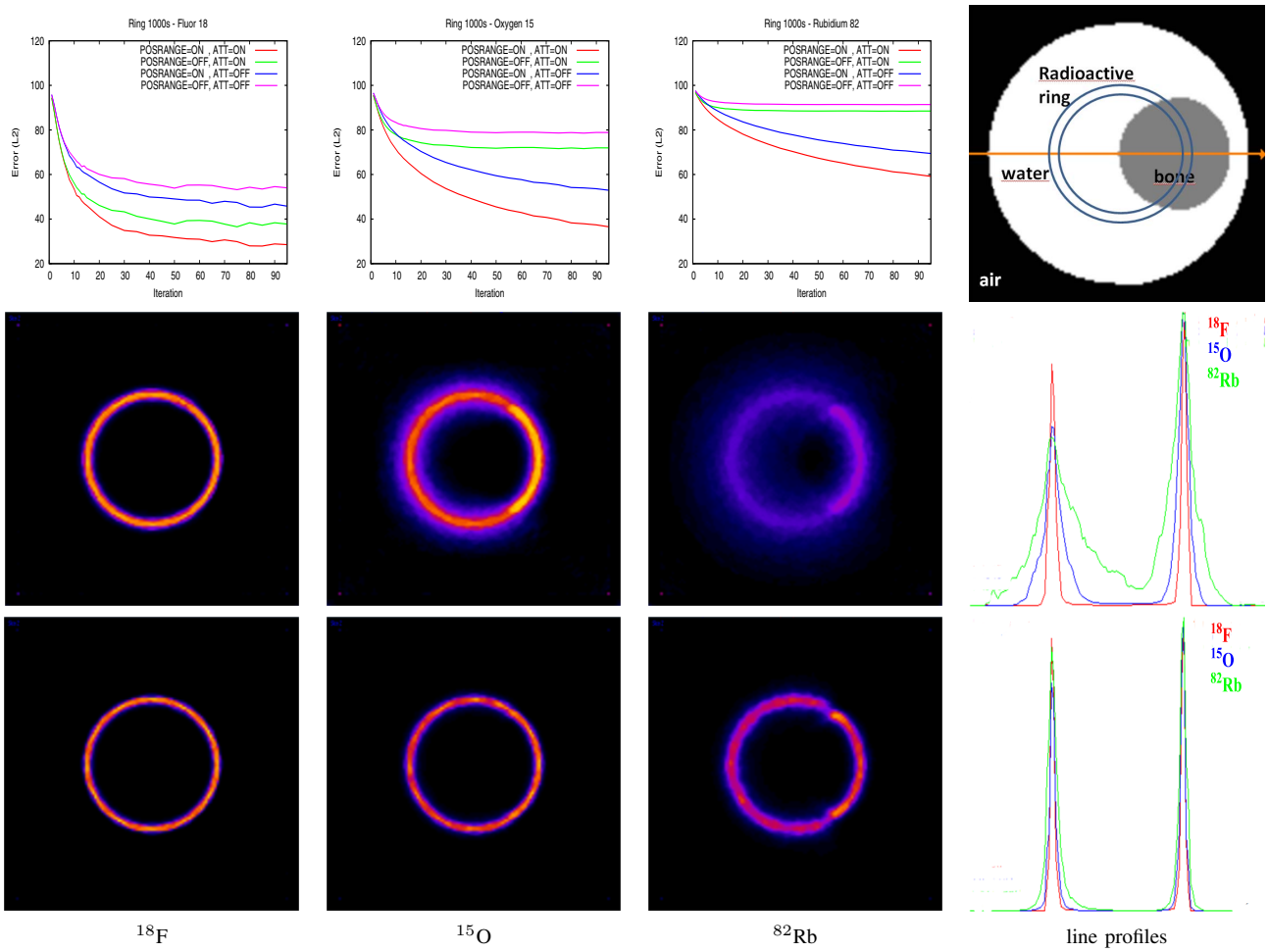

Fig. 4. Reconstruction of the ring. The upper row contains the L2 error curves and the material map. The images of the middle and lower rows show the reconstructions without and with positron range compensation, respectively. Columns correspond to different isotopes.

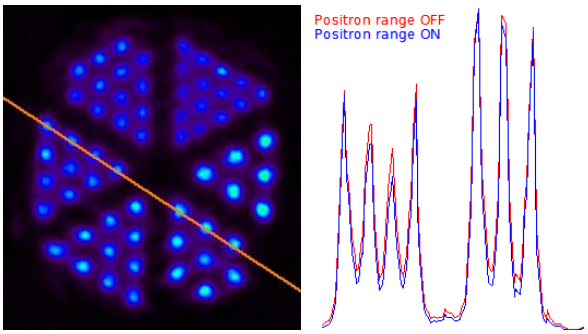

Fig. 5. Micro Derenzo phantom reconstruction and the line profiles showing the difference produced by positron range compensation.

with and without positron range compensation. The transversal slice and the line profiles are shown by Fig. 5.

Fig. 6 shows the reconstruction of a small animal IQ phantom and the line profiles obtained with and without positron range.

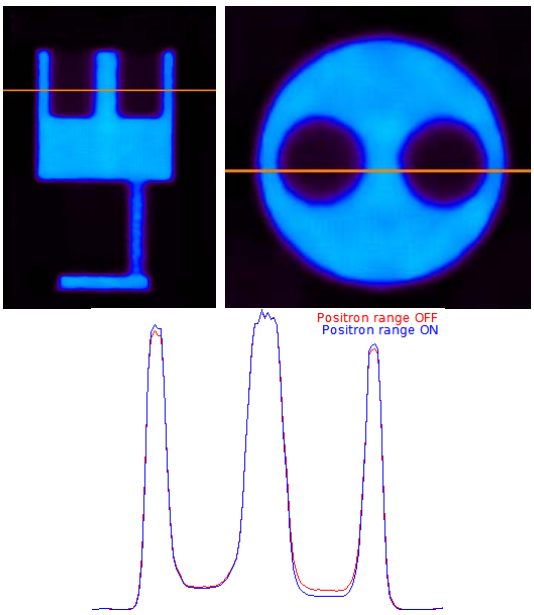

Fig. 6. Small animal IQ phantom reconstruction and line profiles comparing to the reconstruction without positron range compensation.

V. Conclusions

This paper presented an efficient positron range compensation algorithm. The positron range calculation in heterogeneous material is decomposed to a series of positron range calculations in homogeneous materials, once for each material type of the examined object, and a final compositing step. Positron range in homogeneous material, in turn, is evaluated in the frequency domain applying FFTs. The complete reconstruction program is implemented in CUDA and runs on the GPU, providing positron range simulations in 0.6 sec and 2 sec at 128^3 and 256^3 resolutions, respectively. We also investigated the need of positron range calculation in the back projection of the ML-EM scheme, and concluded that it can be safely skipped, making the reconstruction faster.

References

[1] J. Cabello and M. Rafecas. Comparison of basis functions for 3D PET reconstruction using a Monte Carlo system matrix. *Physics in Medicine and Biology*, 57(7):1759–1777, 2012.

[2] J. Cal-González, J.L. Herraiz, S. Espana, M. Desco, J.J. Vaquero, and J.M. Udias. Positron range effects in high resolution 3d pet imaging. In *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 2788–2791, 2009.

[3] J. Cal-González, J.L. Herraiz, S. Espana, E. Vicente, E. Herranz, M. Desco, J.J. Vaquero, and J.M. Udias. Study of CT-based positron range correction in high resolution 3D PET imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648(Supplement 1):S172 – S175, 2011.

[4] S. E. Derenzo. Mathematical removal of positron range blurring in high resolution tomography. *IEEE Transactions on Nuclear Science*, 33:546–549, 1986.

[5] S. Jan and et al. GATE: A simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*, 49(19):4543–4561, 2004. http://www.opengatecollaboration.org.

[6] C. S. Levin and E. J. Hoffmann. Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution. *Physics in Medicine and Biology*, 44:781–799, 1999.

[7] M. Magdics, L. Szirmay-Kalos, B. Tóth, D. Légrády, Á. Cserkaszky, L. Balkay, B. Domonkos, D. Völgyes, G. Patay, P. Major, J. Lantos, and T. Bükki. Performance Evaluation of Scatter Modeling of the GPU-based "Tera-Tomo" 3D PET Reconstruction. In *IEEE Nuclear Science Symposium and Medical Imaging*, pages 4086–4088, 2011.

[8] Mediso. Nanoscan-PET/CT. http://www.mediso.com/products.php ?fid=2,11&pid=86.

[9] NVIDIA. http://developer.nvidia.com/cuda. In *The CUDA Homepage*, 2007.

[10] M. R. Palmer and G. L. Brownell. Annihilation density distribution calculations for medically important positron emitters. *IEEE Transactions on Medical Imaging*, 11:373–378, 1992.

[11] A. J. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clinics*, 2(2):173–190, 2007.

[12] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1:113–122, 1982.

[13] G. Zeng and G. Gullberg. Unmatched projector/backprojector pairs in an iterative reconstruction algorithm. *IEEE Transactions on Medical Imaging*, 19(5):548–555, 2000.