Elsevier Editorial System(tm) for Micron
Manuscript Draft

Manuscript Number: JMIC-D-13-00223R1

Title: Autonomous reconstruction and segmentation of tomographic data

Article Type: SI: David C.H. Cockayne SI

Keywords: tomography; tomographic reconstruction; segmentation

Corresponding Author: Dr. Markus Wollgarten,

Corresponding Author's Institution: Helmholtz Zentrum Berlin für Materialien und Energie

First Author: Markus Wollgarten

Order of Authors: Markus Wollgarten; Michael Habeck, Dr.

Abstract: A Bayesian approach to reconstruction and segmentation of tomographic data is outlined and further detailed for the case of absorption tomography. The algorithm allows the quantification of reconstruction errors and segmentation confidence. Calculation results for various experimental settings (number of projections, incident dose, different materials) are shown and discussed.

We are pleased to dedicate this paper to the memory of David J. H. Cockayne.

Highlights

Simultaneous tomographic reconstruction and segmentation algorithm

Fully self-contained

Probabilistic error metrics for both reconstructed densities and segmentation on a voxel basis

Autonomous reconstruction and segmentation of tomographic data

Markus Wollgarten[a,*], Michael Habeck[b]

[a] *Helmholtz Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, D-14109 Berlin, Germany*
[b] *Institute for Mathematical Stochastics, Georg-August-University of Göttingen, Goldschmidtstrasse 7, D-37077 Göttingen,Germany*

Abstract

A Bayesian approach to reconstruction and segmentation of tomographic data is outlined and further detailed for the case of absorption tomography. The algorithm allows the quantification of reconstruction errors and segmentation confidence. Calculation results for various experimental settings (number of projections, incident dose, different materials) are shown and discussed.

We are pleased to dedicate this paper to the memory of David J. H. Cockayne.

Keywords: tomography, tomographic reconstruction, segmentation

1. Introduction

During the last years, three dimensional (3D) investigation of materials has attracted more and more interest. Advances made on the acquisition side, especially in electron tomography of materials [1, 2], are accompanied by new computational methods for reconstruction and segmentation of volumes from a tilt series of projected images [3, 4, 5]. However, although methods become more and more sophisticated, still most of them require substantial user intervention to tune the parameters of the reconstruction procedure and assist algorithms in finding the correct segmentation thresholds. Moreover, for quantification of the final data only estimates of the resolution have been available [6, 7, 8]. Only recently attention has been given to errors and artefacts in (SIRT) reconstructions that can deteriorate quantification [9, 10] and guidelines for the estimation of error bounds in binary tomography have been described [11]. Nevertheless, a (voxelwise) error metric for the reconstructed density as well as a confidence measure for the segmentation is still missing. Therefore future reconstruction methods should aim to minimize user bias and to provide means to assess the confidence of the reconstructed data.

This paper outlines a new statistical approach to tomographic reconstruction and segmentation. Our major focus is to assess the confidence of the reconstruction and segmentation parameters. This requires several new concepts in comparison to standard reconstruction procedures. First we develop a probabilistic model that quantifies the experimental errors and comprises a hierarchical prior probability for the reconstruction parameters. The prior probability differs from the regularizers that are typically used in image reconstruction [12] and introduces segmentation parameters, which assign the voxels to different materials. By means of Bayes's theorem we obtain a *joint* posterior probability for both reconstruction and segmentation parameters. The second innovation is how we explore the posterior distribution. Rather than locating its maximum by means of optimization procedures we *sample* from the posterior using a Monte Carlo algorithm. Our sampling procedure not only draws highly probable reconstruction and segmentation parameters but also explores their uncertainty. Using the posterior samples we can compute the mean reconstructed volume and segmentation as well as assess their reliability using a local entropy-based confidence metric. We develop and demonstrate our algorithm for problems in absorption tomography, but the approach is more generally applicable and not restricted to this particular application.

2. Theory/calculation

We apply the framework of Bayesian probability to absorption tomography. To this end we first develop a probabilistic model for the observed data, which in our context are the measured intensities D_h where index h enumerates all combinations of detector pixel positions and projection angles; the complete data set is $D = \{D_h\}$. The probabilistic model comprises a *likelihood function* that models the experimental noise (section 2.1) and *a prior probability* that implements data-independent properties of a plausible reconstruction (section 2.2). The resulting posterior probability is explored using a Monte Carlo algorithm (section 2.4). The reconstruction and segmentation parameters generated with the Monte Carlo algorithm can be used to assess the confidence of the reconstruction (section 2.5).

*Corresponding author
Email address: wollgarten@helmholtz-berlin.de (Markus Wollgarten)

2.1. Likelihood function for absorption tomography

According to the Beer-Lambert law the measured intensities D are related to the absorption coefficients $\mu = \{\mu_m\}$ through:

$$I_h(\mu) = I_h = I_0\, e^{-\sum_m t_{hm}\mu_m} \tag{1}$$

where I_0 is the dose (given in e.g. electrons per detector pixel) and t_{hm} are the lengths of the beam path by which sample voxel of index m is transversed [13]. The probability of observing intensities D_h is given by a Poisson distribution [12]. Because we assume that the experimental errors are independent, the probability of observing the full set of intensities D given the absorption coefficients is:

$$P(D|\mu) = \prod_h P(D_h|\mu) = \prod_h \frac{I_h(\mu)^{D_h}}{D_h!} e^{-I_h(\mu)}. \tag{2}$$

The Poisson model (2) relates the unknown reconstruction parameters μ to the observed intensities D. General properties of a meaningful reconstruction can be implemented by means of a prior probability on μ.

2.2. Prior probability

For materials science problems, the absorption coefficients μ_m should ideally take distinct values (e.g. for vacuum, metal particles, etc.). We allow them to vary continuously but demand that they should roughly correspond to the material's absorption coefficients c_k where $k = 1, \ldots, K$ enumerates the different materials. The prior distribution of the absorption coefficients, $p(\mu|c_k, \sigma_k)$, can be any uni-modal probability distribution that scatters about a maximum value close to the material's absorption coefficient c_k. Here we assume that the components can be modelled with an exponential distribution in case of vacuum with a mean absorption coefficient c_1 and several Gaussian distributions with mean c_k and standard deviation σ_k $(2 < k < K)$. In case there is evidence that the prior distribution of the absorption coefficient is not captured significantly well with a Gaussian distribution (e.g. because it is skewed), the Gaussian should be replaced by a more appropriate model. In the most general case, the prior probabilities of the different materials are unknown and have to be estimated simultaneously with the absorption coefficients of the individual voxels.

2.2.1. Segmentation

In order to utilize the prior knowledge we would need to know the material contained in every voxel. To express this knowledge formally, we introduce the discrete parameters $z_m \in \{1, \ldots, K\}$ where z_m is a label indicating which material voxel m contains. Given z_m, the *a priori* probability for the absorption coefficient of the m-th voxel follows the prior probability $p(\mu_m|c_{z_m}, \sigma_{z_m})$. The parameters z_m encode a segmentation of the volume into the different materials. Because the correct segmentation is unknown, we have to estimate the segmentation parameters during the reconstruction. That is, we estimate μ and z simultaneously. To do so, we also need to express our prior knowledge about the segmentation. The simplest model assumes that we only know the number of materials K and that every voxel can contain each of the different materials with equal probability $p(z_m = k) = 1/K$. The joint prior probability of the reconstruction and segmentation is then:

$$p(\mu_m, z_m|c, \sigma) = p(\mu_m|c_{z_m}, \sigma_{z_m})/K \tag{3}$$

For this simple prior probability, we can remove the segmentation parameters z_m by summation:

$$\begin{aligned} p(\mu_m|c, \sigma) &= \sum_{z_m=1}^{K} p(\mu_m, \sigma_m|c, \sigma) \\ &= \frac{1}{K}\sum_{k=1}^{K} p(\mu_m|c_k, \sigma_k). \end{aligned} \tag{4}$$

Equation (4) is a *mixture model* with equal probability for the K components corresponding to the materials.

2.2.2. Potts model

The assumption that the segmentation parameters z_m are statistically independent is not realistic because it neglects correlations that result from the fact that the specimen typically consists of spatially extended objects. We capture these correlations with a K-state Potts model [14]:

$$p(z|J) = \frac{1}{Z(J)} \exp\left\{ J \sum_{\langle mm' \rangle} \delta(z_m, z_{m'}) \right\} \tag{5}$$

where $\delta(i, j)$ is the Kronecker delta and $\sum_{\langle mm' \rangle}$ indicates a double sum over all voxels and each voxel's nearest neighbors on a 3D grid; $Z(J)$ is the normalizing partition function. The 2D and 3D Potts models are standard models in statistical physics to study critical behavior. The interaction coupling constant J has to be chosen such that the desired degree of spatial continuity in the segmentation is achieved. For $J = 0$ we recover the simple prior probability neglecting spatial correlations in the segmentation (Eq. 4). Representative segmentations drawn from the

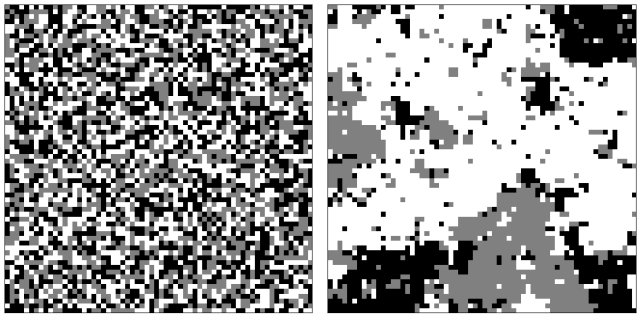


Figure 1: Random configuration generated from the 3-state Potts prior (5) without ($J = 0$; left panel) and with nearest neighbor coupling ($J = 1$; right).

prior distribution (Eq. 5) are shown in Figure 1 with and without nearest neighbor coupling.

In summary, the segmentation parameters are introduced by means of a hierarchical prior probability on the absorption coefficients μ_m. The sample densities are decomposed into classes corresponding to the material composition of the specimen. The unknown class assignments are spatially coupled using the Potts model (5).

2.3. Posterior probability

By using Bayes' theorem we obtain a *joint* posterior probability for the reconstruction and segmentation parameters:

$$p(\mu, z|c, \sigma, J, D) \propto p(D|\mu)\, p(\mu|z, c, \sigma)\, p(z|J). \qquad (6)$$

Therefore, our approach integrates tomographic reconstruction and segmentation into a single probabilistic model. In the following we assume that the parameters characterizing the material absorption coefficients, c_k and σ_k, as well as the coupling constant J of the Potts model are known. Both assumptions could be relaxed but this is beyond the scope of this article [15].

2.4. Posterior sampling algorithm

We aim to not only compute high-quality reconstructions and segmentations but to also assess their confidence. For this purpose we use a Monte Carlo algorithm to draw highly probable reconstruction and segmentation parameters, μ_m and z_m, from their joint posterior distribution (6); we thereby solve the reconstruction and the segmentation problem simultaneously. In contrast to algebraic techniques, our approach does not determine a single optimum or most probable result but generates an *ensemble* of highly probable solutions. The generated specimen densities and voxel labellings allow us to calculate statistical figures of merit for the reconstruction and segmentation.

We use a Gibbs sampler [16] to generate random samples from the full posterior probability distribution (6). Gibbs sampling is an iterative Markov chain Monte Carlo algorithm that cycles over successive parameter updates based on the conditional posterior distributions:

$$\mu^{(l+1)} \sim p(D|\mu)\, p(\mu|z^{(l)}, c, \sigma) \qquad (7)$$

$$z^{(l+1)} \sim p(\mu^{(l+1)}|z, c, \sigma)\, p(z|J) \qquad (8)$$

In step (7), the symbol "$\sim$" indicates that the absorption coefficients are generated from $p(D|\mu)\, p(\mu|z, c, \sigma)$ where z is set to the current segmentation; the next subsection explains this part of the algorithm in more detail. The second step (8) involves sampling from the Potts model under an external field given by $\log p(\mu_m|c_k, \sigma_k)$ for voxel m and label k. We update the segmentation by using Metropolis Monte Carlo (MMC) [17].

2.4.1. Hamiltonian Monte Carlo

In sampling step (7), we use Hamiltonian Monte Carlo (HMC) [18, 19] to update the absorption coefficients μ_m conditioned on the current segmentation $z^{(l)}$. HMC is an efficient method to generate random samples from probability distributions over *continuous* variables. The difference to standard MMC is that additional momentum parameters are introduced for every μ_m. A new set of absorption coefficient μ is proposed by integrating Hamilton's equations of motion using the leapfrog algorithm [19]. During leapfrog integration, the absorption coefficients are guided by the gradient of $-[\log p(D|\mu) + \log p(\mu|z, c, \sigma)]$. Thereby the HMC sampler takes the parameter dependencies and the shape of the conditional posterior distribution into account, which suppresses the random walk behavior from which standard MMC algorithms typically suffer.

2.4.2. Replica-exchange Monte Carlo

The Gibbs algorithm [Eqs. (7) and (8)] explores the joint posterior distribution only locally and is likely to get trapped in a local mode. Already the generation of random labellings z from the Potts model is a challenging task. The Replica-exchange Monte Carlo (RMC) algorithm [20, 21] circumvents this problem by "heating up" the posterior distribution and drawing samples from multiple versions of the system at different temperatures $T = 1/\beta$. If the temperature is chosen high enough, the Gibbs sampler can escape from local modes and samples the heated posterior ergodically. Intermediate distributions bridge between the high-temperature distribution and the true posterior distribution (6). Exchange swaps between neighboring systems allow for the diffusion of states between the high- and low-temperature systems and therefore propagate large rearrangements in the configurations down to the target posterior distribution. RMC is in spirit similar to simulated annealing but has the additional advantage that it samples from the posterior distribution rather than optimizing it. That is, RMC produces μ and z samples whose variability and relative proportions match the posterior distribution.

2.5. Segmentation and error metrics

The posterior samples generated by our algorithm can be used to quantify the quality of the reconstruction in a clear and statistically sound way. First, we can obtain a distribution of absorption coefficient for every voxel and compute its mean absorption coefficient and standard deviation. In addition, the labels z_m provide the segmentation of the continuous absorption coefficient. The posterior probability that the m-th voxel contains material k is the frequency:

$$p_{km} = \frac{1}{L} \sum_{l=1}^{L} \delta(k, z_m^{(l)}) \qquad (9)$$

where l enumerates all L samples generated with our algorithm such that $z_m^{(l)}$ is the class assignment of voxel m in

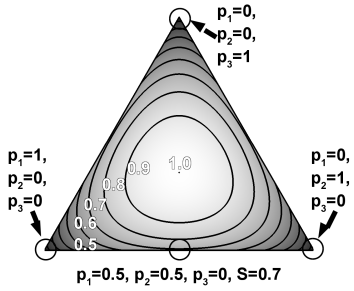


Figure 2: Visualization of the (normalized) segmentation entropy for a three component segmentation.

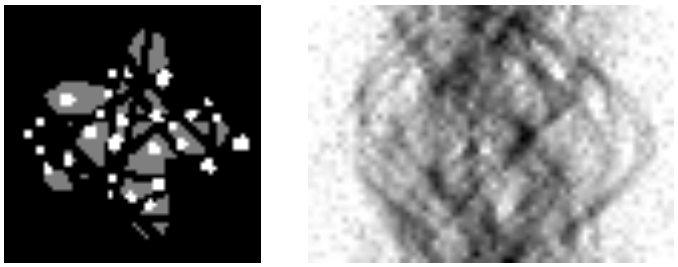


Figure 3: Left: Phantom used for simulation of projections for later reconstruction using the presented algorithm (black: vacuum, grey: carbon black, white: metal). Right: sinogram comprising 45 projections at 4 degrees intervals for the very phantom using a dose of 10^3 counts/pixel.

the l-th sampling iteration. The most probable segmentation is obtained by voxel-wise determination of the most frequent indicator:

$$l_m = \mathrm{argmax}_k \; p_{km}. \tag{10}$$

The confidence in the segmentation can be found by calculating the (normalized) segmentation entropy for each voxel:

$$S_m = -\sum_{k=1}^{K} p_{km} \log p_{km} / \log K. \tag{11}$$

For a three component system this metric is visualized in Fig. 2 from which it is evident that an entropy value of less than 0.7 is required to avoid ambiguity ($p_{1m} = p_{2m} = 0.5$, $p_{3m} = 0$).

3. Results

3.1. Reconstruction of phantom data

To test our algorithm we set up a two-dimensional phantom of 64 by 64 voxels, i. e. a slice of a three-dimensional volume. (Fig. 3). It resembles in its composition and morphology the specimens studied in catalyst research; in addition to vacuum voxels (absorption coefficient ≈ 0.0 voxel^{-1}) there are voxels for carbon black (≈ 0.01 voxel^{-1}) and a metal (≈ 0.03 voxel^{-1}). The absorption coefficients have been chosen in accordance with experimental data and settings obtained from the study of Ru catalysts on carbon black [22]. Projections onto an one-dimensional detector have been calculated for various combinations of projection angles and incident intensities. We refer to the different simulations parameters by <no. of projections>-<angular interval>-<logarithm of dose>, i. e. 180-1-3 refers to 180 projections at 1° intervals and 10^3 counts/pixel.

The absorption coefficients are initialized by an iterative procedure:

$$\mu_m^{(n+1)} = \mu_m^{(n)} \left. \frac{B(\{D_h\})}{B(P(\{\mu_m^{(n)}\}))} \right|_m, \tag{12}$$

where $P(\cdot)$ denotes the forward and $B(\cdot)$ the backward projection function. The initial configuration $\{\mu_m^{(0)}\}$ is filled with random values; iteration (12) is stopped after 500 steps. The mean absorption coefficients of the mixture model c_k are fixed to the values used for the phantom [15]. The standard deviation parameter, σ_k, for the Gaussian mixture model components are determined by subclassing the result of Eq. 12 using the k-means algorithm and calculating the standard deviation of the absorption coefficient of the individual classes.

We use a total of eight replicas to simulate the posterior distribution. In the first four replicas, J is kept at zero (i.e. neighboring voxels are completely uncoupled *a priori*), while β increases linearly from 0.2 to 1.0. In the second half of the chain, J is raised linearly from 0 to 1 (i.e. neighboring voxels are increasingly coupled). After the RMC algorithm has converged, we calculate the mean absorption coefficient and its standard deviation for every class (Figure 7) by averaging over the ensemble of generated segmentation parameters $z_m^{(l)}$ and absorption coefficients $\mu_m^{(l)}$.

3.2. Characteristic results

Figure 4 depicts snapshots of the continuous absorption coefficient for different replicas as found in the course of the calculation. Figures 5 and 6 show all quantities that can be deduced for each voxel from the ensemble of possible solutions: mean absorption coefficient and its standard deviation, segmentation (Eq. 10), segmentation entropy (Eq. 11) and, as the solution is known for the present case, correct segmentation. For each setting, 5000 samples were drawn and statistically analyzed. Besides visual representation of the reconstruction and its error metrics we calculate global quantities that allow us to follow the reconstruction process in a clear manner. Using the segmentation, the slice of continuous absorption coefficients can be subdivided into label classes (three in our application). The class mean and its standard deviation can be determined and plotted against the iteration index (Fig. 7). The development of the segmentation entropy is another informative diagnostic. It is expected that the entropy adopts stable values while accumulation of samples

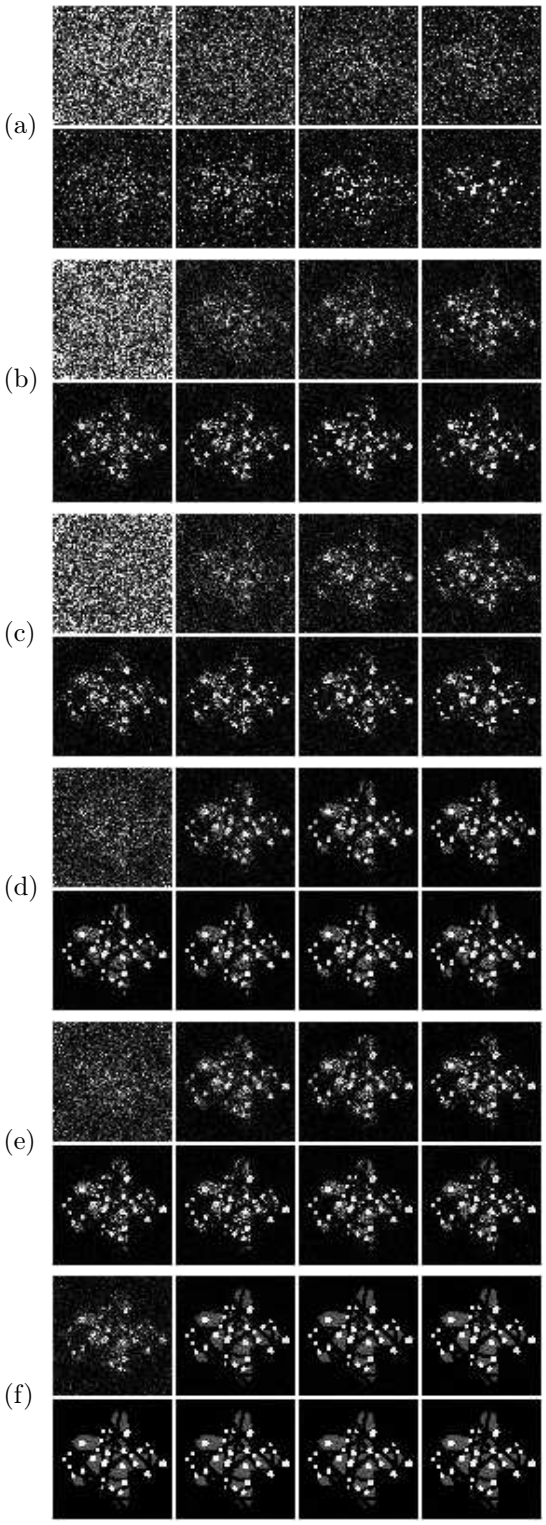


Figure 4: Replica snapshots showing the continuous absorption coefficient (black: 0.0 voxel^{-1}, white: 0.03 voxel^{-1}). Replicas are ordered by β and J left to right and top to bottom. Calculation settings are 20-9-2 (a), 180-1-2 (b), 20-9-3 (c), 180-1-3 (d), 20-9-4 (e) and 180-1-4 (f).

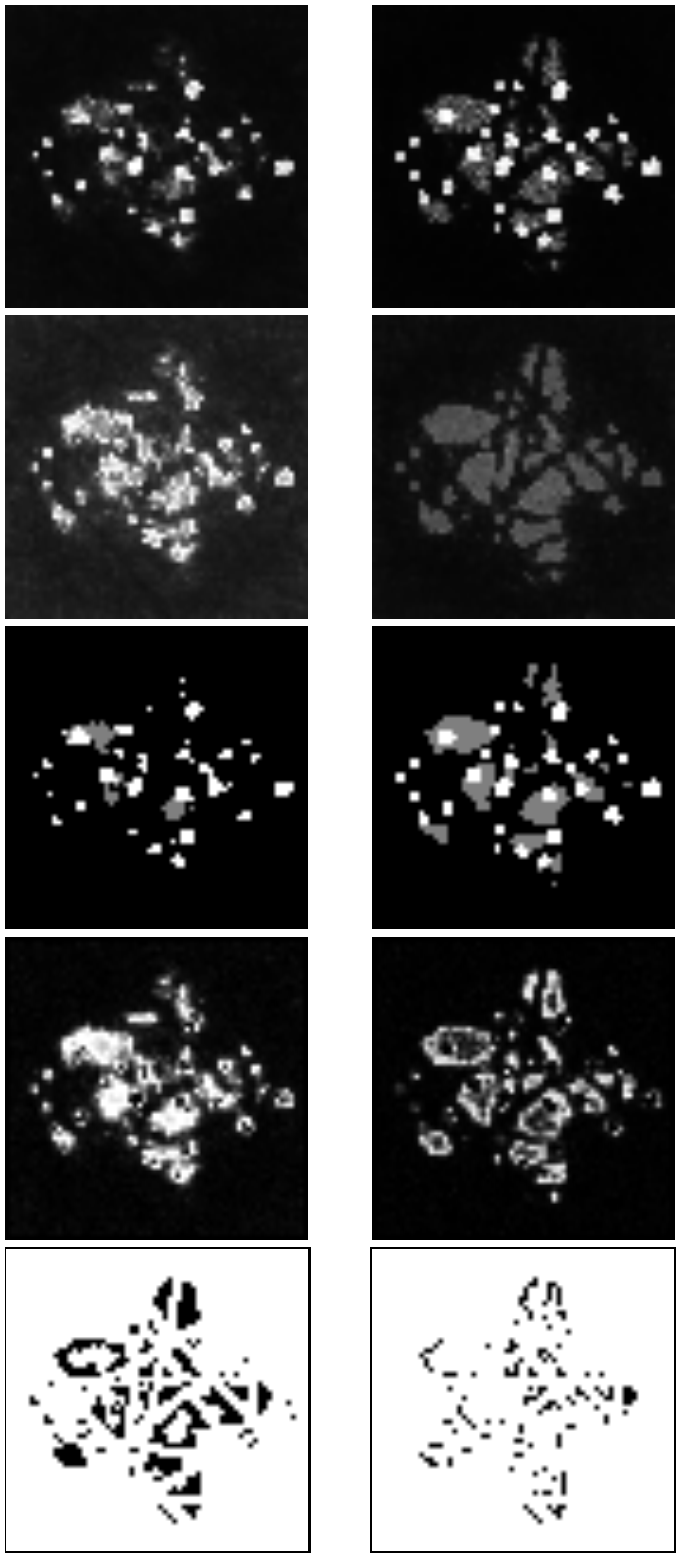


Figure 5: T. t. b.: mean continuous absorption coefficient (range displayed: 0 - 0.03 voxel^{-1}), its standard deviation (0 - 0.01 voxel^{-1}), segmentation, segmentation entropy (0 - 0.7) and correct labeling (white: correct); left column: setting 20-9-3, right column: setting 180-1-3.

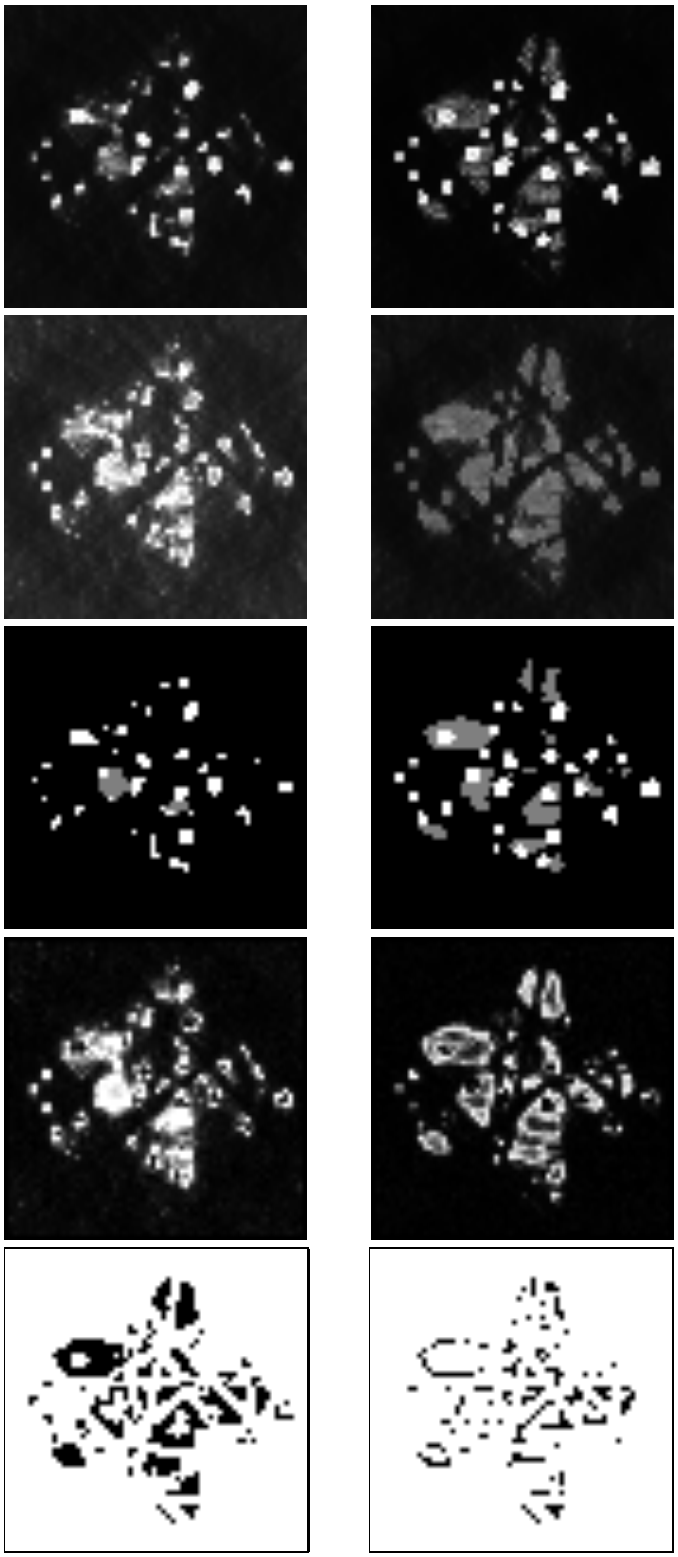


Figure 6: T. t. b.: mean continuous absorption coefficient (range displayed: 0 - 0.03 voxel^{-1}), its standard deviation (0 - 0.01 voxel^{-1}), segmentation, segmentation entropy (0 - 0.7) and correct labeling (white: correct); left column: setting 16-9-3, right column: setting 144-1-3.

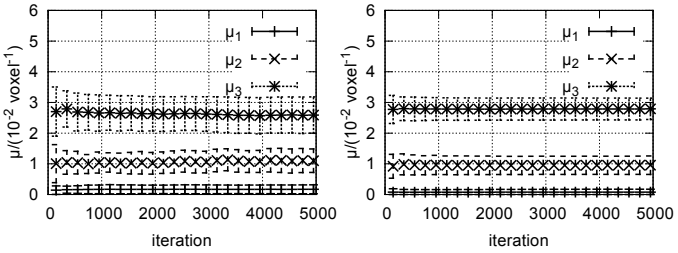


Figure 7: Development of the classified absorption coefficients μ_c over sample accumulation. Left: 20-9-3, right: 180-1-3

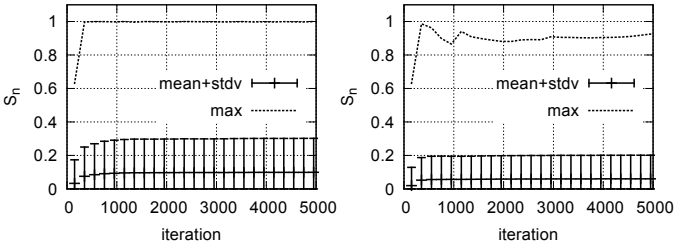


Figure 8: Development of the segmentation entropy over sample accumulation. Left: 20-9-3, right: 180-1-3

progresses. In Figure 8 the mean, standard deviation and maximum value of the segmentation entropy are plotted for the last replica in the chain. In addition, we calculate the fraction of correctly labeled voxels for each material (Fig. 9 and 10).

4. Discussion

Our approach differs in many aspects from other reconstruction methods. First, absorption coefficients in physical units are retrieved, not grayscale values. As for DART, segmentation is part of the reconstruction process. Currently we assume that the absorption coefficients of the materials are known. This assumption is not very strong because absorption coefficients can be measured by other means [15]. They could also be estimated as part of an extended algorithm but this is beyond the scope of the present paper. The most important difference to other reconstruction techniques is that it is possible to assess reconstruction errors and confidence metrics both globally and at the voxel level. We will now compare the re-

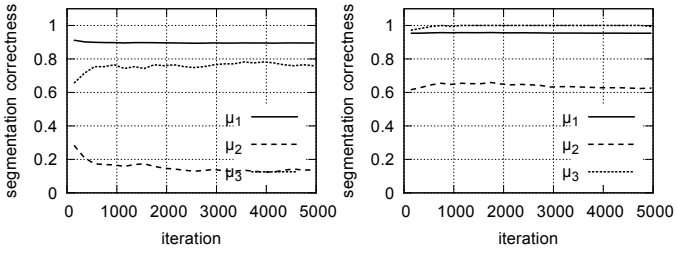


Figure 9: Fraction of correctly identified voxel labels as function of the number of drawn samples. Left: 20-9-3, right: 180-1-3.

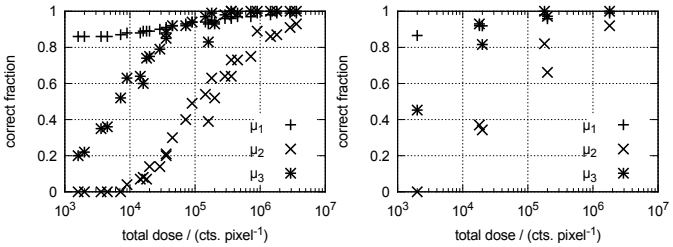


Figure 10: Class wise fraction of correctly identified voxel labels (i.e. "vacuum on a vacuum site" etc.) after 5000 samples have been drawn. μ_1 denotes the vacuum voxels, μ_2 carbon black and μ_3 represents the metal particles. Dose values are given by the product of dose for a single projection times the number of projections. Left: absorption coefficients for the phantom 0.0, 0.01 and 0.03 voxel^{-1}, right: absorption coefficients for the phantom 0.0, 0.02 and 0.06 voxel^{-1}.

constructions that we computed for different experimental conditions.

We find agreement between the different possibilities to assess the reliability of the reconstruction. As illustrated in Figs. 5 and 6 voxels with high standard deviation of the continuous absorption coefficient are also identified as problematic in the segmentation entropy map and *vice versa*. The problematic spots are often at the boundary between voxels with different absorption coefficient. This is to be expected, because the continuous specimen has been discretised via the detector pixelation and the "voxelation" of the model slice. In addition, in case of poor illumination and/or a small number of projections, plausible solutions are expected to be spread much wider in parameter space.

Our algorithm automatically takes care of the different experimental conditions. Figure 11 illustrates this for two voxels by plotting all values of the continuous absorption coefficient that have been visited during HMC. The voxel denoted as "center" is located inside a carbon particle, whereas the voxel termed as "edge" is located at the edge. It is clearly visible that for a smaller dose and/or fewer projections, the range of values is generally larger than under better experimental conditions. In particular, for setting 20-9-3, the absorption coefficient of the edge voxel scatters more strongly, while the coefficient for the center voxel is spread much less. Accordingly, the segmentation entropy is higher for the boundary voxel (see e.g. entropy plot in Figures 5, right column). In addition, the averaged values of the absorption coefficient and of the entropy (Figure 7 and 2) demonstrate, that better experimental parameters improve the confidence in the reconstruction/segmentation.

Up to this point, error and confidence quantification has been achieved without knowledge of the ground truth. This is another difference to existing methods, because this knowledge is used in almost every publication on reconstruction algorithms and the assessment of their reliability. Sometimes even the noise is not taken into account (see e.g. [23]). The present calculations are done for a known model configuration. Therefore, we can *additionally* compare the results to the ground truth. Generally we find good agreement between wrong labels and high values both for the standard deviation and segmentation entropy (see Figs. 5 and 6). Ambiguous edge voxels can have severe consequences because deduced sizes (and volumes) of particles have to be discussed with caution. An error of $\delta r = 1$ voxel for a particle of radius $r = 4$ results in a surface error of about 50 % and a volumetric error of about 100 %. Our algorithm identifies these regions, which allows the user to quantify their impact on, for example, size distributions.

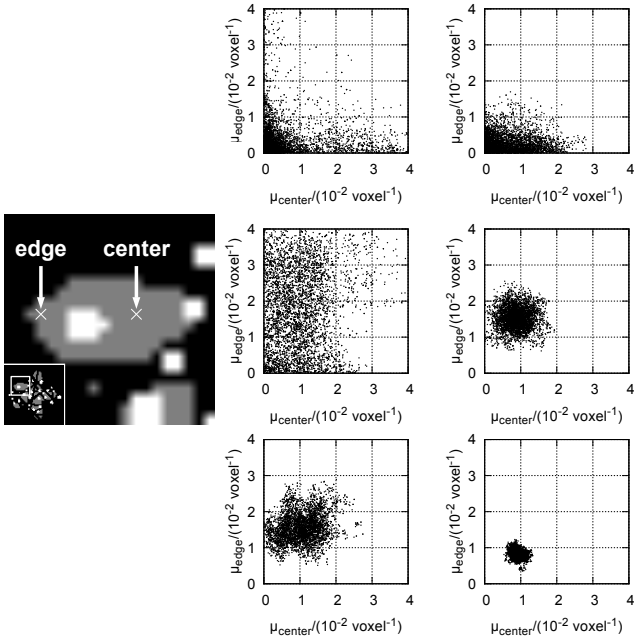


Figure 11: Scatter plots of the continuous absorption coefficient for two voxels as marked on the model image. Settings left column: 20-9-2, 20-9-3, 20-9-4; right column: 180-1-2 180-1-3, and 180-1-4.

In preparation of a real experiment, different conditions can be simulated to estimate possible reconstruction problems with some prior knowledge about the expected specimen morphology. For example, our calculations show that even for 90 projections and $I_0 = 1000$ cts./pixel, a typical setting in electron tomography, a large percentage of voxels is still labeled incorrectly. Nevertheless, such data are used in extensive data analyses (see e.g. [24]). On the other hand, a missing wedge does not necessarily lead to artefacts in the reconstruction. Figure 6 shows results which have be obtained using projection angles covering only 144°, a typical range for a standard tomography holder in transmission electron microscopy. Elongated particles are not visible. However, the reconstruction and segmentation is worse than the results for a full 180° tilting range (Figure 5). We attribute this to the reduced amount of experimental data resulting in worse statistics. A generalization to smaller ranges is not valid. Even the range of 144° might give other results for a phantom with changed morphology and/or absorption coefficients(see below).

It is noticeable that the reconstruction of the vacuum region is already quite good for settings, where the regions filled with matter are still reconstructed rather poorly. This demonstrates that the reconstruction depends strongly on the number of different materials involved. In this sense, data sets from binary specimens are much "easier" to reconstruct and segment, which should be taken into account when different reconstruction and segmentation approaches are compared.

It has to be pointed out that the reconstruction results presented here cannot be generalized in simple terms like more projections lead to better reconstructions. In regard of the reliability of the segmentation, different results are obtained if e.g. doubled absorption coefficients are used for the model and its reconstruction (0.02 voxel^{-1} and 0.06 voxel^{-1}, see Fig. 10). Stronger absorption leads to higher signal variations in the recorded projections with respect to intensity variations by noise (defined by the incident dose I_0). This results in a higher number of correctly labeled voxels. Of course, variation of the absorption coefficients is limited in real experiments (e.g. by varying the beam energy). If possible one should tailor the experimental conditions such that the detected intensities range from I_0 down to almost zero. However, zero count detection should be avoided, since opaque objects lead to ambiguous reconstructions.

5. Conclusions

The presented algorithm is fully self-contained and requires only known parameter values (absorption coefficients of the sample species, dose). No further user interaction or parameter settings are required. Tomographic reconstruction and segmentation are carried out simultaneously; quantitative assessment of reconstruction errors and confidence evaluation are built-in. Example calculations demonstrate the importance of the dose parameter and an optimized range of measured intensities. Although this may seem self-evident, a number of publications, especially in the field of electron tomography in materials science, exist where these experimental parameters are not reported, which makes it difficult to compare the quality of experimental data and results deduced from them. We therefore suggest that for better assessment of electron tomography data the declaration of I_0 and the dynamic range within the projection images should be mandatory, in the same way as it is common practice to indicate the projection directions.

References

[1] P. Midgley, E. Ward, A. Hungría, J. Thomas, Nanotomography in the chemical, biological and materials sciences, Chem. Soc. Rev. 36 (9) (2007) 1477–1494.

[2] J. M. Thomas, P. A. Midgley, C. Ducati, R. K. Leary, Nanoscale electron tomography and atomic scale high-resolution electron microscopy of nanoparticles and nanoclusters: A short survey, Prog. Nat. Sci.: Mater. Int. 23 (3) (2013) 222 – 234.

[3] A. Lange, M. P. Hentschel, A. Kupsch, Computertomographische Rekonstruktion mit DIRECTT: 2D-Modellrechnungen im Vergleich zur gefilterten Rückprojektion, MP Mater. Test. 50 (2008) 272.

[4] K. J. Batenburg, J. Sijbers, Dart: A practical reconstruction algorithm for discrete tomography, IEEE Trans. Image Process. 20 (2011) 2542–2553.

[5] R. Leary, Z. Saghi, P. A. Midgley, D. J. Holland, Compressed sensing electron tomography, Ultramic. 131 (0) (2013) 70 – 91.

[6] R. A. Crowther, D. J. DeRosier, A. K. ., The reconstruction of a three-dimensional structure from projections and its application to electron microscopy, in: Proc. Roy. Soc. A, Vol. 317(1530), 1970, p. 319–340.

[7] M. van Heel, M. Schatz, Fourier shell correlation threshold criteria, J. Struct. Bio. 151 (3) (2005) 250–262.

[8] H. Y. Liao, J. Frank, Definition and estimation of resolution in single-particle reconstructions, Structure 18 (7) (2010) 768–775.

[9] H. Friedrich, M. R. McCartney, P. R. Buseck, Comparison of intensity distributions in tomograms from BF TEM, ADF STEM, HAADF STEM, and calculated tilt series., Ultramic. 106 (1) (2005) 18 – 27.

[10] C. Kübel, M. Godehardt, R. Cieslinski, S. Rozeveld, Towards a quantitative understanding in electron tomography, Microscopy and Microanalysis 15 (2009) 602–603.

[11] W. Fortes, J. Sijbers, K. J. Batenburg, Practical error bounds for binary tomography, in: V. Cnudde (Ed.), Tomography of Materials and Structures - Book of Abstracts - Talks, 1st Int. Conf. on Tomography of Materials and Structures, July 1-5, 2013 Ghent (Belgium), 2013, pp. 97–100.

[12] K. Sauer, C. Bouman, A local update strategy for iterative reconstruction from projections, IEEE Trans. Sig. Process. 41 (1993) 534–548.

[13] A. C. Kak, M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, 1988.

[14] F.-Y. Wu, The potts model, Rev. Mod. Phys. 54 (1) (1982) 235.

[15] To be lifted in a forthcoming paper; the absorption coefficients c_k could be determined e. g. by measuring the absorption of respective thin films of known thickness.

[16] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (6) (1984) 721–41.

[17] N. Metropolis, M. Rosenbluth, A. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, J . Chem. Phys. 21 (6) (1957) 1087–1092.

[18] S. Duane, A. Kennedy, B. J. Pendleton, D. Roweth, Hybrid Monte Carlo, Phys. Lett. B 195 (2) (1987) 216 – 222.

[19] R. Neal, MCMC Using Hamiltonian Dynamics, CRC Press, 2011. doi:10.1201/b10905-6.
URL `http://dx.doi.org/10.1201/b10905-6`

[20] R. H. Swendsen, J.-S. Wang, Replica Monte Carlo simulation of spin glasses, Phys. Rev. Lett. 57 (21) (1986) 2607–2609.

[21] M. Habeck, M. Nilges, W. Rieping, Replica-exchange Monte Carlo scheme for Bayesian data analysis, Phys. Rev. Lett. 94 (2005) 018105.

[22] M. Wollgarten, R. Grothausmann, P. Bogdanoff, G. Zehl, I. Dorbandt, S. Fiechter, J. Banhart, Ruthenium deposition on CO_2-treated and untreated carbon black investigated by electron tomography, in: S. Richter, A. Schwedt (Eds.), EMC 2008 14th European Microscopy Congress 1–5 September 2008, Aachen, Germany, Springer Berlin Heidelberg, 2008, pp. 279–280.

[23] X. Wang, R. Lockwood, M. Malac, H. Furukawa, P. Li, A. Meldrum, Reconstruction and visualization of nanoparticle composites by transmission electron tomography, Ultramic. 113 (2012) 96–105.

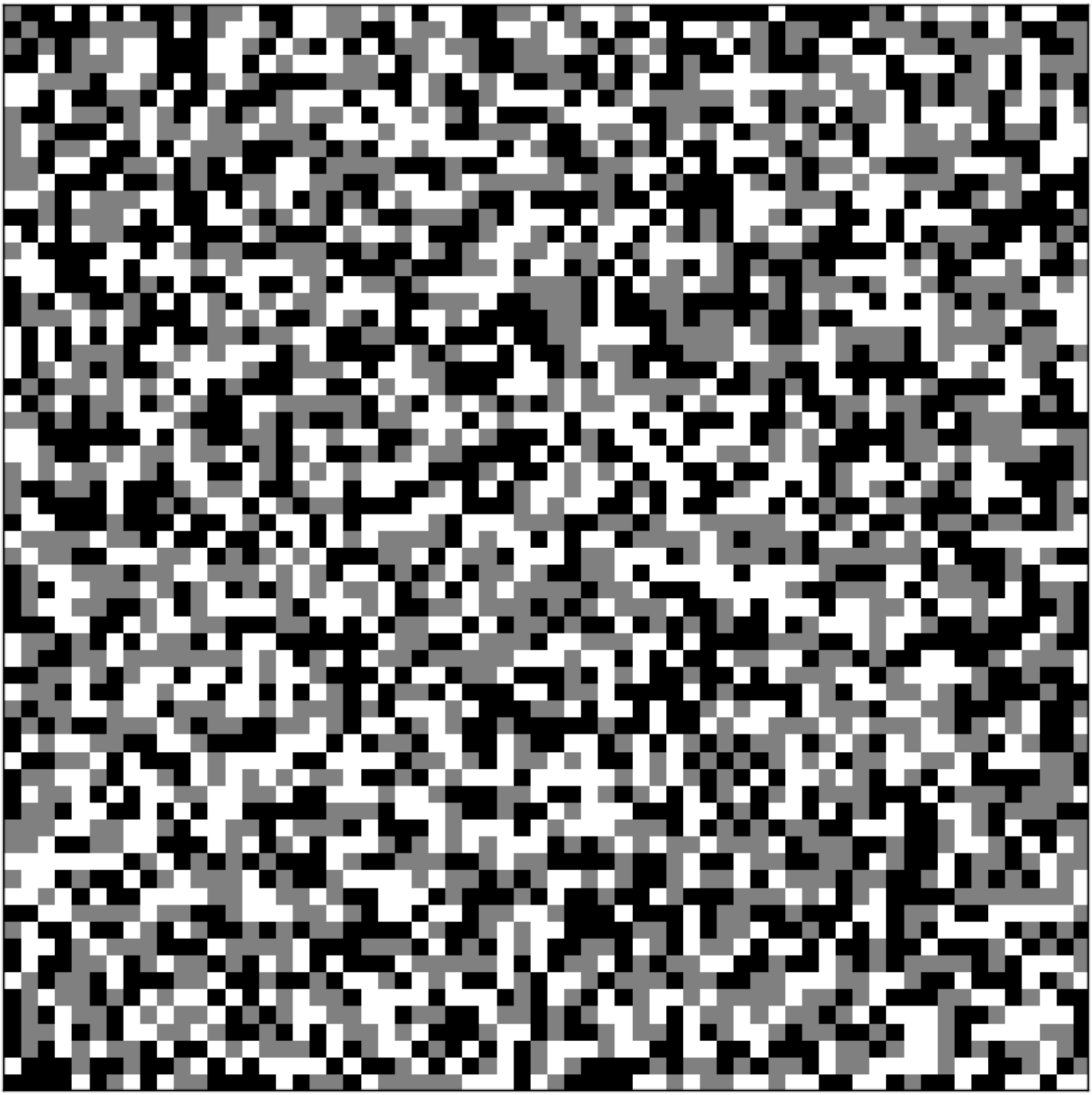
[24] R. Grothausmann, G. Zehl, I. Manke, S. Fiechter, P. Bogdanoff, I. Dorbandt, A. Kupsch, A. Lange, M. P. Hentschel, G. Schumacher, J. Banhart, Quantitative structural assessment of heterogeneous catalysts by electron tomography, J. American Chem. Soc. 133 (45) (2011) 18161–18171.

highlights.tex

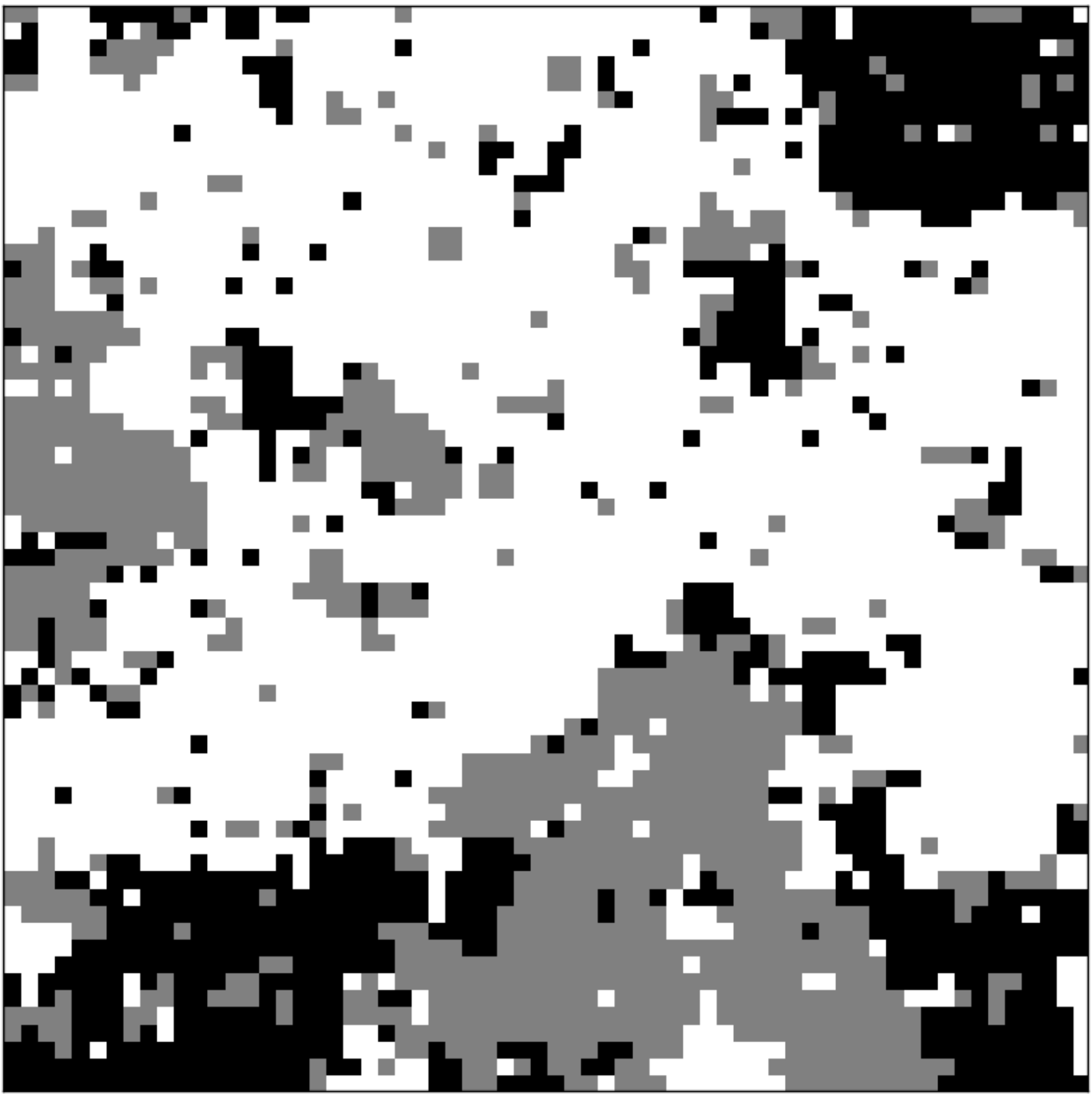
Click here to view linked References

```
This is TeX, Version 3.1415926 (Web2C 2009) (format=tex 2011.5.3)  10 FEB 2014 13:13
**highlights.tex
(./highlights.tex
! Undefined control sequence.
l.1 \section
        *{Highlights}
The control sequence at the end of the top line
of your error message was never \def'ed. If you have
misspelled it (e.g., `\hobx'), type `I' and the correct
spelling (e.g., `I\hbox'). Otherwise just continue,
and I'll forget about whatever was undefined.

! Undefined control sequence.
l.2 \begin
       {itemize}
The control sequence at the end of the top line
of your error message was never \def'ed. If you have
misspelled it (e.g., `\hobx'), type `I' and the correct
spelling (e.g., `I\hbox'). Otherwise just continue,
and I'll forget about whatever was undefined.

[1] )
Output written on highlights.dvi (1 page, 480 bytes).
```

potts_0_0.eps



potts_100_0.eps



***Detailed Response to Reviewers**

1.
The description of the method should be more explicit and easier to follow for audience without extensive mathematics background. A paragraph briefly describing the concepts would be helpful. How does the single step segmentation + reconstruction produce 3D volume ? Can this be described in a simple and explicit manner ?

In our method, reconstruction and segmentation are integrated into a unifying probabilistic model for the projection data. The segmentation comes in as an auxiliary quantity that allows us to couple neighboring absorption coefficients. This is implemented as a prior probability and different from the regularizers that are typically used in image reconstruction. We revised the introduction and theory sections in order to make this point clearer and more comprehensible.

2.
The section 2.4 should be rewritten so that the relevance to tomographic segmentation and reconstruction becomes clear.

Figure 10 - I suspect that the left column is for pixel at the edge while the right column is for pixel inside the metal particle? The figure should be labeled more clearly so that there is no guessing on what is what.

We revised section 2.4 to clarify how the segmentation and reconstruction is done in our approach.

We also changed Figure 10 and its caption (in addition, other pixels are used to make the effect even more evident).
Changed Text in 'Discussion' at end of first paragraph.

3.
Determination of particle boundaries (segmentation) does not appear to be much better than standard methods including simple. Or at least it is not clearly obvious why the current method would be more reliable than manually selecting a range of greyscale values from the histogram as the object boundary on a volume reconstructed by standard FBP method. The results in the reference X. Wang et al Ultramicroscopy 113 p. 96 may be a suitable to comment on.

The main benefit of our algorithm is not so much that it produces highly accurate segmentations but rather that it quantifies the reliability of the estimated

parameters without knowledge of the ground truth. This knowledge is, however, utilized in almost every publication on reconstruction algorithms and their reliability. Often even the noise is not taken into account.
We try to clarify this point and added/changed text in the discussion at the beginning of the second paragraph.

4.
Further on the point above: at the end the appearance (and volume) of particles appears to depend on the choice of the parameters for the algorithm. Figure 10 seems to demonstrate the concern: how does one know what are the correct parameters ? Even in simple FBP reconstructed volume one can define various metrics of correctness of the reconstruction.

This is probably a misunderstanding. The parameters changed in Figure 10 are experimental parameters (number of projections, angular interval and zero intensity) and not parameters of the algorithm. The only parameter we are (at the moment) free to set is the interaction constant J (for this, please see point 6 and 9).

The only metric we know, for a single reconstruction by (W)FBP is the Fourier shell correlation, which can be calculated without knowing the ground truth. But this gives a measure for the resolution only, not for the reliability of the absorption coefficient of a single voxel and its assigned labeling.

5.
The error in determination of edge voxels was discussed for example in X. Wang et al Ultramicroscopy 113 p. 96. The authors should consider adding the reference.

As far as we understand the paper of Wang et al., they attribute problems for the reconstruction of edge voxels to artifacts from filtering during Fourier back projection. This has no relation to our approach, since filters are not used.

6.
Parameter J in Eq. 5 seems to be almost arbitrary fudge factor. Is that so ?

J is not a fudge factor. It controls the coupling strength between the segmentation parameters of neighboring voxels and plays a role similar to the regularization parameter in SIRT. In general, it is a delicate problem to choose this parameter because the Potts model undergoes a phase transition at a critical J value above

which the segmentations no longer appear to be random but exhibit some clustering depending on the strength of J (see Figure 1 in the paper). However, it is possible to estimate this parameter in a data-driven fashion (please confer point 9 for further details). Another possibility would be to use cross-validation to determine the optimal J.

7.
How sensitive is the method to accurate alignment of the individual projections ? The accuracy of alignment seems to be of sufficient concern to spend the effort to fabricate precise fiducial markers (see for example Hayashida et. al. Micron 50 p. 29)

Indeed, one might expect some deteriorating influence from inaccurate alignment, also from defective values of the projection angles. But as pointed out in answer to comment 10, a different set of absorption coefficients and/or specimen morphology can lead to other results in terms of confidence. Thus, also here we are reluctant to check possible influences of all kinds as we do not want to promote generalization of our results. We rather would like to encourage that such tests are done, but keeping in mind certain experimental/specimen conditions.

8.
Similarly how important is a precise measurement of tilt angle for correct reconstruction ? The discussion in Hayashida et al Rev. Sci Instr 82 103706 may be relevant.

---

see previous comment. In addition, we would suggest to invest into an accurate protractor as has been used by Hayashida et al. We think a precise measurement of the tilting angles would make a discussion of related artifacts obsolete.

9.
In the reconstruction, it is assumed that the materials absorption constants are known as prior knowledge and the authors claim that their determination doesn't fall in the scope of this paper and they refer to a forthcoming paper . I understand that it may be difficult to give a complete explanation how they want to determine these constants, but I believe they must already include a small paragraph on this topic. Also the coupling constant J of the Potts model has to be known in advance.

The coupling parameters as well as the parameters of the distribution of the materials absorption constants can be estimated in a data-driven fashion. The latter step is

straightforward as it only involves standard parameter updates used to fit mixture models. Estimation of J is more involved but also possible. J can be viewed as a regularization constant and the particular difficulty to estimate this parameter stems from the fact that we cannot evaluate the partition function Z(J) [cf. Eq. (5)] analytically. However, in Mechelke and Habeck (Phys Rev E, 2012) we explain one possibility to determine this parameter using Bayesian model comparison. For the Potts prior also a simpler approach should be possible because we can estimate Z(J) accurately using, for example, the Wang-Landau algorithm.

In addition when the chemical elements in the specimen are known, one could determine the absorption coefficients experimentally by measuring the absorption of thin films of the respective species deposited with a known thickness.

10.
In the last column of figure 5, the correct labeling of the segmented pixels is shown. From this figure, it seems that a non-negligible amount of the pixels is classified incorrectly, even with a large number of projection data (tilt interval of 1º). This is only discussed briefly in the discussion section of the paper, but it would also be interesting to see a graph where the number of misclassified pixels is displayed as a function of tilt interval. Maybe you can also compare the quality of the reconstruction to the quality of more conventional reconstruction algorithms such as SIRT or DART. I also suggest to enlarge figure 5 to improve the visibility.

Yes indeed, there are a number of misclassified voxels. However, a graph showing the amount of correctly identified voxels is only of limited use as a generalization is not possible. For similar experimental parameters one might obtain much better results if the absorption coefficients are different, i.e. higher, leading to a better signal to noise ratio/contrast, which in turn reduces the volume of highly probable solutions within the parameter space.

The comparison to other algorithms/papers is also not straightforward. A large number of articles neglect noise and/or deal only with binary phantoms. For the remaining cases employing SIRT, ART and (w)FBP segmentation has to be done manually or by a heuristic approach. In our algorithm, the segmentation is inherent. DART also estimates a segmentation, but does not provide quantitative error/confidence metrics (only an overall estimate for binary tomography (Fortes et al. 2013), which we would like to stress as the main achievement of our approach.

11.
A very common problem in electron tomography is the presence of missing wedge artifacts. I therefore believe that the authors must also include this in their simulation experiments and discuss the impact on the quality of the reconstruction.

Yes, indeed. We added results for a limited range of 144°. Apparently, here artifacts are not visible. However, we are sure, that this cannot be generalized, even for the 144° range. We detailed this in the discussion.

12.
In the paper, the authors claim that the distribution of the absorption coefficient values are assumed to be Gaussian distributions. Can the authors comment when this assumption is valid and what may change when the density of the different materials varies largely inside the samples.

The assumption of a Gaussian distribution is not essential (besides, we assume an exponential distribution for the vacuum). If there is evidence for a more suitable distribution such as, for example, a log-normal distribution, it should be used instead of the Gaussian. We added a comment in the revised section 2.2. A formal justification for the use of a Gaussian distribution is given by a maximum entropy argument: The Gaussian distribution is the least biasing distribution assuming that we now the materials mean absorption coefficient and its variance.

Regarding the number of possibly influencing parameters, we would like to see that the authors of other reconstruction/segmentation approaches had spent time to investigate only a part of the possible influential aspects that you mention. Being serious again, regrettably, a number of papers deal with noise-free binary systems only. In fact, the absorption coefficient of the two materials used in our simulation is already quite different bearing in mind the exponential attenuation. The effect on the reconstruction is that the vacuum is reconstructed more accurately than the metal particle and finally the carbon. This is demonstrated by Fig. 9, but is also evident from Fig. 5.
If the absorption coefficient of a given material is not constant (e.g. gradients in an alloy of miscible elements) we expect that the standard deviation of the classified absorption coefficients increases and probably classification/segmentation collapses into a single class (not tested yet).

---

other changes:

Enlarged Figs. 4 and 5

Fig. 6 (now 7) changed caption to:
Development of the >classified< absorption coefficients >μ_c< over sample accumulation. Left: 20-9-3, right: 180-1-3

Changed unit in plots from pixel-1 (used for the detector only) to voxel-1 (used for volume elements of the reconstruction/model/sample)