

3D imaging of semiconductor components by discrete laminography

K.J. Batenburg^{*,†}, W.J. Palenstijn[†] and J. Sijbers[†]

^{*}*Centrum Wiskunde & Informatica, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands*

[†]*iMinds-Vision Lab, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium*

Abstract. X-ray laminography is a powerful technique for quality control of semiconductor components. Despite the advantages of nondestructive 3D imaging over 2D techniques based on sectioning, the acquisition time is still a major obstacle for practical use of the technique. In this paper, we consider the application of *Discrete Tomography* to laminography data, which can potentially reduce the scanning time while still maintaining a high reconstruction quality. By incorporating prior knowledge in the reconstruction algorithm about the materials present in the scanned object, far more accurate reconstructions can be obtained from the same measured data compared to classical reconstruction methods. We present a series of simulation experiments that illustrate the potential of the approach.

Keywords: Laminography, discrete tomography

PACS: 81.70.Tx

INTRODUCTION

Precise quality control inspections of soldering and assembly of electronic devices have become priority items in the electronics manufacturing industry. Many inspection systems for such devices make use of X-rays to form images (radiographs) that exhibit features representing the internal structure of the devices and connections. Interpretation of a single radiograph, however, is not obvious as depth information is completely lost. In an attempt to compensate for this shortcoming, laminographic systems have been built in which the object is viewed from a plurality of angles. The additional views enable these systems to partially resolve the ambiguities present in the X-ray radiograph.

In Fig. 1, a laminographic setup is depicted. The X-ray beam is rotated with respect to a vertical axis. The magnification and field of view are controlled by projecting the beam at different angles with respect to the axis of rotation. The detector rotates at the same speed as the beam, but 180° out of phase. By this arrangement, a single focal plane is created through which the beam travels for the entire duration of one rotation.

The focal plane can be simply reconstructed by shifting and adding the constituent projection images to bring structures into registration. The focal plane at an arbitrary depth can be determined and reconstructed [1]. Although this traditional shift-and-add tomosynthesis is simple and computationally effective, it suffers from significant blurring artefact due to the superposition of objects from other planes. Indeed, radiograph contributions from out-of-plane material, noise, and rotation of the X-ray source around the object during each laminograph formation, all contribute to imperfections in the representation of the object within the focal plane. Therefore, the development of im-

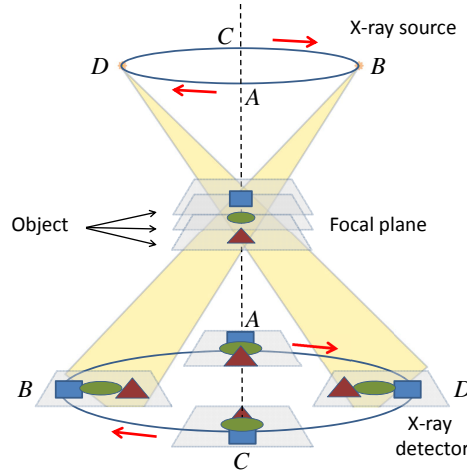


FIGURE 1. Laminography setup

proved reconstruction algorithms to correct for the out-of-plane blur is essential. Many approaches to reduce the blur artefact have been introduced [2, 3, 4]. The filtered back-projection approach, which consists of applying a filtering step to the projection data followed by a backprojection of the filtered data, is commonly used in practice. Iterative methods, such as the Simultaneous Iterative Reconstruction Technique (SIRT) result in fewer artefacts, at the expense of a longer reconstruction time [5]. Matrix Inversion Tomosynthesis (MITS) uses known imaging geometry and linear systems theory to deterministically separate in-plane detail from residual tomographic blur [6]. Despite these efforts, laminographic reconstruction still suffers from a limited depth resolution and various artefacts, mainly due to an inherent lack of a sufficient number of radiographs, which makes the reconstruction problem underdetermined.

The problem of reconstructing an image from a small number of radiographs has recently attracted considerable interest in the field of Compressed Sensing [7, 8]. In particular, it was proved that if the image is sparse, it can be reconstructed accurately from a small number of measurements with very high probability. In many images of objects that occur in practice, the image itself is not sparse, yet the boundary of the object is relatively small compared to the total number of pixels. In such cases, sparsity of the gradient image can be exploited by Total Variation Minimization [9, 10]. In this paper, we consider a different type of prior knowledge, where it is assumed that the object to be reconstructed consists of a small number (i.e., 2-5) of different materials, each corresponding to a characteristic, approximately constant gray level in the reconstruction. This type of reconstruction problem is known as *Discrete Tomography* (DT) [11]. Recently, the discrete algebraic reconstruction technique (DART), was introduced as an algorithm that can incorporate such prior knowledge [12]. The DART technique is an iterative reconstruction method that assumes the objects to be piecewise uniform with known densities. This reconstruction method has proved to be effective in reconstructing high quality images from only a limited number of radiographs [13, 14, 15].

In this paper, the use of DART for X-ray laminography of through-silicon Vertical

Interconnect Accesses (VIAs) is explored. We first provide an outline of the algorithmic ideas behind DART and discuss the particular obstacles that must be overcome when using DART for laminography applications. Subsequently, a series of simulation experiments is described that have been performed to assess the feasibility of using discrete tomography for the reconstruction of through-silicon VIAs from X-ray laminography data. The results of these experiments indicate that DART has strong potential for reducing the number of projection images required, and therefore the required scanning time.

METHODS

In this section, we outline the DART algorithm and discuss the specific considerations of applying DART to the reconstruction of copper through-silicon VIAs (TSVs) from X-ray laminography data. DART requires prior knowledge of the gray levels in the reconstructed image, which correspond to the different materials in the object. By effectively using this prior knowledge, a more accurate reconstruction can be obtained compared to algorithms that do not restrict the gray levels. For the experiments presented in this paper, we focus on the reconstruction of a simulated TSV for which the complete structure exhibits three gray levels, for void, silicon and copper.

As a starting point, we briefly review the iterative SIRT algorithm, which computes a grayscale reconstruction from a tomography (or laminography) dataset. A high level overview of the DART algorithm, which uses SIRT as a subroutine, is subsequently presented. The reader is referred to [12] for details about the implementation of DART.

The SIRT algorithm

The SIRT algorithm is a standard reconstruction technique for reconstructing grayscale images from a number of X-ray images, collected from different angles. It computes an approximate solution of the linear system $\mathbf{W}\mathbf{x} = \mathbf{p}$, where the vector \mathbf{x} contains the gray level (also called pixel value) for each pixel, the vector \mathbf{p} contains the measured projection data, and the matrix \mathbf{W} represents the projection operation (i.e., computing the product $\mathbf{W}\mathbf{x}$ yields the projections corresponding to the image \mathbf{x}). If no exact solution of this system exists, SIRT computes a solution for which the L^2 norm of the difference between the computed projection and the measured data, $\mathbf{W}\mathbf{x} - \mathbf{p}$ (also called projection error), is minimal, i.e., a least-squares solution (see [5] for details). Subsequently, each image pixel value is updated by adding a weighted average of the projection difference for all lines that intersect this pixel.

An important feature for using this algorithm within DART, is the fact that SIRT can also compute a grayscale reconstruction on a *subset* of the image pixels, where the intensity in the projections is only distributed among the selected subset. As we will see in the next section, DART uses this feature to focus the reconstruction on the object boundary.

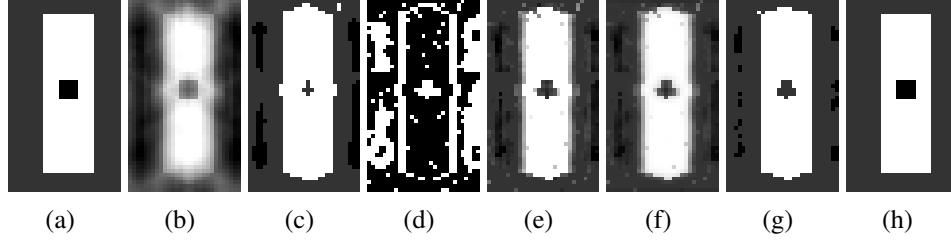


FIGURE 2. Illustrations of the different stages within an iteration of DART.

Overview of DART

In this section, we will introduce the DART algorithm and give an overview of the basic algorithmic steps. The DART algorithm is an iterative method, which starts from an initial image estimate and then repeatedly updates this image to improve the quality of the reconstruction.

As an example, suppose that we want to reconstruct a silicon sample containing a number of TSVs embedded in the silicon, some of which possibly containing voids. A cross-section of one of these TSVs is shown in Fig. 2(a).

Before applying DART, one needs an estimate of the number of gray levels in the reconstruction (i.e., the number of materials in the sample), as well as the actual gray levels. In this case we assume that three different densities are present: void, silicon and copper. An estimate of the actual graylevels can be hard to obtain, in particular when only a small number of projections are available. As a first step before applying DART, a conventional SIRT reconstruction is usually computed to obtain information about the materials in the sample and their gray levels. A SIRT reconstruction of the TSV is shown in Fig. 2(b).

Although the TSVs have a constant density, the SIRT reconstruction exhibits a spectrum of gray values. In addition, the shape of the reconstructed structure is clearly distorted, due to the limited angular range of the available projections. The range of gray levels in the reconstruction immediately presents a problem when the reconstructed image needs to be segmented. Segmentation is commonly performed by thresholding, but it is not obvious at all which threshold should be chosen in this case. The DART algorithm starts from a thresholded SIRT reconstruction, and then iteratively improves upon the current segmentation. Although a threshold also has to be chosen for DART, its choice is of minor importance to the final result. In this example, we choose the thresholds to be exactly in the middle between the gray level of the background and copper, and between copper and silicon, respectively. Fig. 2(c) shows the thresholded SIRT reconstruction. The thresholded reconstruction shows that pixels of the interior of the object that are not too close to the boundary are assigned the correct segmentation class (void, silicon, or copper). Pixels that are close to the boundary can be detected automatically from the thresholded SIRT reconstruction, by checking if any of the surrounding pixels belongs to a different segmentation class. We refer to these pixels as boundary pixels, and to the remaining pixels as non-boundary pixels. Fig. 2(d) shows the set of boundary pixels computed from Fig. 2(c). Note that even some non-boundary pixels have the wrong gray

level in the thresholded SIRT reconstruction, when compared to the original TSV image.

We now turn back to the SIRT reconstruction in Fig. 2(b). The non-boundary pixels are assigned the gray level that corresponds to their respective material class (void, silicon, or copper). Next, the SIRT algorithm is used again, but only the *boundary* pixels are allowed to vary. The non-boundary pixels are kept fixed at their discrete levels. In this way, the number of variables in the linear equation system $\mathbf{W}\mathbf{x} = \mathbf{p}$ is significantly reduced, while the number of equations remains the same. Fig. 2(e) shows the result after 10 SIRT iterations for the boundary pixels. The SIRT step on the boundary can cause heavy fluctuations of the gray levels. A smoothing step is performed on the boundary to remove this roughness. The result is shown in Fig. 2(f). Applying a stronger smoothness filter leads to less noise and smoother boundaries in the reconstruction, while losing some fine single-pixel details. We now repeat the same steps as before, starting from the reconstruction in Fig. 2(f). The result of the threshold step is shown in Fig. 2(g). It is already very clear that the quality of the segmentation has improved considerably in just one iteration. By performing several iterations, the reconstruction quality is further improved, as shown in Fig. 2(h) for 10 full DART iterations.

DART for laminography

Compared to electron tomography, where the DART algorithm has already been applied successfully to the reconstruction of a range of nanomaterials [13, 14, 15], the laminography case is substantially more challenging from an implementation point-of-view. As the viewing directions for the series of projection images do not share a common plane, a fully-3D volume representation of the object is needed to compute the projections. Subdividing the volume into a stack of slices that are each processed independently is not possible in this case. As a consequence, the entire image volume must be kept in memory all at once, which imposes significant memory requirements.

Computing a series of projections for a given voxel object is computationally expensive. As each voxel contributes to each projection, the total number of mathematical operations that must be performed to compute a series of projections is proportional to the product of the number of voxels and the number of projection images. This computation can easily take several hours for a series of 200 projections of a $1024 \times 1024 \times 1024$ voxel volume, when run on a standard PC. In recent years, Graphics Processing Units (GPUs) have developed into highly parallel multi-processors that can process up to 1000 operations simultaneously. This high degree of parallelism is possible for the tomography computation, as the *same* operations must be performed to a large number of data elements.

In [16], we presented a GPU-based algorithm for the projection and backprojection operations that attains a speedup of two orders of magnitude compared to a conventional CPU implementation. By using the 3D version of this code for laminography reconstruction, the running time for a complete 3D DART reconstruction of a $500 \times 500 \times 50$ volume can be reduced to around 15 minutes on an NVIDIA GTX580 GPU. A key limitation of this GPU-based approach is the fact that for maximal efficiency, the entire voxel volume and the series of projections must be stored in the RAM memory on the

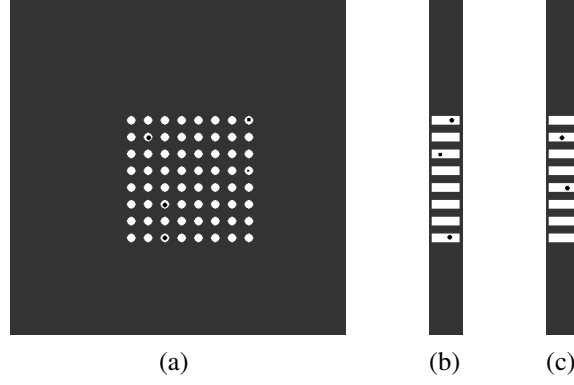


FIGURE 3. Cross-sections of the simulation phantom.

graphics card itself. This memory is limited to 6GB–12GB for the high-end NVIDIA Tesla cards, limiting the size of volumes for which this high efficiency can be obtained.

EXPERIMENTS

To validate the use of DART for X-ray laminography of TSVs, a series of simulation experiments was set up. The experiments are based on a simulated 3D phantom object, which represents a slab of silicon containing a regular grid of copper TSVs, some of which contain voids. This results in a voxelized phantom containing three gray levels for the voids, silicon, and copper.

The phantom volume contains a stack of 50 slices, each consisting of 500×500 voxels, corresponding to a thin slab of silicon. The TSVs have a diameter of 11 voxels, whereas voids contained in these TSVs have a diameter of 6 voxels. Cross-sections of the phantom in three directions are shown in Fig. 3. The projection image recorded for each angle covers only a part of the silicon slab. As a result, only the central part of the slab is covered by all projections. Fig. 4(a-b) illustrate the coverage of different parts of the phantom by the set of projections, where the white region is covered by all angles and the black region is not covered at all. The experimental results presented in the next section focus on a particular Region-of-Interest (ROI) within the silicon slab that is covered by all angles. The extent of this region is depicted in Fig. 4(c-d).

Synthetic projection data of the phantom were generated by simulating a parallel beam laminography setup, as follows: First, the Radon transform of the phantom was computed, resulting in a series of projection images, where the width of a detector cell was taken to be the same as the voxel size of the phantom. Next, (noiseless) CT projection data were generated where a mono-energetic X-ray beam was assumed. The projections were then perturbed with Poisson distributed noise where the number of counts per detector element I_0 (flat field) was varied from 10^4 (high noise) up to 10^6 (low noise). Next, the linearized noisy projection data was obtained by dividing the CT projection data by the flatfield intensity and computing the negative logarithm. In this way, simulated projection images were obtained for varying signal-to-noise ratios. Finally, the simulated, noisy CT images were reconstructed using both the SIRT

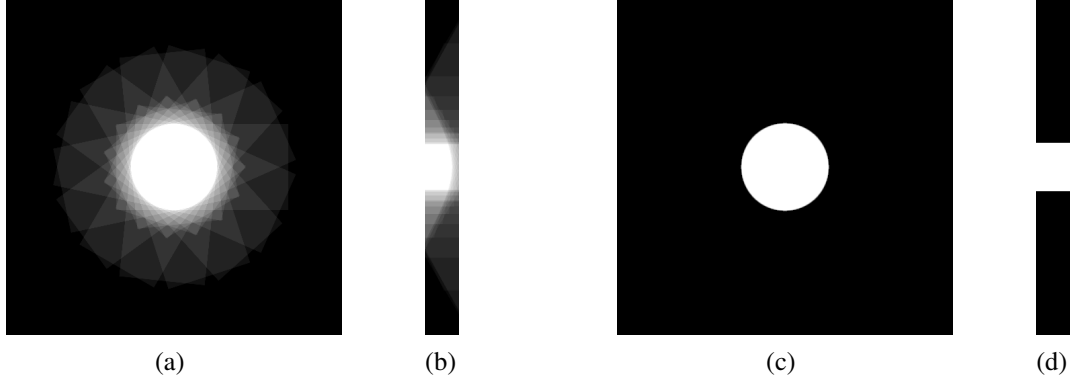


FIGURE 4. (a-b): Illustration of the coverage of the cross-section shown on the left by the series of projections. (c-d) Region-of-Interest that is covered by all projections.

algorithm and the DART algorithm.

The SIRT reconstructions were thresholded to obtain an image containing just three gray levels, using the well-known method of Otsu [17]. The resolution of the reconstruction depends strongly on the position within the reconstructed volume, varying both in depth and in the lateral position within the slab. In the presentation of the results we focus on two slices through the reconstructed volume, parallel to the silicon slab (xy-plane) and through the slab (xz-plane).

To compare the quality of the reconstructions, the *Number of Misclassified Pixels* (NMP) is used, which is defined as the number of misclassified pixels in the ROI of the segmented reconstruction.

RESULTS

Two series of experiments have been performed, comparing the reconstruction quality of SIRT and DART, where the SIRT algorithm is considered as a representative method from the current class of "relatively advanced" methods applied in X-ray laminography. The first set of experiments is aimed at investigating the dependency of the reconstruction quality on the number of projection images, while the second set is aimed at investigating the influence of noise.

Varying the number of projections

In a first series of experimental results, we provide a qualitative impression of the reconstruction quality obtained by both SIRT and DART, when varying the number of projection images. Fig. 5 shows reconstructed cross-sections through the silicon in two directions, where the number k of projections is varied from $k = 5$ up to $k = 60$.

From the results shown in Fig. 5, it is clear that within the region of silicon that is covered by all projections, DART allows to reduce the number of projections down to 10 while still obtaining reconstructions of good visual quality.

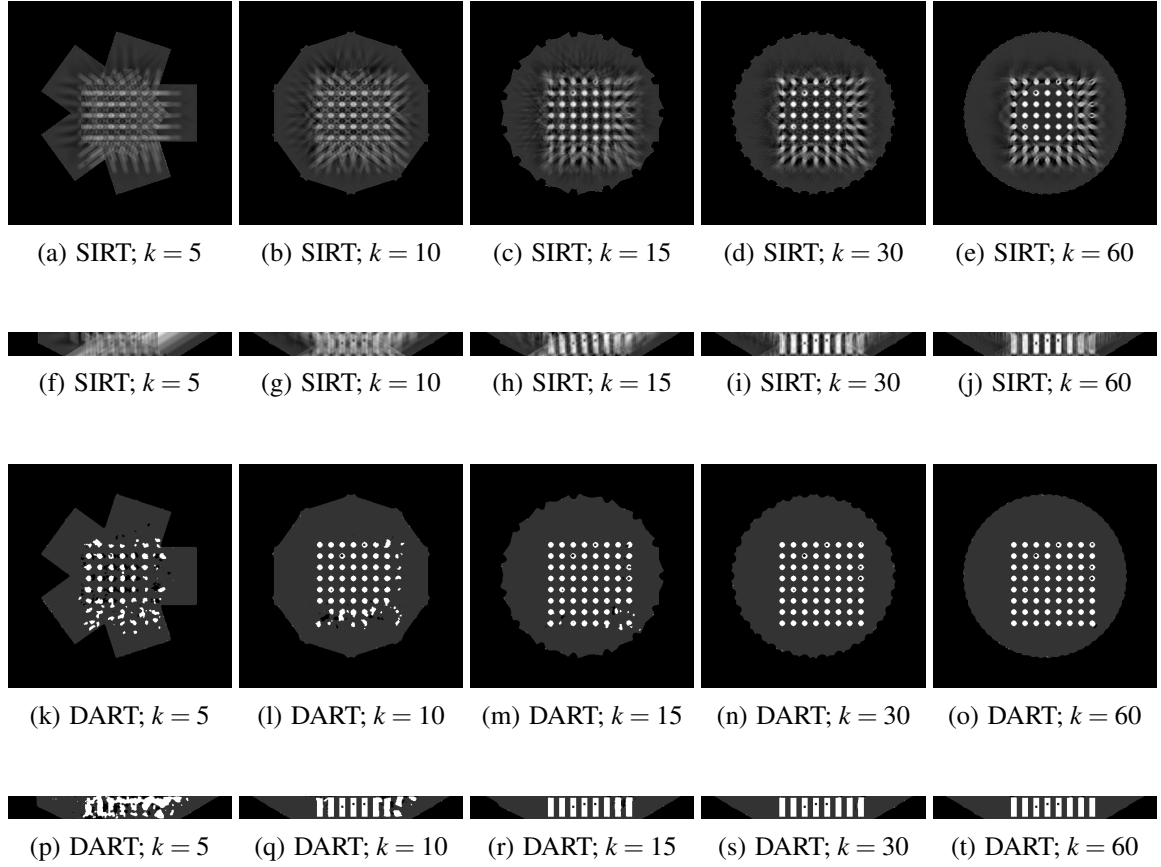


FIGURE 5. Cross-sections of the reconstructed volume computed by SIRT and DART from a varying number of projections; First row: xy-slice, SIRT; Second row: xz-slice, SIRT; Third row: xy-slice, DART; Fourth row: xz-slice, DART.

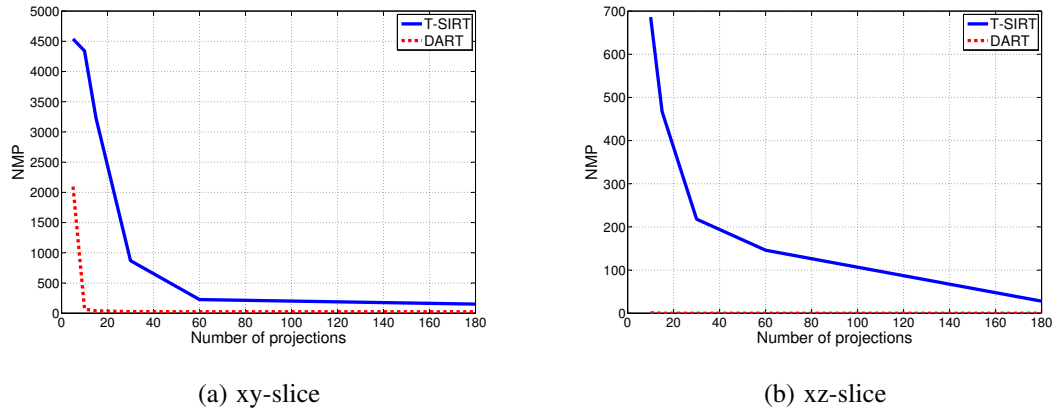


FIGURE 6. Number of Misclassified Pixels (NMP) in the reconstructed ROI for a varying number of projections, for both SIRT and DART. Left: error in xy-slice; Right: error in xz-slice.

In Fig. 6, a *quantitative* comparison is made between the SIRT and DART reconstructions, plotting the NMP in the ROI as a function of the number of projections. Fig. 6a shows the NMP in a slice parallel to the silicon slab (xy-slice), while Fig. 6b shows the results for a slice in the direction of the VIAs (xz-slice). It can be observed that even when using only 10 projections, the error for DART already drops to a level which is not attained by SIRT, even when using as many as 180 projections.

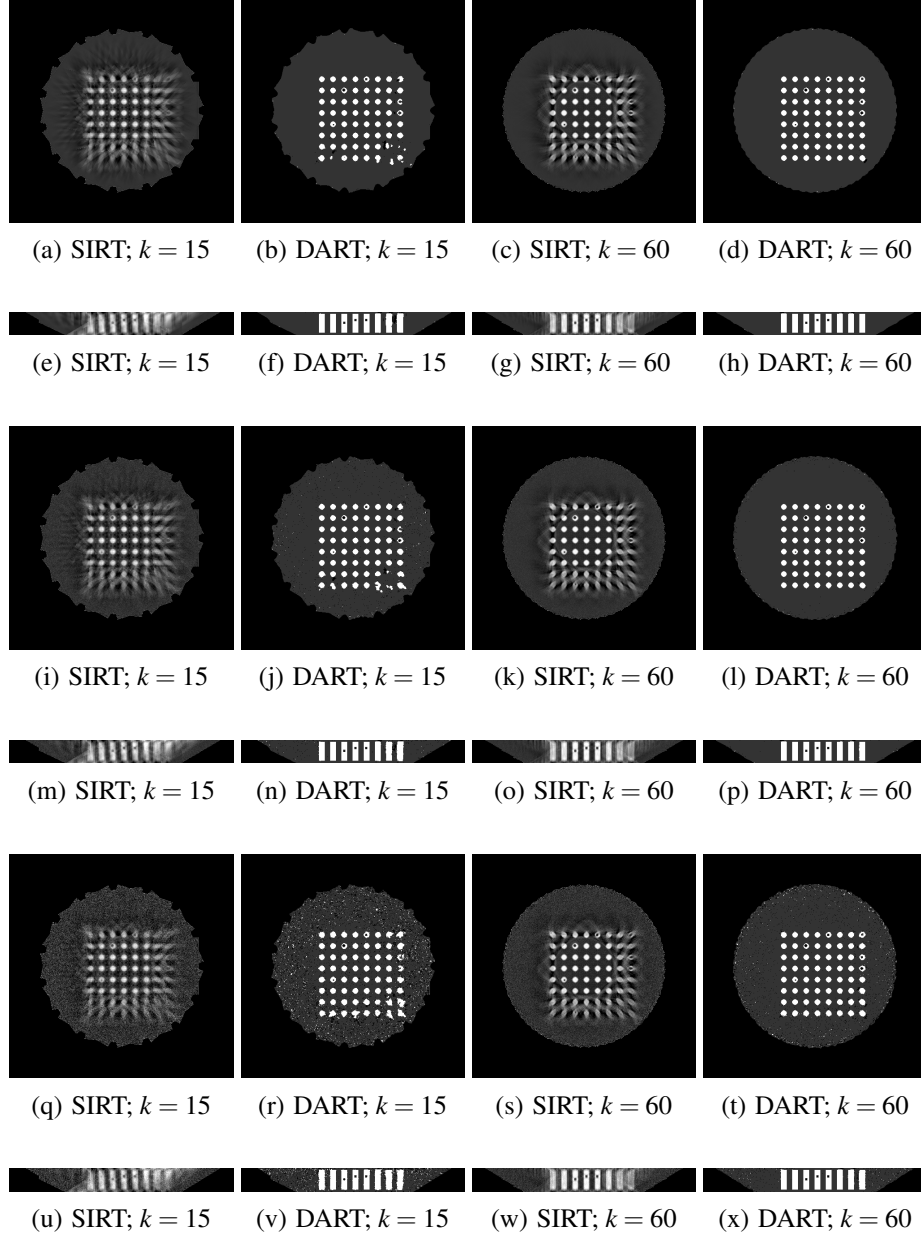


FIGURE 7. Cross-sections (xy-slice) of the reconstructed volume computed by SIRT and DART from varying noise levels; Row 1-2: no noise; Row 3-4: moderate noise ($I_0 = 100000$); Row 5-6: high noise ($I_0 = 10000$).

Varying the noise level

In a second series of experiments, the robustness of DART is compared to SIRT for varying noise levels in the projection data, for two sets of projection angles ($k = 15$ and $k = 60$). A qualitative impression of the reconstruction results is shown in Fig. 7.

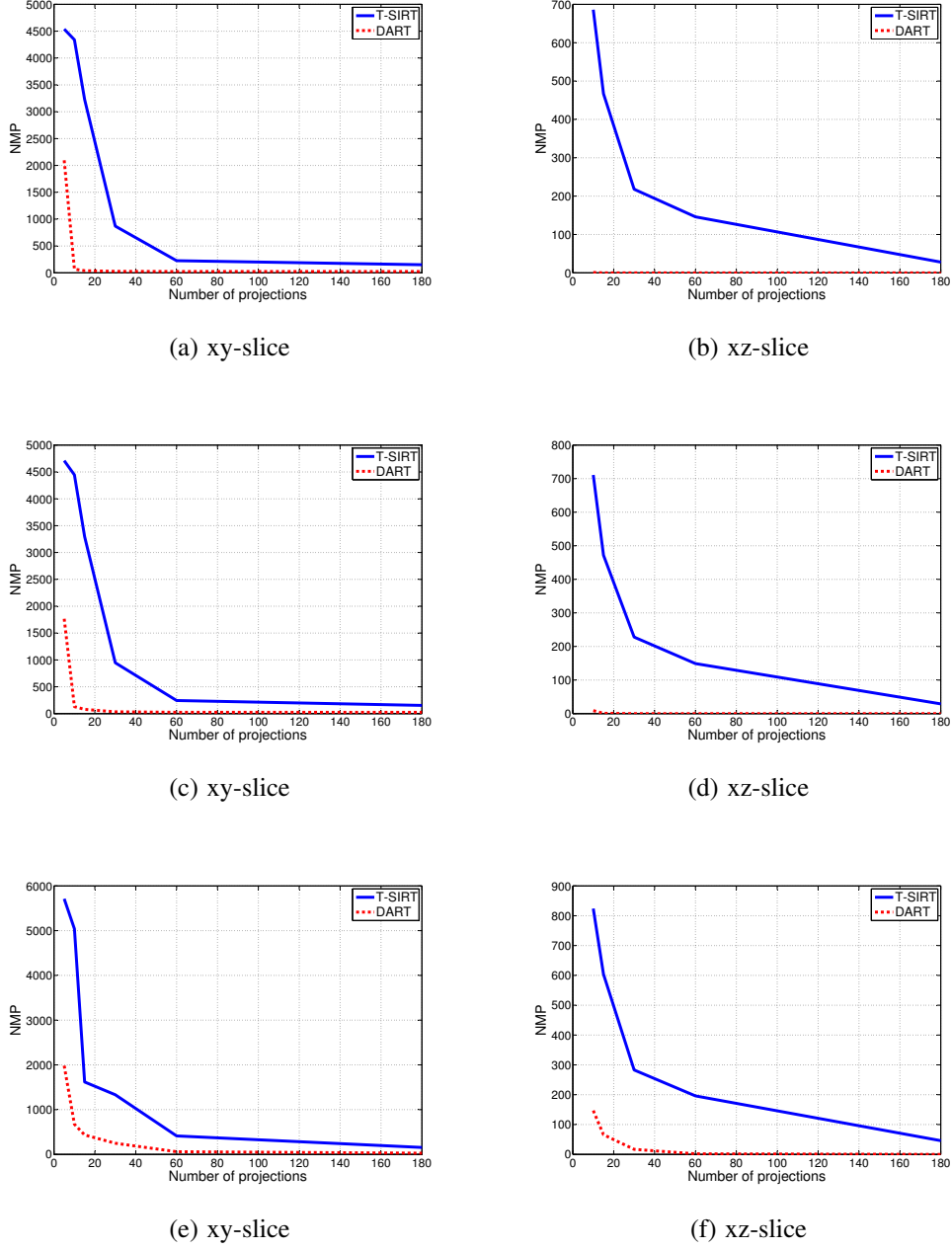


FIGURE 8. Number of Misclassified Pixels (NMP) in the reconstructed ROI for a varying number of projections, for both SIRT and DART. Left: error in xy-slice; Right: error in xz-slice. Row 1: no noise; Row 2: moderate noise; Row 3: high noise.

The results in Fig. 7 illustrate that even if the projection data contains a fair amount of noise, the ROI can still be reconstructed accurately using DART, even from a low number of $k = 15$ projections.

Fig. 8 shows a quantitative comparison of SIRT and DART for a varying number of angles and for three different noise levels: no noise, moderate noise ($I_0 = 100000$) and high noise ($I_0 = 10000$). In all cases, DART clearly outperforms SIRT in terms of both the NMP and the visual quality of the reconstruction.

DISCUSSION AND CONCLUSIONS

The results presented in this paper provide a first indication that discrete tomography could be successfully applied to the reconstruction of electrical components from X-ray laminography measurements. To assess the true practical value of the technique in this domain, the DART algorithm must now be tested on experimental datasets. These real-world experiments will include the effects of mechanical instabilities, beam-hardening and other physical sources of error that cause a mismatch with the idealized linear projection model employed in DART.

However, given the fact that DART has already proven itself on experimental data in several domains, including the highly challenging domain of electron tomography, we are positive about the prospects of such experiments, which may eventually lead to a strong reduction in the number of projections required for a full quality assessment, and therefore in a reduction of the scanning time.

ACKNOWLEDGMENTS

KJB acknowledges support from the Netherlands Fund for Scientific Research (NWO), project number 639.072.005.

REFERENCES

1. Z. Kolitsi, G. Panayiotakis, V. Anastassopoulos, A. Scodras, and N. Pallikarakis, *Medical Physics* **19**, 1045–1050 (1992).
2. D. N. Ghosh Roy, R. A. Kruger, B. Yih, and P. Del Rio, *Medical Physics* **12**, 65–70 (1985).
3. D. P. Chakraborty, M. V. Yester, G. T. Barnes, and A. V. Lakshminarayana, *Radiology* **150**, 225–229 (1984).
4. Z. Kolitsi, G. Panayiotakis, and N. Pallikarakis, *Medical Physics* **20**, 47–50 (1993).
5. J. Gregor, and T. Benson, *IEEE Trans. Image Proc.* **27**, 918–924 (2008).
6. J. T. Dobbins III, and D. J. Godfrey, *Physics in Medicine and Biology* **48**, R65–R106 (2003).
7. D. L. Donoho, *IEEE Transactions on Information Transfer* **52**, 1289–1306 (2006).
8. E. Y. Sidky, C.-M. Kao, and X. Pan, *J. X-ray Sci. Tech.* **14**, 119–139 (2006).
9. A. Chambolle, *Journal of Mathematical Imaging and Vision* **20**, 89–97 (2004).
10. E. Y. Sidky, and X. Pan, *Phys. Med. Biol.* **53**, 4777–4807 (2008).
11. G. T. Herman, and A. Kuba, editors, *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999.
12. K. J. Batenburg, and J. Sijbers, *IEEE Transactions on Image Processing* **20**, 2542–2553 (2011).
13. S. Bals, K. J. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo, *Nano Letters* **7**, 3669–3674 (2007).

14. K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, *Ultramicroscopy* **109**, 730–740 (2009).
15. T. Roelandts, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, *Ultramicroscopy* **26**, 96–105 (2012).
16. W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, *Journal of structural biology* **176**, 250–253 (2011).
17. N. Otsu, *IEEE Trans. Systems, Man, and Cybernetics* **9**, 62–66 (1979).