

# Efficient Monte Carlo Methods for Emission Tomography

László Szirmay-Kalos, Balázs Tóth, Gábor Jakab, Péter Gudics, and Milán Magdics

Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, Hungary  
<http://cg.iit.bme.hu>

## Abstract

*Iterative Positron Emission Tomography (PET) reconstruction computes projections between the voxel space and the LOR space, which are mathematically equivalent to the evaluation of multi-dimensional integrals. These integrals are elements of the System Matrix (SM) and can be obtained either by deterministic quadrature or Monte Carlo (MC) methods. Due to the enormous size of the SM, it cannot be stored but integral estimation should be repeated whenever matrix elements are needed. In this paper we show that it is worth using random SM estimates, because this way errors made in projections can compensate each other and do not accumulate to unacceptable values which can happen in case of deterministic approximation.*

## 1. Introduction

In Positron Emission Tomography (PET) we need to find the spatial positron emission density. At a positron–electron annihilation, two oppositely directed photons are generated. Assuming that the electron and the positron are “not moving” before collision, the energy  $E$  of the photons can be obtained from the rest mass  $m_e$  of the colliding particles and the speed of light  $c$ ,  $E = m_e c^2 = 511$  keV. As these photons fly in the medium, they might collide with the electrons of the material. The probability that collision happens in unit distance is the *cross section*  $\sigma$ . During such collision the photon may get scattered, absorbed according to the *photoelectric effect* and new photon pair may be generated, but in our energy range and in human body only incoherent, i.e. Compton scattering is relevant. When scattering happens, there is a unique correspondence between the relative scattered energy and the cosine of the scattering angle, as defined by the *Compton formula*:

$$\varepsilon = \frac{1}{1 + \varepsilon_0(1 - \cos\theta)} \implies \cos\theta = 1 - \frac{1 - \varepsilon}{\varepsilon_0 \varepsilon},$$

where  $\varepsilon = E_1/E_0$  expresses the ratio of the scattered energy  $E_1$  and the incident energy  $E_0$ , and  $\varepsilon_0 = E_0/(m_e c^2)$  is the incident energy relative to the energy of the electron.

The differential of the *scattering cross section*, i.e. the probability density that the photon is scattered from direc-

tion  $\vec{\omega}$  to  $\vec{\omega}'$ , is given by the *Klein-Nishina formula*<sup>12</sup>:

$$\frac{d\sigma}{d\omega} \propto \varepsilon + \varepsilon^3 - \varepsilon^2 \sin^2 \theta$$

where the proportionality ratio includes the classical electron radius and the electron density of the material. Instead of using these physical parameters explicitly, we use the measured cross section of Compton scattering on energy level 511 keV, i.e.  $\varepsilon_0 = 1$  for the representation of the material. From this, the phase function which is supposed to be normalized can be found as:

$$P_{\text{KN}}(\cos\theta) = \frac{\varepsilon + \varepsilon^3 - \varepsilon^2 \sin^2 \theta}{\int_{\Omega} \varepsilon + \varepsilon^3 - \varepsilon^2 \sin^2 \theta d\omega}.$$

The energy dependence of the Compton scattering cross section can be computed from the scaling factor in the Klein-Nishina formula:

$$\sigma(\varepsilon_0) = \sigma(1) \cdot \frac{\int_{\Omega} \varepsilon(\varepsilon_0) + \varepsilon^3(\varepsilon_0) - \varepsilon^2(\varepsilon_0) \sin^2 \theta d\omega}{\int_{\Omega} \varepsilon(1) + \varepsilon^3(1) - \varepsilon^2(1) \sin^2 \theta d\omega}.$$

As photons travel in the considered volume, they may get scattered several times before they leave the volume or are captured by a detector.

A PET/CT collects the numbers  $\mathbf{y} = (y_1, y_2, \dots, y_{N_{\text{LOR}}})$  of simultaneous photon incidents in detector pairs, also called *Lines Of Responses* or *LORs*, and obtains the *material map* of the examined object by a CT scan. The output of the reconstruction method is the *tracer density* function  $x(\vec{v})$ ,

which is approximated in a *finite function series* form:

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

where  $\mathbf{x} = (x_1, x_2, \dots, x_{N_{\text{voxel}}})$  are unknown coefficients and  $b_V(\vec{v})$  ( $V = 1, \dots, N_{\text{voxel}}$ ) are *basis functions*, 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*  $\mathcal{T}(\vec{v} \rightarrow L)$  defining the probability that a radioactive decay happened in  $\vec{v}$  is detected by LOR  $L$ .

The Maximum Likelihood – Expectation Maximization (ML-EM) scheme searches tracer density coefficients  $x_1, \dots, x_{N_{\text{voxel}}}$  that maximize the probability of measurement results  $y_1, \dots, y_{N_{\text{LOR}}}$  by an iterative algorithm<sup>5</sup>, which alternates simulations, called forward projections, and corrective steps based on the computed and measured values.

*Forward projection* computes the expectation value of the number of hits in each LOR  $L$   $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{N_{\text{LOR}}})$ :

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

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 (SM)*:

$$\mathbf{A}_{LV} = \int_{\mathcal{V}} b_V(\vec{v}) \mathcal{T}(\vec{v} \rightarrow L) d\mathbf{v} \quad (3)$$

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:

$$x_V^{(n+1)} = x_V^{(n)} \cdot \frac{\sum_L \mathbf{A}_{LV} \frac{y_L}{\tilde{y}_L^{(n)}}}{\sum_L \mathbf{A}_{LV}}, \quad (4)$$

where

$$\tilde{y}_L^{(n)} = \sum_{V'=1}^{N_{\text{voxel}}} \mathbf{A}_{LV'} x_{V'}^{(n)}$$

is the result of forward projecting the current estimate. With a more compact matrix notation, we can also write

$$\mathbf{x}^{(n+1)} = \langle x_V^{(n)} \rangle \cdot \bar{\mathbf{A}}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}^{(n)}} \quad (5)$$

where  $\langle x_V^{(n)} \rangle$  is an  $N_{\text{voxel}}^2$  element diagonal matrix of current voxel values,

$$\bar{\mathbf{A}}_{LV} = \frac{\mathbf{A}_{LV}}{\sum_{L'} \mathbf{A}_{L'V}}$$

is the *normalized SM*, and vector division is interpreted element-wise manner.

This iteration converges to a fixed point  $\mathbf{x}^*$  where voxel

values are not modified by this formula, i.e. the iteration solves the following equation:

$$\bar{\mathbf{A}}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}^*} = \mathbf{1}.$$

In order to study the convergence properties, let us express the activity estimate in step  $n$  as  $\mathbf{x}^{(n)} = \mathbf{x}^* + \Delta \mathbf{x}^{(n)}$ , i.e. with the difference from the fixed point. Substituting this into the iteration formula and replacing the terms by first order Taylor's approximations, we obtain<sup>4</sup>:

$$\Delta \mathbf{x}^{(n+1)} \approx \left( \mathbf{1} - \langle x_V^* \rangle \cdot \bar{\mathbf{A}}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{A} \right) \cdot \Delta \mathbf{x}^{(n)},$$

where  $\langle \frac{y_L}{\tilde{y}_L^2} \rangle$  is a  $N_{\text{LOR}}^2$  element diagonal matrix of ratios  $\frac{y_L}{\tilde{y}_L}$ . The iteration is convergent if

$$\mathbf{T} = \mathbf{1} - \langle x_V^* \rangle \cdot \bar{\mathbf{A}}^T \cdot \langle \frac{y_L}{\tilde{y}_L^2} \rangle \cdot \mathbf{A}$$

is a *contraction*.

## 2. Error analysis

To compute forward and back projections, we should consider all points where positrons can be generated and all possible particle paths that can lead to an event in LOR  $L$ . A particle path can be described by a sequence of particle–matter interaction points, thus potential contribution  $\mathcal{T}(\vec{v} \rightarrow L)$  of positron emission point  $\vec{v}$  to LOR  $L$  is a high-dimensional integral, so are the expected LOR hits in Eq. 2 and SM elements in Eq. 3.

In tomography the size of the SM is enormous since both  $N_{\text{voxel}}$  and  $N_{\text{LOR}}$  may exceed  $10^8$ , thus matrix elements cannot be pre-computed and stored, but must be re-computed each time when a matrix element is needed. The standard ML-EM reconstruction scheme is based on the assumption that SM elements as well as forward projections  $\tilde{y}_L$  storing the expected number of hits in LORs can be precisely computed. However, this is not the case since re-computation involves numerical quadrature. Deterministic quadrature results in estimations of deterministic error while Monte Carlo (MC) methods result in random values involving random approximation error. To show why MC methods offer a better solution, we first analyze deterministic approximation.

### 2.1. Deterministic approximation

Deterministic approximation makes a similar error in each iteration step, and thus errors may accumulate during the iteration sequence. To formally analyze this issue, let us first consider that SM estimations may be different in forward projection and back projection, and due to the numerical errors both differ from the exact matrix  $\mathbf{A}$ . Let us denote the forward projection SM by  $\mathbf{F} = \mathbf{A} + \Delta \mathbf{F}$  and the normalized back projection SM by  $\bar{\mathbf{B}} = \bar{\mathbf{A}} + \Delta \bar{\mathbf{B}}$ . The ML-EM iteration

scheme using these matrices is

$$\mathbf{x}^{(n+1)} = \langle x_V^{(n)} \rangle \cdot \bar{\mathbf{B}}^T \cdot \frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}^{(n)}}, \quad (6)$$

where

$$\bar{\mathbf{B}}_{LV} = \frac{\mathbf{B}_{LV}}{\sum_{L'} \mathbf{B}_{L'V}}$$

is the normalized back projector matrix.

The question is how the application of approximate matrices modifies the fixed point  $\mathbf{x}^*$  of the iteration scheme. Let us express the activity estimate in step  $n$  as  $\mathbf{x}^{(n)} = \mathbf{x}^* + \Delta\mathbf{x}^{(n)}$ . Substituting this into the iteration formula and replacing the terms by first order Taylor's approximations, we obtain:

$$\Delta\mathbf{x}^{(n+1)} \approx \mathbf{T} \cdot \Delta\mathbf{x}^{(n)} + \langle x_V^* \rangle \cdot (\Delta\mathbf{b} - \Delta\mathbf{f})$$

where

$$\Delta\mathbf{f} = \bar{\mathbf{A}}^T \cdot \left\langle \frac{y_L}{\tilde{y}_L^2} \right\rangle \cdot \Delta\mathbf{F} \cdot \mathbf{x}$$

is the error due to the forward projection estimation, and

$$\Delta\mathbf{b} = \Delta\bar{\mathbf{B}}^T \cdot \frac{\mathbf{y}}{\tilde{y}}$$

is the error due to the back projection estimation.

The limiting value will be different from  $\mathbf{x}^*$  due to the errors of forward and back projections:

$$\Delta\mathbf{x}^{(\infty)} = \left( \bar{\mathbf{A}}^T \cdot \left\langle \frac{y_L}{\tilde{y}_L^2} \right\rangle \cdot \bar{\mathbf{A}} \right)^{-1} \cdot (\Delta\mathbf{b} - \Delta\mathbf{f}). \quad (7)$$

According to this formula, matrix  $\left( \bar{\mathbf{A}}^T \cdot \left\langle \frac{y_L}{\tilde{y}_L^2} \right\rangle \cdot \bar{\mathbf{A}} \right)^{-1}$  expresses *error accumulation*, i.e. how the error made in a single step is scaled up during iteration. If the execution of a forward and then a back projection for a point source, i.e. multiplying with matrix  $\bar{\mathbf{A}}^T \cdot \bar{\mathbf{A}}$  is far from the identity, or many LORs have small or even zero measured value  $y_L$ , then error accumulation can be prohibitively large even if the error of a single step is acceptable.

## 2.2. Random approximation

The error accumulation problem of deterministic approximations can be attacked by applying Monte Carlo quadrature to re-compute projections, because an unbiased Monte Carlo quadrature causes a random error of zero mean, so errors made in different iteration steps can hopefully compensate each other. In this case, projector matrices  $\mathbf{F}^{(n)}$  and  $\bar{\mathbf{B}}^{(n)}$  are realizations of random variables and have a different value in each iteration step. Note that as we have to re-compute the matrix elements anyway, the costs of repeating the previous computation or obtaining a statistically independent new estimation are the same.

If projections are computed with *unbiased estimators*,

then the expectations of the random projection matrices will be equal to the exact ones:

$$E[\mathbf{F}] = \mathbf{A}, \quad E[\bar{\mathbf{B}}] = \bar{\mathbf{A}}.$$

Using random approximation instead of the deterministic approximation of expectation of LOR value  $\tilde{y}_L$ , we obtain a random variable  $\hat{y}_L$  that only approximates the expected value. This random variable depends on the random numbers used to compute the MC estimate, thus it can change in every iteration step<sup>6</sup>. This random and varying error makes the iteration not convergent but the iterated value will fluctuate around the exact solution. To get an accurate reconstruction, the center of the fluctuation should be identical or close to the real solution and its amplitude should be small when actual estimate  $\mathbf{x}$  is close to fixed point  $\mathbf{x}^*$ .

### 2.2.1. Center of the fluctuation

The center of the fluctuation is the real solution if iterating from the fixed point, the expectation of executing a forward and a back projection for the real solution is still the real solution, i.e.

$$E\left[\bar{\mathbf{B}}^T \cdot \frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}}\right] = \bar{\mathbf{A}}^T \cdot \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}}.$$

Unfortunately, this requirement is not met even by unbiased projectors.

If the forward projector and the back projector are statistically independent, then the expectation of their product is the product of their expectations:

$$E\left[\bar{\mathbf{B}}^T \cdot \frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}}\right] = E[\bar{\mathbf{B}}^T] \cdot E\left[\frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}}\right] = \bar{\mathbf{A}}^T \cdot E\left[\frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}}\right]. \quad (8)$$

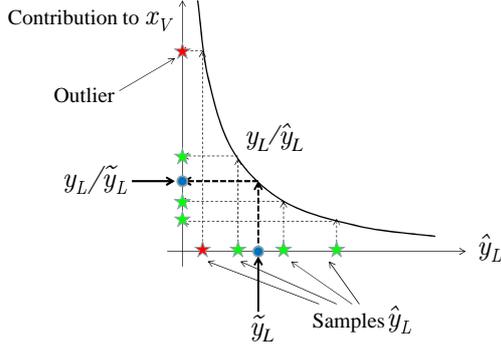
Note that the second factor is generally not equal to  $\frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}}$  since the forward projection result is in the denominator, thus its non-linear, reciprocal function determines the expectation value:

$$E\left[\frac{\mathbf{y}}{\mathbf{F} \cdot \mathbf{x}}\right] \neq \frac{\mathbf{y}}{E[\mathbf{F} \cdot \mathbf{x}]} = \frac{\mathbf{y}}{\mathbf{A} \cdot \mathbf{x}}.$$

To examine this for a single element of the vector, let us consider the expectation of the ratio of measured and computed hits,  $y_L/\hat{y}_L$ . According to the relation of harmonic and arithmetic means, or equivalently to the Jensen's inequality taking into account that  $1/\hat{y}_L$  is a convex function, we obtain:

$$E\left[\frac{y_L}{\hat{y}_L}\right] \geq \frac{y_L}{E[\hat{y}_L]} = \frac{y_L}{\tilde{y}_L}. \quad (9)$$

This inequality states that  $y_L/\tilde{y}_L$  has a random estimator of positive bias<sup>9</sup>. An intuitive graphical interpretation of this result is shown by Fig. 1. Here we assume that the iteration is already close to the fixed point, so different estimates are around the expected detector hit corresponding to the maximum likelihood. Note that the division in the back projection may amplify forward projection error causing large fluctuations, especially when  $\tilde{y}_L$  is close to zero.



**Figure 1:** Expected LOR hit number  $\tilde{y}_L$  is approximated by random samples  $\hat{y}_L$ , which have mean  $\tilde{y}_L$ . These random samples are shown on the horizontal axis. Back projection computes ratio  $y_L/\hat{y}_L$  to obtain voxel updates, which is a non-linear, convex function, resulting in voxel values that may be much higher than the correct value  $y_L/\tilde{y}_L$ . These overshooting samples are responsible for a positive bias and occasionally cause a large random increase in the voxel value.

This bias can be tolerated if the forward projector has low variance, thus the generated values  $\hat{y}_L$  are in a small interval where the application of a non-linear reciprocal function can be well approximated by a linear one. Another approach is the modification of the ML-EM scheme in order to correct the distorted sample distributions to restore unbiasedness even after the application of the reciprocal function<sup>9</sup>.

If the forward projector and back projector are not statistically independent even the factorization of Equ. 8 fails in addition to the problem of non-linear operations. From the point of view of having the random process oscillating around the real solution, we can conclude that the forward projector and the back projector should preferably be independent and the forward projector should have small variance.

### 2.2.2. Amplitude of the fluctuation

The second requirement of accurate reconstruction in addition to the correct center of the fluctuation is that the fluctuation should have small amplitude, i.e. the variance of applying a complete iteration step is small, especially when the process is close to the fixed point.

To get small variance, the following factors need to be taken into account. The forward projector should be of small variance especially where the LOR value is small, because this LOR value will be the denominator in the back projection formula. The slope of the  $1/\tilde{y}_L$  function is  $-1/\tilde{y}_L^2$ , which scales up the variance of the forward projector especially when  $\tilde{y}_L$  gets close to zero.

The variance of the back projector is included in the variance of the result without any amplification. As the back

projector matrix elements are in the numerator and the forward projector matrix elements in the denominator, the variance can also be reduced if they are made correlated. When, due to the random approximation of the forward projector a matrix element is overestimated, and thus the corresponding LOR value in the denominator gets greater than needed, the modified voxel value can be made more accurate by simultaneously increasing the matrix element in the numerator, which represents the back projector. So, from the point of view of the oscillation, it seems advantageous to use the same projector for back projection as used for forward projection of the same iteration step. Establishing such a correlation is easy if the same algorithm is used to compute the forward projection and the back projection, only the seed of the random number generation should be set to the same value before back projection as was set before forward projection of this iteration step.

### 2.2.3. Optimal randomization

Analyzing the mean and the variance of a single ML-EM step involving random projectors, we noted that the accuracy of forward projection is more crucial than that of the back projection, but for the independence or correlation of forward and back projectors, unbiasedness and low variance resulted in different requirements. Unbiasedness requires statistically *independent* forward and back projectors, but low variance due to error compensation needs *correlated* forward and back projectors.

Matrix elements are integrals of Equ. 2 where the integrand is a product of source intensity  $x(\vec{v})$  and scanner sensitivity  $\mathcal{T}(\vec{v} \rightarrow L)$  and integration happens in path space where a “point” corresponds to a path of particles from the emission to the absorption in the detectors. The variance of the MC quadrature depends on the number and distribution of the samples and on the variation of the integrand<sup>10, 7</sup>. There are many possibilities to generate sample paths, which differ in the direction of path building and also in whether roughly the same number of samples is used for each LOR integral, or the sampling process prefers some LORs to others and allocates most of the samples to the preferred ones.

If natural phenomena are directly simulated, then annihilation points are sampled proportionally to their emission density, each sample path is generated with its real probability and can cause a single hit, thus the number of samples in LORs will be proportional to the expected values that are computed. This method is called *voxel driven*<sup>8</sup>. This means that different LORs will be calculated with a similar absolute error. However, when roughly the same number of samples are allocated for each LOR, their error depends just on the variation of their corresponding integral. If the variation is proportional to the integrand, then different LORs are computed with the same relative error. This approach is called *LOR driven*.

Whether it is worth trading more bias for less variance,

i.e. using correlated projectors rather than independent ones, depends on the level of fluctuation. This level can be very high when low contribution LORs are estimated with similar absolute error than high contribution LORs. In the extreme case it can happen that a LOR value is approximated by zero in forward projection, which results in an infinite fluctuation unless the matrix elements corresponding to this LOR are also zero in back projection. Thus, voxel driven methods seem to be better with correlated projectors, but LOR driven approaches prefer lower bias provided by independent projectors.

### 3. Photon tracing

In *Photon Tracing*, first annihilation point  $\vec{v}$  is sampled, then the paths of the two annihilation photons are generated mimicking the free path length and scattering in the real material. If annihilation points are sampled proportional to the activity<sup>3,2</sup>, then we have a voxel driven approach. A voxel driven approach initiates

$$N_V = \frac{x_V N_{PT}}{\sum_{V'=1}^{N_{\text{voxel}}} x_{V'}}$$

number of photons from voxel  $V$  where  $N_{PT}$  is the total number of paths initiated from all photons. When the same

$$N_V = \frac{N_{PT}}{N_{\text{voxel}}}$$

annihilation sample points are allocated to each voxel, then the method is LOR driven.

The paths of the two annihilation photons are obtained with scanner sensitivity  $\mathcal{T}(\vec{v} \rightarrow L)$ . To do this, an initial direction is drawn from uniform distribution. Two photons are started from the annihilation point and their free paths are sampled to find the photon-material interaction points. At scattering, a new direction is generated mimicking the Klein-Nishina formula, and the photon energy is adjusted according to the Compton law. When one of the photons leaves the detector or its energy drops below the discrimination threshold, the photon pair is lost and no LOR is contributed. If photons hit the detector surfaces of LOR  $L$ , the simulation of this path is terminated and the affected SM element  $A_{LV}$  is given a contribution equal to  $1/N_V$ .

Photon tracing, as MC methods in general, results in random SM elements, i.e. projections. The MC simulation is repeated in forward and back projections in each iteration step. The correlation or independence between the forward and back projections can be controlled by whether or not the seed of the pseudo random number generator is reset between these projections. As stated, the accuracy of forward projection is more important, therefore we propose two techniques to increase the accuracy of forward projections without increasing the number of samples, i.e. computation time.

### 4. Statistical filtering

Recall that the classical ML-EM scheme works with two values in a LOR, the measured value  $y_L$  and its mean  $\tilde{y}_L$  computed from the actual activity estimate. The expected value is a scalar determined by the activity, the measured value is a realization of a random variable of Poisson distribution having mean  $\tilde{y}_L$ . Based on the concept of maximum likelihood estimation, the activity estimate is found so that the joint probability of measured values given the expectations obtained from the activity has a maximum.

This classical view should be altered and we should accept the fact that expectation  $\tilde{y}_L$  cannot be accurately computed, we can only get random samples  $\hat{y}_L$  approximating the expectation. When a LOR is processed we have two random samples, measured value  $y_L$  and random estimate  $\hat{y}_L$  of its expectation. Fluctuations of the activity can be suppressed if those estimates  $\hat{y}_L$  that are unacceptable outliers are replaced by some robust estimate.

To detect whether the pair of measured value  $y_L$  and expected value approximation  $\hat{y}_L$  is acceptable or an outlier, we should check whether  $y_L$  could be a reasonable realization of a Poisson distributed random process of mean  $\hat{y}_L$ .

Given an observation  $y_L$ , the confidence interval for mean  $\tilde{y}_L$  with confidence level  $1 - \alpha$  is given by the following fairly sharp approximation<sup>1,11</sup>:

$$F(y_L) \leq \tilde{y}_L \leq F(y_L + 1)$$

where

$$F(y_L) = y_L \left( 1 - \frac{1}{9y_L} - \frac{z_{\alpha/2}}{3\sqrt{y_L}} \right)$$

and  $z_{\alpha/2}$  is the standard normal deviate with upper tail integral  $\alpha/2$ . It means that when our approximation of the expected value satisfies inequality

$$\hat{y}_L < y_L \left( 1 - \frac{1}{9y_L} - \frac{z_{\alpha/2}}{3\sqrt{y_L}} \right),$$

then we cannot be confident in this approximation and therefore we correct it. We set  $z_{\alpha/2} = 2.1$  because it guarantees that  $\hat{y}_L = 0$  is outside the confidence interval if measured value  $y_L \geq 1$ . This corresponds to 98% percent confidence.

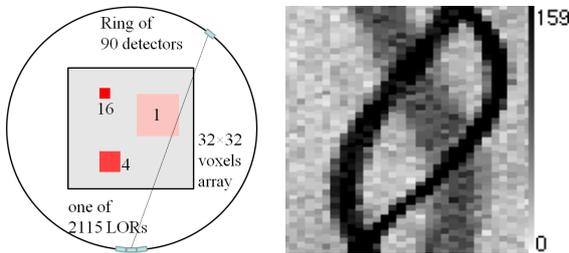
When the current expected value approximation is out of the confidence interval, we should replace it with an acceptable value or completely ignore this LOR in this iteration step. We first try to substitute value  $\hat{y}_L$  by the average of the approximations of this LOR in previous iteration cycles. Note that when the algorithm is close to the converged state, the expected LOR hit does not change too much, so averaging the previous estimates helps decrease the variance of the estimator by trading variance to a small bias. We accept the average when it is in the confidence interval. If even the average is out of the confidence interval, then this LOR is skipped during back projection.

## 5. LOR space blurring

The result of forward projection is the random estimation of LOR hits  $\hat{y}_L$ , which is usually an unbiased estimate having higher variance. As stated, low variance LOR values are essential especially when the expectation is close to zero. Thus, it is worth trading some bias for reduced variance. We can assume that neighboring LORs get similar number of hits, thus variance can be reduced by 4D spatial blurring, which means that each LOR value is replaced by the average of neighboring LOR values. We used  $3 \times 3$  size uniform blurring kernel.

## 6. Results

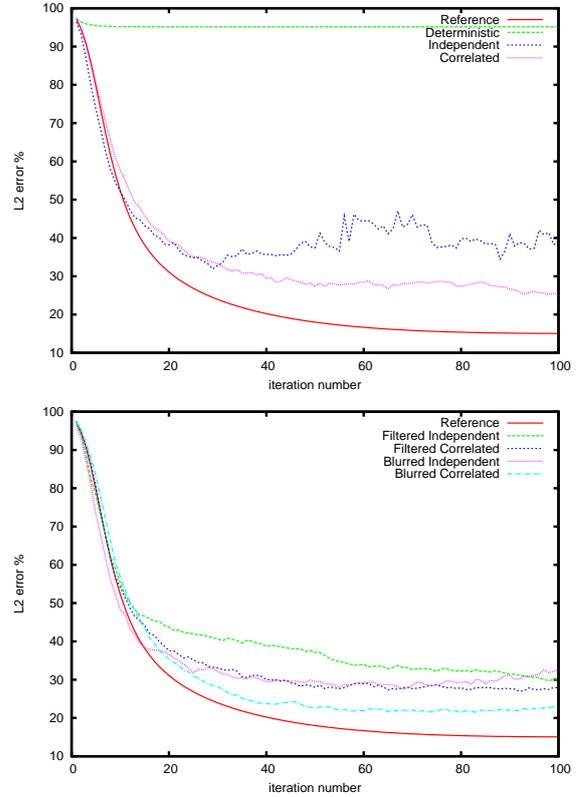
We examine a simple *flat-land* problem describing a 2D PET where  $N_{\text{LOR}} = 2115$  and  $N_{\text{voxel}} = 1024$  (Fig. 2). The reason of the application of the flat-land model is that the 2D planar case makes the SM of reasonable size so the proposed methods can be compared to ground truth solutions when the SM is pre-computed and re-used in iteration steps (recall that this is not possible in fully 3D PET because of the prohibiting size). In order to test scattering in the flat-land model, the Compton law and the Klein-Nishina phase function should be converted to be consistent with the planar case. This means that we use only the scattering angle  $\theta$  and assume that photons remain in the flat-land even after scattering. In 3D, a uniform random rotation around the original direction is also involved, which is now replaced by a random mirroring of 0.5 probability, i.e. by a random decision whether or not  $\theta$  is multiplied by  $-1$ .



**Figure 2:** 2D tomograph model: The detector ring contains 90 detector crystals and each of them is of size 2.2 in voxel units and participates in 47 LORs connecting this crystal to crystals being in the opposite half circle, thus the total number of LORs is  $90 \times 47/2 = 2115$ . The voxel array to be reconstructed is in the middle of the ring and has  $32 \times 32$  resolution, i.e. 1024 voxels. The ground truth voxel array has three hot squares of activity densities 1, 4, and 16 and of sizes  $8^2$ ,  $4^2$ , and  $2^2$ .

The reference activity is a simple function defined by three hot rectangles of Fig. 2. The measured values are obtained by sampling Poisson distributed random variables set-

ting their means to the product of the SM and the reference activity (right of Fig. 2).

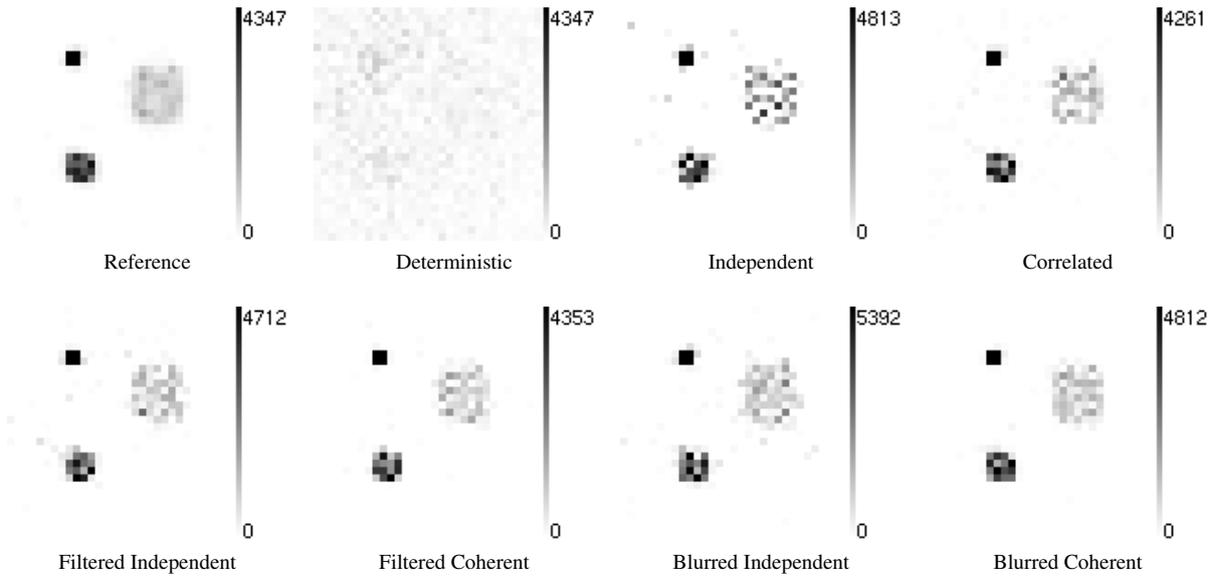


**Figure 3:**  $L_2$  errors of a voxel driven direct MC method. We used  $10^4$  samples in all cases.

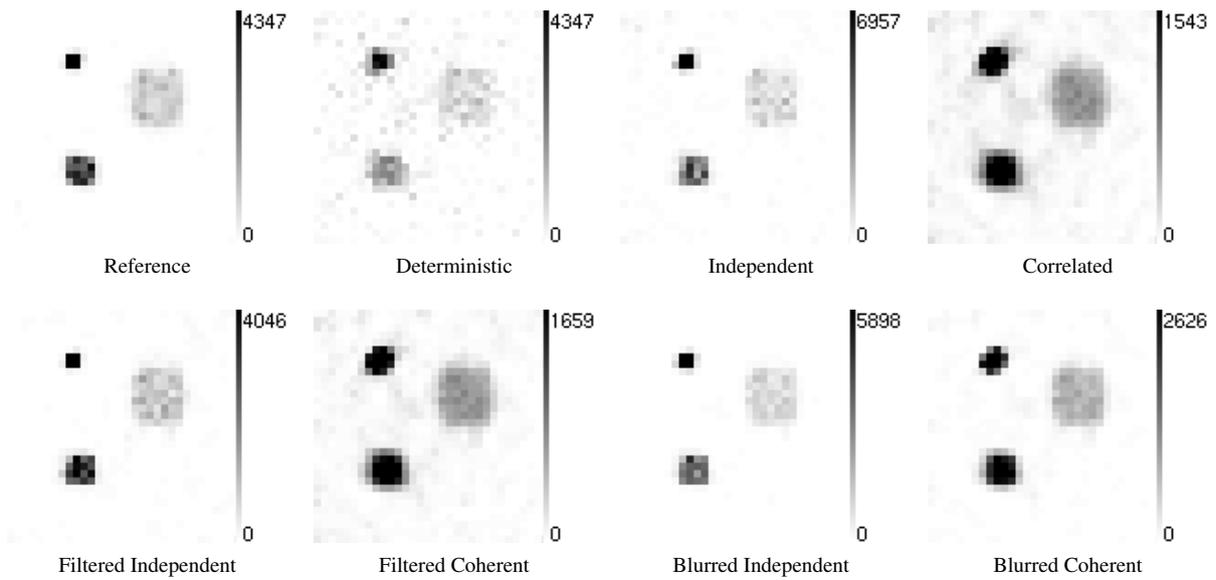
The  $L_2$  error curves are shown by Figs. 3 and 4 for voxel driven and LOR driven methods, respectively, and Figs. 5 and 6 show the reconstructed volumes. As in the case of the given tomograph model where all events are captured and of this non uniform phantom the voxel driven method is more efficient, we used  $10^4$  MC samples in the voxel driven method and  $3 \cdot 10^5$  samples for the LOR driven method. The reference matrix is obtained by  $2 \cdot 10^6$  LOR driven samples. The measurement is simulated with  $5 \cdot 10^4$  samples.

Deterministic iteration, where the same matrix is used in all forward and back projections, has unacceptably poor accuracy if the number of samples is not particularly high. For voxel driven sampling, Correlated projectors provide better solution than Independent projectors. Independent projectors are significantly improved by either statistical filtering or blurring, but only blurring helps Correlated projectors.

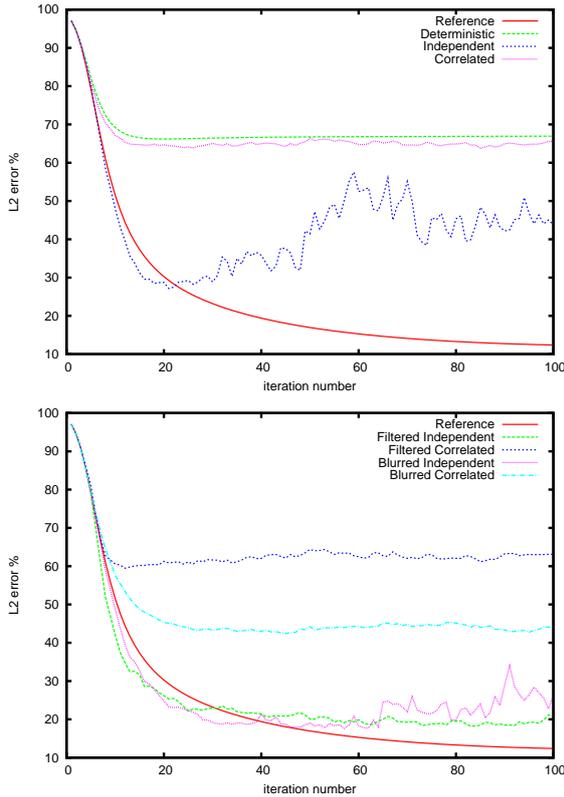
In case of LOR driven sampling, Correlated projectors are almost as bad as Deterministic projectors due to the added bias and blurring helps but filtering does not. Independent



**Figure 5:** Reconstructions of the voxel driven method.



**Figure 6:** Reconstructions of the LOR driven method.



**Figure 4:**  $L_2$  errors of a LOR driven direct MC method. We used  $3 \cdot 10^5$  samples in all cases.

projectors are much better here and their high fluctuation is successfully reduced by both filtering and blurring.

## 7. Conclusions

This paper examined why it is worth using MC estimates to compute forward and back projections in iterative PET reconstruction. We also analyzed the questions whether forward and back projections should be statistically independent or correlated and proposed two techniques to improve the accuracy. These techniques are basically filtering, but statistical filtering operates in the time domain while blurring in the spatial LOR domain.

We concluded that voxel driven methods are worth combining with correlated sampling but LOR driven methods with independent projections. In the future, we examine how the benefits of both approaches can be obtained. Additionally, we also develop more sophisticated LOR blurring methods and instead of applying the same kernel for all contributions, we plan to increase the kernel size depending on the number of scattering events occurred on the photon path.

## Acknowledgement

This work has been supported by OTKA K-104476.

## References

1. N.E. Breslow and N.E. Day. *Statistical Methods in Cancer Research: Volume 2 – The Design and Analysis of Cohort Studies*. International Agency for Research on Cancer. ISBN 978-92-832-0182-3, 1987.
2. B. Csébfalvi. An evaluation of prefiltered reconstruction schemes for volume rendering. *IEEE Trans. on Vis. and Comp. Graph.*, 14(2):289–301, 2008.
3. B. Csébfalvi and L. Szirmay-Kalos. Monte carlo volume rendering. In *Proceedings of the 14th IEEE Visualization Conference (VIS'03)*, pages 449–456, 2003.
4. M. Magdics, L. Szirmay-Kalos, B. Tóth, and A. Penzov. Analysis and control of the accuracy and convergence of the ML-EM iteration. *LECTURE NOTES IN COMPUTER SCIENCE*, 8353:147–154, 2014.
5. L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.
6. L. Szirmay-Kalos. Stochastic iteration for non-diffuse global illumination. *Computer Graphics Forum*, 18(3):233–244, 1999.
7. L. Szirmay-Kalos. *Monte-Carlo Methods in Global Illumination — Photo-realistic Rendering with Randomization*. VDM, Verlag Dr. Müller, Saarbrücken, 2008.
8. L. Szirmay-Kalos, M. Magdics, and B. Tóth. Multiple importance sampling for PET. *IEEE Trans Med Imaging*, 33, 2014.
9. L. Szirmay-Kalos, M. Magdics, B. Tóth, and T. Bükki. Averaging and metropolis iterations for positron emission tomography. *IEEE Trans Med Imaging*, 32(3):589–600, 2013.
10. L. Szirmay-Kalos and L. Szécsi. Deterministic importance sampling with error diffusion. *Computer Graphics Forum*, 28(4):1056–1064, 2009.
11. Wikipedia. [http://en.wikipedia.org/wiki/poisson\\_distribution](http://en.wikipedia.org/wiki/poisson_distribution), 2013.
12. C. N. Yang. The Klein-Nishina formula & quantum electrodynamics. *Lect. Notes Phys.*, 746:393–397, 2008.