2014

# Compressed sampling strategies for tomography

Yan Kaganovsky
*Duke University*

Daheng Li
*Washington University in St Louis*

Andrew Holmgren
*Duke University*

HyunJu Jeon
*Duke University*

Kenneth P. MacCabe
*Duke University*

*See next page for additional authors*

## Recommended Citation

Kaganovsky, Yan; Li, Daheng; Holmgren, Andrew; Jeon, HyunJu; MacCabe, Kenneth P.; Politte, David G.; O'Sullivan, Joseph A.; Carin, Lawrence; and Brady, David J., ,"Compressed sampling strategies for tomography." Journal of the Optical Society of America A: Optics and Image Science, and Vision. 31,7. 1369-1394. (2014).
https://digitalcommons.wustl.edu/open_access_pubs/3537

## Authors

Yan Kaganovsky, Daheng Li, Andrew Holmgren, HyunJu Jeon, Kenneth P. MacCabe, David G. Politte, Joseph A. O'Sullivan, Lawrence Carin, and David J. Brady

# Compressed sampling strategies for tomography

**Yan Kaganovsky,[1,]* Daheng Li,[2] Andrew Holmgren,[1] HyungJu Jeon,[1] Kenneth P. MacCabe,[1]**
**David G. Politte,[3] Joseph A. O'Sullivan,[2] Lawrence Carin,[1] and David J. Brady[1]**

[1]*Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA*
[2]*Department of Electrical and Systems Engineering, Washington University, St. Louis, Missouri 63130, USA*
[3]*Mallinckrodt Institute of Radiology, Washington University, St. Louis, Missouri 63110, USA*
**Corresponding author: yan.kaganovsky@duke.edu*

Received February 18, 2014; revised April 19, 2014; accepted April 20, 2014;
posted April 25, 2014 (Doc. ID 206542); published June 9, 2014

We investigate new sampling strategies for projection tomography, enabling one to employ fewer measurements than expected from classical sampling theory without significant loss of information. Inspired by compressed sensing, our approach is based on the understanding that many real objects are compressible in some known representation, implying that the number of degrees of freedom defining an object is often much smaller than the number of pixels/voxels. We propose a new approach based on quasi-random detector subsampling, whereas previous approaches only addressed subsampling with respect to source location (view angle). The performance of different sampling strategies is considered using object-independent figures of merit, and also based on reconstructions for specific objects, with synthetic and real data. The proposed approach can be implemented using a structured illumination of the interrogated object or the detector array by placing a coded aperture/mask at the source or detector side, respectively. Advantages of the proposed approach include (i) for structured illumination of the detector array, it leads to fewer detector pixels and allows one to integrate detectors for scattered radiation in the unused space; (ii) for structured illumination of the object, it leads to a reduced radiation dose for patients in medical scans; (iii) in the latter case, the blocking of rays reduces scattered radiation while keeping the same energy in the transmitted rays, resulting in a higher signal-to-noise ratio than that achieved by lowering exposure times or the energy of the source; (iv) compared to view-angle subsampling, it allows one to use fewer measurements for the same image quality, or leads to better image quality for the same number of measurements. The proposed approach can also be combined with view-angle subsampling.

*OCIS codes:* (110.6960) Tomography; (110.7440) X-ray imaging; (340.7430) X-ray coded apertures; (340.7440) X-ray imaging; (110.6955) Tomographic imaging.
http://dx.doi.org/10.1364/JOSAA.31.001369

## 1. INTRODUCTION

Computerized tomographic imaging [1] or projection tomography is the process of estimating a multidimensional function/distribution from measurements of its line integrals. The function of interest usually represents some physical property of an object, and here we refer to it as simply the "object." Sampling in tomography [2] has classically relied on a simplified description of the relationship between angular sampling and the desired resolution, assuming the object is described by images or image volumes and represented on a rectangular grid. The spacing between angles should be no bigger than the pixel/voxel size divided by twice the largest dimension.

There has been significant recent interest in reducing the number of measurements for inferring an underlying signal/image, leading to the field of compressed sensing (CS). The seminal work of Candès *et al.* [3] and Donoho [4] demonstrated that if the underlying signal/image is sparse in its native basis, or in some orthonormal basis (e.g., wavelets), then the number of required measurements for perfect signal/image recovery may be much smaller than that anticipated by classical sampling theory. If the image or basis coefficients are not exactly sparse but rather are compressible (many coefficients are negligibly small, but not exactly zero), then highly accurate signals/images can still be reconstructed. In the following, for simplicity we refer to images as being "sparse," with the understanding that in practical settings near-sparsity ("compressibility") is more common.

The idea behind CS for imaging is that we have a very large set of basis images, of which only a small *but unknown* set has non-negligible coefficients. Therefore, the measurements need to be maximally sensitive to as many of the basis images as possible, and images that differ in a small number of basis images must be as distinguishable as possible in measurement space. The optimal measurements are therefore as unstructured as possible in terms of their sensitivity to basis images, but they should also maximally separate sparse representations in measurement space. The reconstruction of the image from these unstructured measurements is enabled by using a sparsity prompting penalty such as $\ell_1$ norm or total variation (TV) [3,4]. This penalty incorporates the prior knowledge that only a small set of coefficients is significant and prevents the problem from being ill posed.

Many studies have considered sampling and compression in tomographic systems. Image estimation from incomplete data has been part of x-ray imaging since the beginning of computed tomography [5]. The earlier studies considered incomplete data due to *physical constraints* that include limited views [5–7] and few view projections [7–10]. Convex optimization strategies used in initial CS studies have been used in tomography for over 30 years [11,12], and the use of the $\ell_1$ norm to invert tomographic data for geometrically sparse objects predates modern CS terminology (e.g., see [8]). Initial CS studies were inspired by some of the earlier work in tomography, as can be seen by Candès *et al.*'s motivating example relating to a few-view tomographic reconstruction in a parallel-beam geometry [13]. CS, in turn, continues to inspire the tomographic community, and some of the later works consider designed undersampling of the object with respect to view angles [7,14–17] and use TV or $\ell_1$ constrained reconstructions; the potential benefit is, of course, reducing radiation dose.

Tomographic measurements are embedded on spaces with lower dimensionality than the physical spaces of the measured object. Typically one seeks to reconstruct 3D objects from measurements distributed on 2D manifolds (for a given source configuration, measurements are performed on a 2D angular grid). To acquire an additional data dimension, one often varies the measurement geometry as a function of time (e.g., considering different source configurations over time, and for each, data are measured on a 2D manifold). This thereby creates a 3D measurement space to match the 3D object space.

One form of compressive measurements attempts to recover a 3D object based on a single 2D measurement. For example, we have shown that 3D hyperspectral [18], diffraction [19], and x-ray scatter [20,21] images may be reconstructed from 2D data. We have also analyzed compressive sampling for reconstruction of 3D objects with conventional optics [22]. Most recently, we have shown that 3D video data cubes may be constructed from 2D frames [23], thus using compressive tomography to perform the reconstruction with respect to time. With specific reference to projection tomography, our previous work has focused on the use of reference structures [24] and coded apertures [25] to physically implement quasi-random unstructured codes consistent with compressive measurements. We have termed this class of compressive measurement systems "compressive tomography," characterized by estimating an $N$-dimensional *physical* object from measurements in $M < N$ dimensional *physical* space.

In this paper we pursue a different means of reducing the quantity of measured data. The framework is based on the more-conventional approach of performing measurements on 2D manifolds, and acquiring a third dimension by varying the source-object orientations (view angles) over time. However, in this setting we seek to reduce the total number of measurements (line integrals) beyond what is considered possible by classical sampling theory. There are two sampling rates that can be controlled to reduce the total number of measurements: (i) reducing the number of detectors in the 2D manifold and (ii) reducing the number of source-object orientations (view angles) considered. Although reducing the number of view angles has been studied due to its straightforward implementation [7,14], reducing the number of detectors has been overlooked, and the goal of this work is to investigate the utility of this approach. Subsampling of either detectors or view angles can be done uniformly, pseudorandomly, or using a prescribed pattern. We compare several sampling strategies using multiple figures of merit. We show that detector subsampling can outperform view-angle subsampling, and in addition has important practical advantages. Detector subsampling can be accomplished by structured illumination, i.e., imposing a structured pattern on the illuminating beam using coded apertures on either the source or detector side, depending on the application.

Our motivation is x-ray imaging where in addition to transmitted photons that travel directly from source to detector, there are also a considerable number of photons that scatter from within the object via various physical mechanisms, such as Compton and coherent scatter, and reach the detectors by passing through a nondirect trajectory. Scatter may be considered as a dominant source of noise, such as in medical scans using cone beam computed tomography [26,27], where the measurement of scatter is done in a separate scan to estimate the noise level, which is then utilized by the reconstruction algorithm to improve image quality. Alternatively, measuring scatter may provide valuable information about the molecular structure of the imaged object [28] and can be used for material discrimination and characterization. For example, this is used in security applications, where dangerous substances such as explosives can be detected [29,30]. Other applications include quality control in industrial manufacturing [31] and illegal drug detection [32]. Regardless of whether it is considered as a source of noise or information, the measurement and estimation of scattered radiation is often desirable in addition to transmitted radiation.

Reducing the number of transmission detectors allows the use of new sensors to exploit and measure scattered radiation during the same scan. We do not explicitly consider measurement of scatter here, but note that the reduction in the number of detectors is motivated in part by the opportunity to measure scattered signals in the newly available sensing space. In medical applications, detector subsampling may be implemented by blocking source rays, thereby exposing the patient to less harmful radiation, which also results in reduced scattered radiation. In this case, the level of scatter will decrease per view angle, while keeping the same transmitted radiation in the remaining rays, leading to improved SNR. An alternative approach to reducing radiation dosage is to reduce the power of the source or to reduce exposure time while keeping all measurements. However, this will further lower the SNR and consequently lower image quality [33].

Detector subsampling can also potentially enable multisource object illumination for compressive tomography. In this case, coded apertures are designed such that any detector pixel only measures a single ray and the illumination pattern of each source undergoes quasi-random detector subsampling. This provides the connection between the strategies studied here and compressive tomography as defined above and will be the subject of future work.

There are several key differences between the compressive measurements explored in this paper for projection tomography and the general framework of CS as described in [3,4,34–36]. Theoretical CS studies [3,4] that provide the strongest results make use of completely random measurements, which are maximally spread out in the space where the object of interest has a sparse representation in some basis (e.g., native basis or wavelet basis), and suggest measuring different linear combinations of *all* basis coefficients. In contrast, in projection tomography, measurements are by nature confined to a line in physical-location space. In addition, most tomographic systems have the property that reducing the number of measurements implies reducing the number of angular samples or line integrals acquired, as opposed to measuring linear combinations of line integrals. This is a restrictive constraint that closely ties tomographic system design to sampling theory. In addition, these systems can have a wide range of designs that depend on the physics and/or geometry of the target, each design providing different additional limitations to how the object can be interrogated.

Several theoretical CS studies [34–36] rely on the assumption that the set of measurements and the set of basis functions (leading to a sparse representation of the object) can both be described as projections onto an orthonormal basis. Instead of using completely random combinations of standard measurements, these studies suggest using a random subset of standard measurements, which is more practical. However, tomographic measurements are linearly dependent and are generally not easily described by projections onto an orthonormal basis. One option is to use the slice-projection theorem, which enables one to describe integrals along parallel lines in terms of radial lines in the 2D Fourier space of the object [13]. Geometries employed in practical x-ray systems greatly differ from the parallel beam geometry (e.g., real systems typically employ fan-beam or cone-beam illuminations). When the object is sampled according to classical sampling theory, it is possible to reorder the line integrals into parallel lines using interpolation [37]. However, this approach *cannot* be done in a straightforward way when the object is *significantly undersampled* in actually used geometries, and the same procedure would result in a large number of missing samples in detector space, which would render the direct Fourier transform on detector measurements quite challenging. Extensions of the slice-projection theorem to practical geometries lead to complicated representations [38].

The remainder of the paper is structured as follows. In Section 2 we state the general sampling problem and present the sampling strategies that will be explored in this paper. We review classical sampling theory, as a point of reference. Details about the forward operator used in this study are discussed in Section 3. In Section 4 we present a singular value decomposition (SVD) analysis, which gives insight into the degree of ill-posedness and sensitivity to noise introduced by the various sampling strategies. It also clarifies which features of the object a system design is most sensitive to, and how much it is matched to sparse objects. In Section 5 we analyze the proposed sampling strategies using CS theory. We present results for synthetic data in Section 6, demonstrating that even with significant undersampling, the object can still be reconstructed with good quality; we also discuss how the proposed approach is related to previously used strategies. The most promising strategies are verified in Section 7 on real data coming from an experimental x-ray laboratory at Duke University and from a Siemens Sensation 16 scanner for human imaging located at St. Louis Children's Hospital. To enable a simple theoretical analysis, we focus mainly on 2D objects, but the results can be easily extended to 3D objects, as demonstrated by the examples for the patient dataset.

## 2. PROBLEM STATEMENT

### A. Preliminaries

For simplicity, we focus on 2D functions, denoted $f(\mathbf{x})$, with $\mathbf{x} = (x_1, x_2)$. Depending on the specific application, the measured line integrals are usually arranged in a certain pattern, which we refer to as a "scanning geometry," or simply "geometry."

In the translate–rotate (parallel-beam) geometry [1], the lines come in parallel bundles and are parametrized by the direction of the beam defined by angle $\varphi$ and the displacement $s$ perpendicular to the central ray in a bundle [see Fig. 1(a)]. The measurements in this configuration correspond to the Radon transform defined by

$$\mathbf{R}f(\varphi, s) = \int_{\mathbf{x}\cdot\boldsymbol{\theta}=s} f(\mathbf{x})\mathbf{dx}, \qquad \boldsymbol{\theta} = (\cos\varphi, \sin\varphi). \tag{1}$$

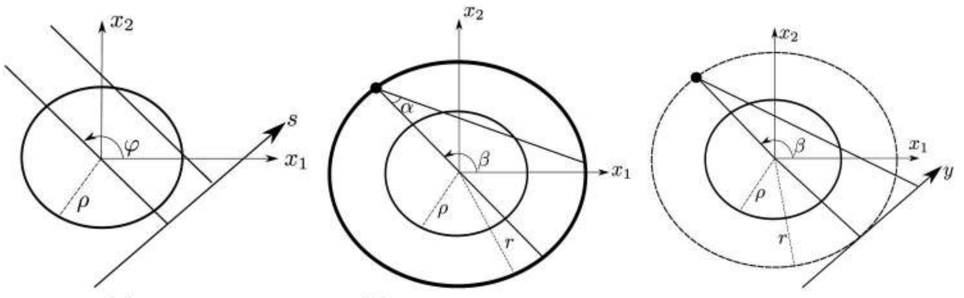
A commonly used physical configuration is the fan-beam geometry, with lines arranged into fans emanating from focal points (point sources) distributed over a circle around the origin; these are parametrized by the angular location of the source $\beta$ and the angle $\alpha$ between each ray and the "central ray" that goes through the origin [see Fig. 1(b)]. This gives rise to the fan-beam transform $\mathbf{D}$, defined by

$$\mathbf{D}f(\beta, \alpha) = \int_{L(\beta,\alpha)} f(\mathbf{x})\mathbf{dx}, \tag{2}$$

where $L(\beta, \alpha)$ is the line connecting the source and detector. The relation to the Radon transform is given by

$$\mathbf{D}f(\beta, \alpha) = \mathbf{R}f(\beta + \alpha - \pi/2, r\sin\alpha), \tag{3}$$

where $r$ is the distance of the source from the origin. An alternative configuration is shown in Fig. 1(c), where a linear displacement is used instead. Note that there are other types of geometries that arise in different applications, e.g., the outer



(a) Parallel (b) Curved Detector Fan (c) Flat Detector Fan

Fig. 1. Different scanning geometries in 2D and the corresponding parameters for the line integrals. The disk of radius $\rho$ defines the image domain where the function $f$ is to be reconstructed from measurements of the line integrals. (a) Parallel beam (translate–rotate) geometry, (b) curved detector fan-beam geometry, and (c) flat detector fan-beam geometry.

circle in Fig. 1(b) could be centered about the source, but we do not aim at giving an extensive list.

In real physical configurations only a discrete set of parameter values can be measured. Although in the continuous case the fan-beam transform is a reparameterization of the Radon transform, in the discrete case the grids of these two geometries will be different. For example, trying to reorder fan-beam measurements with only a few view angles into parallel lines will result in many missing detector and view angle samples in the parallel-beam geometry. The Radon transform plays a central role in classical sampling theory, as discussed in Section 2.C and Appendix A, but the translate–rotate geometry is no longer used in modern computed tomography (CT) systems, and we shall consider here a sampled fan-beam transform.

## B. Sampling Strategies

The canonical 2D geometries described in Section 2.A constitute a natural starting point for any study of new sampling strategies, as they have already been studied extensively and are well understood [2]. In this work we shall focus on the fan-beam geometry, due to its practical use in x-ray CT systems. The same concept is also applicable to spiral CT with evidence provided in Section 7.B. There are two variables according to which undersampling is possible: (1) detector location and (2) view angle ["view angle" is the angle between the line from the source to the origin and a reference axis, e.g., the angle $\beta$ in Fig. 1(b)]. We shall compare view and detector undersampling while setting the variable that is not being undersampled to the minimal required resolution according to classical sampling theory. There are four basic sampling strategies that are considered, as summarized in Fig. 2: (i) uniform view (UV) subsampling, in which views are selected uniformly by keeping only every $N$th view angle, and the complete detector set is used; (ii) random view (RV) subsampling, in which views are selected randomly while the complete detector set is used; (iii) uniform detector (UD) subsampling, in which detectors are selected uniformly by keeping only every $N$ th detector, and the complete view angle set is used; and (iv) random detector (RD) subsampling, in which detectors are selected randomly while the complete set of view angles is used. The view subsampling strategy (UV) has been previously considered in [7]. The nonstructured (random) detector subsampling has not been considered before, to the best of our knowledge. We compare to the case where both view and detector sampling are done according to the minimal resolutions dictated by classical sampling theory, which we refer to as "complete measurements."

As explained in the introduction, the nonstructured (quasi-random) detector subsampling is more suitable for sparse representations of the object, and clear evidence for that will be given in Sections 4–7. We will consider a few variations. (i) Dynamic random detector (DRD) subsampling is where detectors are selected randomly per view angle with a different selection for each view angle. (ii) Semidynamic random detector (SDRD) subsampling is the same as DRD, but with a number of different detector arrays (selections) that is less than the number of views so the same detector selections are reused in a periodic way. This can correspond to using a few different masks, which is easier to implement in practice. (iii) Static random detector (SRD) subsampling is the same as DRD, but with the same detector selection used for all views.

## C. Sampling Conditions Based on the Peterson–Middleton Theorem

Sampling theory for tomography originated from classical sampling theory, which permits the reconstruction of a band-limited function from its values on a regular grid or lattice. However, the imaged domain is always of finite support, and strictly speaking there are no band-limited functions of compact support. Therefore, a widely used approach is to consider *essentially* band-limited functions instead, i.e., the frequency content outside some frequency band is considered negligible. Given a desired resolution of some size, sampling conditions are derived for stable reconstruction of details down to that size. Here "*stable*" means that a small error in measurements does not result in an arbitrary large error in the reconstructed function.

Assume the object of interest $f$ is contained inside a disk of radius $\rho$ outside which $f = 0$. Sufficient sampling conditions for several different 2D geometries are summarized in Table 1 for reconstructing details of down to size $\delta$. There are two different resolutions that are required according to this theory, one for the view-angle intervals and one for detector intervals (see Fig. 1 for the definitions of the parameters).

For completeness, in Appendix A we outline the derivation of the sampling conditions. The overview of the derivations emphasizes that nowhere in the theory is it assumed that the object to be measured may be represented sparsely, and therefore the theory does not account for all aspects of the signal model leveraged here.

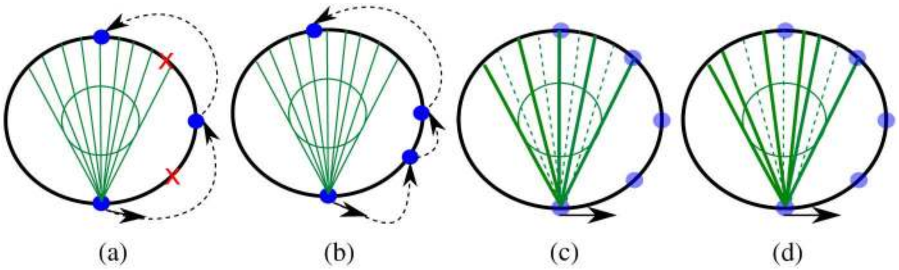


Fig. 2. Different sampling strategies under consideration demonstrated on a fan-beam geometry. The blue circles denote source locations (corresponding to view angles), and lines denote source–detector pairs (for clarity, only the lines for one source location are shown for each case). The solid and dashed lines correspond to measured and nonmeasured line integrals, respectively. The large circle represents the source trajectory, and the small circle represents the imaged domain. (a) Uniform view (UV) subsampling, (b) random view (RV)-angle subsampling, (c) uniform detector (UD) sampling, and (d) random detector (RD) subsampling.

**Table 1. Sufficient Sampling Conditions for Reconstructing Details down to Size $\delta$ for the Scanning Geometries in Fig. 1 According to Classical Sampling Theory**

| Geometry | View Interval | Detector Interval |
|---|---|---|
| Parallel | $\Delta\varphi = \delta/2\rho$ | $\Delta s = \delta/2$ |
| Fan—curved detector | $\Delta\beta = (r + \rho/r)\delta/2\rho$ | $\Delta\alpha = \delta/2r$ |
| Fan—flat detector | $\Delta\beta = (\delta/2\rho)(r + \rho/r)(1 - \rho^2/r^2)$ | $\Delta y = \delta/2$ |

## 3. FORWARD MODEL STUDIED

In all the following studies we use the flat detector fan-beam geometry detailed in Fig. 3, which corresponds to the experimental system used in Section 7. The noiseless measured line integrals (sinogram data) are given in the matrix form

$$\mathbf{y} = \mathbf{H}\mathbf{f}, \tag{4}$$

where $\mathbf{f} \in \mathbb{R}_{+}^{N}$ represents the object of interest, which is discretized using a 2D Cartesian grid (image) and rearranged as a vector. In x-ray CT, $\mathbf{f}$ is the attenuation of x rays per unit length inside the imaged object in units of $\mathrm{mm}^{-1}$ (the full physical model for x-ray imaging will be discussed in Section 6). $\mathbf{H} \in \mathbb{R}_{+}^{M\times N}$ contains the intersection length of each ray with each pixel inside the image (for x-ray CT, it is in units of mm), with row index corresponding to line integral and column index corresponding to image pixel. A complete dataset is chosen according to the classical sampling theory of Section 2.

In the following theoretical study, $\mathbf{f}$ is represented on a $128 \times 128$ ($N = 128^2$) Cartesian grid corresponding to a rectangular region of size 116 mm × 116 mm, but the image domain is restricted to a disk of radius $\rho = 58$ mm in order to be consistent with the theory presented in Section 2.C. In the system in Fig. 3, the detector panel and source are not equally spaced from the axis of rotation as assumed by theory ($R_x \neq r$ in Fig. 3), so we calculate the sampling rates for a symmetric system where these distances are equal to $r$ and then project back to the given detector panel location. According to Table 1, the minimal resolution for detectors is 0.377 mm and for views is 0.494°, which for the given detector size of 193 mm and a full scan of 360° corresponds to 512 detectors and 727 views, a total of $M = 372{,}224$ rays. These sampling rates are used as a starting point from which we downsample according to different sampling strategies, and we refer to them as the "complete set" of measurements. Note that the off-axis displacement $\Delta$ in Fig. 3 is very small, and its effect on the sampling rates can be neglected. This displacement is often used in order to break the symmetry in the line integrals when source and detector locations are exchanged and to make sure there is no redundancy in measurements when source locations cover the entire 360° of the circle. We would also like to note that the number of measurements in the complete set is much greater than the number of unknowns (pixels/voxels) due to the ill-posed nature of the problem [39].

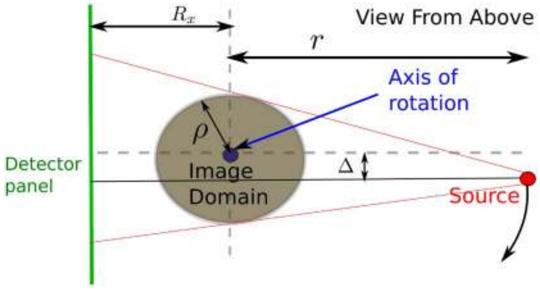


Fig. 3. Fan-beam geometry with flat detector panel used throughout the analysis and simulations, which corresponds to the experimental system described in Section 7. $\rho = 58$ mm, $\Delta = 1.2$ mm, $r = 484.6$ mm, $R_x = 290.2$ mm.

## 4. SINGULAR VALUE ANALYSIS

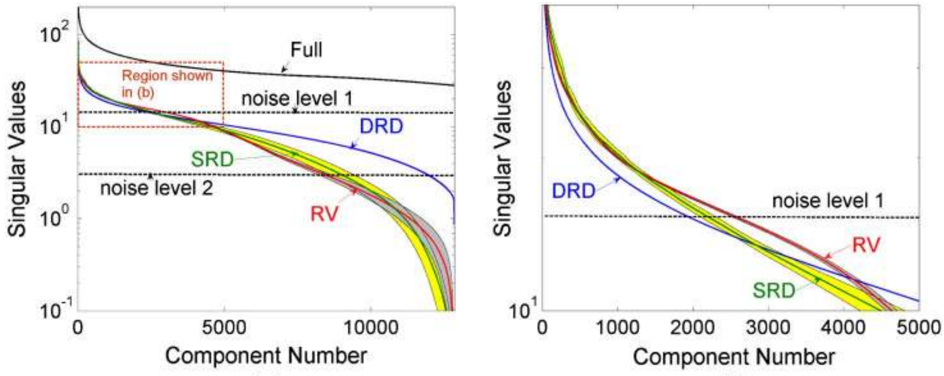
The SVD of $\mathbf{H}$ is a valuable tool for analyzing ill-posed problems [2]. When comparing two competing sampling strategies for a given noise level, if nothing is known about the object that is being measured, then the strategy that has more components lying above that noise level is preferred, since it captures more orthogonal components of the object. However, many times some prior information about the object is known. In this case, it is not enough to judge the performance based on the number of singular components above the noise level; one needs to consider which features of the object the system is most sensitive to, which are given by the right singular vectors corresponding to the highest singular values. Ideally, the features that the system is most sensitive to will belong to the class of objects that is of interest. An alternative approach is to look at the features that cannot be recovered reliably, which represent the effective null space of the system when noise is present.

In this work, we assume the measured object can be represented by an image that is sparse in some known representation. In the following section we shall show that for a sparse object we would like the singular components that are most sensitive to noise to be as incoherent (noise-like) as possible, i.e., as noncompressible as possible. In this way, we make sure that the effective null space of the system does not include sparse or compressible objects that are of interest, so we can distinguish between two different sparse objects in measurement (sinogram) space. A more detailed explanation will be provided in Section 4.B.

### A. Singular Values

Here we calculate the singular values of $\mathbf{H}$ in Eq. (4) corresponding to different random sampling strategies discussed in Section 2.B. We start with the minimal resolution required by classical sampling theory for both views *and* detectors that defines the full $\mathbf{H}$ matrix that corresponds to the complete set of measurements. Then, the image $\mathbf{f}$ is undersampled according to the different strategies, and this determines the remaining rows of $\mathbf{H}$ that are used to compute the SVD. The selected measurements are drawn uniformly at random using standard pseudorandom number generators.

Figure 4 presents the singular values of $\mathbf{H}$ for random subsampling by a factor of 16. The random selection of measurements was repeated 20 times for each subsampling strategy, and the mean and standard deviation for this ensemble were



(a) (b)

Fig. 4. Singular values for undersampling by a factor of 16 using the different random subsampling strategies described in Section 2.B. RV, random view; SRD, static random detector; DRD, dynamic random detector. The mean and standard deviation (std) for 20 different random selections are indicated by solid lines and shaded strips, respectively. For clarity the boundaries of the shaded strips correspond to 1 std above and below the mean (gray strip for RV and yellow strip for SRD). Note that the means for DRD and SDRD are indistinguishable from each other in this scale and therefore the latter is omitted here. Also, the std for DRD and SDRD are indistinguishable from the mean of DRD and are therefore omitted here. Also included are the singular values for the complete dataset ("Full"). The forward model is described in Fig. 3, and a 128 × 128 image resolution is used. A circular domain was used to be consistent with classical sampling theory. (a) All singular components and (b) only first 5000 components.

computed. The forward operator on the Haar wavelet domain obtained by right multiplying **H** by the discrete inverse wavelet transform matrix (DWTM) has the same singular values due to the unitary property of the DWTM. Two different examples for noise levels are shown in Fig. 4(a) and indicated by two dashed lines denoted "noise level 1" and "noise level 2." For the higher noise level (no. 1), RV sampling has more singular components above noise level than DRD sampling, and therefore the former is preferred in this regime [see also Fig. 4(b) for a zoom-in view of the first 5000 singular components]. However, for the lower noise level (no. 2), DRD sampling is much better than RV, and the difference in the number of components that are above noise level is significantly greater than for the higher noise level. This analysis shows that if nothing is known about the object, the DRD approach is better than RV for the lower noise regime (at least for the x-ray geometry considered). A SDRD sampling scheme that uses only 10 different detector arrays (or masks) performs very similarly to the DRD approach and offers potential simplification of system design (note that SDRD is indistinguishable from DRD within the scale of Fig. 4 and was therefore omitted). In the SDRD approach, the mask is still changed for every view, but there are less masks than views, so the masks are reused periodically (10 masks instead of 727 for DRD). The number of required masks should change according to the complexity of the imaged object; however, this goes beyond the scope of this work (see discussion in Section 8). It is also interesting to note that the differences between SRD sampling (where the same detector array is used for all views) and RV sampling are minor compared to DRD. In the low-noise regime SRD and RV are the worst subsampling strategies according to this analysis.

The next natural question to ask is this: do we necessarily gain from sampling randomly instead of uniformly? In order to answer this question, Fig. 5 compares the singular values of **H** with uniform sampling versus different random sampling approaches. It can be seen in Fig. 5(a) that UV sampling is comparable to RV sampling, so there is no advantage to random sampling. A similar result can be seen in Fig. 5(b) for detector sampling, where UD sampling is only slightly better than SRD sampling. However, DRD sampling is significantly better than UD and SRD, at least in the low-noise regime. The condition number (ratio of largest singular value to smallest singular value) for DRD is almost two orders of magnitude smaller than for UD or SRD. What sets DRD apart from the other approaches is the fact that different detectors are sampled at each view angle. In conclusion, according to this analysis, random sampling for detectors leads to better conditioning of the forward operator, i.e., increased robustness to noise, but only provided it is done dynamically, i.e., changing the used detectors from one view angle (source location) to the other.

### B. Singular Vectors

An important requirement for a system measuring sparse objects is to be able to distinguish between two different sparse objects via their corresponding measurements. Assume the measurements can be described as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{5}$$

where $\mathbf{\Phi}$ is the discrete linear operator on the domain in which the object representation $\mathbf{x}$ is sparse (in our example it is the wavelet domain). In order to illustrate the concept in the simplest way, we shall assume a strictly sparse image, i.e., the number of nonzero elements denoted $\|\mathbf{x}\|_0$ is small. In order for two different sparse images $\mathbf{x}_1, \mathbf{x}_2$ with $\|\mathbf{x}_{1,2}\|_0 \leq S$ to yield different measurements, one must require that the null space of $\mathbf{\Phi}$ does not contain objects with less than $2S$ elements, since

$$\mathbf{\Phi}\mathbf{x}_1 \neq \mathbf{\Phi}\mathbf{x}_2 \Rightarrow \mathbf{\Phi}(\mathbf{x}_1 - \mathbf{x}_2) \neq 0 \tag{6}$$

and since $\|\mathbf{x}_1 - \mathbf{x}_2\|_0 \leq 2S$. This property can be quantified using the coherence metric discussed in Section 5.A, but it can also be illustrated here using the singular components. We shall look at the features that cannot be recovered reliably, which represent the effective null space of the system when noise is present. We would like those components in the null space to be as incoherent (noise-like) as possible in the sparsifying domain, i.e., to be as noncompressible as possible, so that the condition in Eq. (6) will hold.

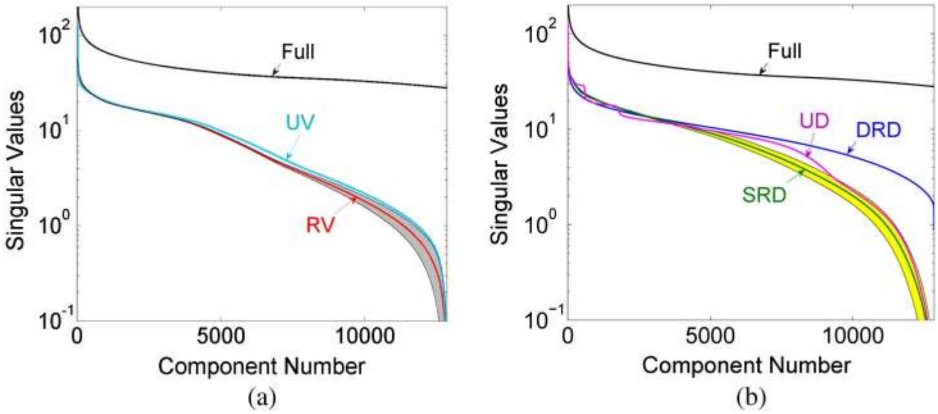


Fig. 5. Singular values for ×16 undersampling using the different sampling strategies described in Section 2.B. Here we compare *uniform* sampling to *random* sampling in the cases of (a) view subsampling and (b) detector subsampling. UV, uniform view; RV, random view; UD, uniform detector; DRD, dynamic random detector; SRD, static random detector. For strategies involving random sampling, the mean and standard deviation (std) for 20 different random selections are indicated by solid lines and shaded strips, respectively. For clarity the boundaries of the shaded strips correspond to 1 std above and below the mean (gray strip for RV and yellow strip for SRD). Note that the means for DRD and SDRD are indistinguishable from each other in this scale and therefore the latter is omitted here. Also, the std for DRD and SDRD are indistinguishable from the mean of DRD and are therefore omitted here. The forward model is described in Fig. 3, and a $128 \times 128$ image resolution is used.

Figures 6–8 show some representative examples of the right singular vectors of **H** for the lowest singular values corresponding to the UD, DRD, and UV sampling strategies, respectively, and presented as images. The right singular vectors for the forward operator on the Haar wavelet domain are obtained by applying the wavelet transform to the singular vector. The wavelet decompositions are shown in Figs. 9–11 for UD, DRD, and UV, respectively. The right singular vectors for UD in Figs. 6(a)–6(d) are clearly compressible using the Haar wavelet basis, as seen in Fig. 9. In view of the discussion above and these results, UD sampling is clearly not suited for compressible objects. The right singular vectors for DRD in Fig. 7 are significantly less compressible than the ones for UD, as seen by comparing Fig. 10 with Fig. 9. In fact, they consist of unstructured noise-like patterns, which is the desired property for CS. This illustrates once again the importance of *random (unstructured)* sampling of detectors. The singular vectors for DRD sampling in Figs. 7(a)–7(d) are also less compressible than the ones for UV sampling in Figs. 8(a)–8(d), as seen by comparing Fig. 10 with Fig. 11, which suggests that DRD is the preferred strategy.

For a complete dataset, the singular vectors resemble those for the parallel beam geometry that are widely known in the literature, since in principle, this is just a reordering of the rays. In that case, the singular vectors are spread across the entire image domain, and vectors corresponding to lower singular values have higher spatial oscillations (higher frequency content) than the vectors for higher singular values. However, it should be noted that the singular vectors for subsampled data can *differ considerably* from the ones corresponding to the complete dataset. For example, UD subsampling results in a more sparse covering of the central region of the image, and therefore this region is most sensitive to noise. It is therefore not surprising that the last singular vectors are confined to this region (see Fig. 6), as opposed to the singular vectors for the complete dataset, which are spread across the entire image domain. Another key difference is that all singular vectors for subsampled data contain both high and low local frequencies.

## 5. COMPRESSIVE SENSING THEORY

Assume the signal can be represented in a given orthonormal basis in $\mathbb{R}^N$:

$$\mathbf{f} = \mathbf{\Psi x}, \tag{7}$$

where the columns of $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ represent the basis functions and $\mathbf{x} \in \mathbb{R}^N$ are the basis weights, assumed to be sparse (compressible). The measurement model is usually assumed to be linear:

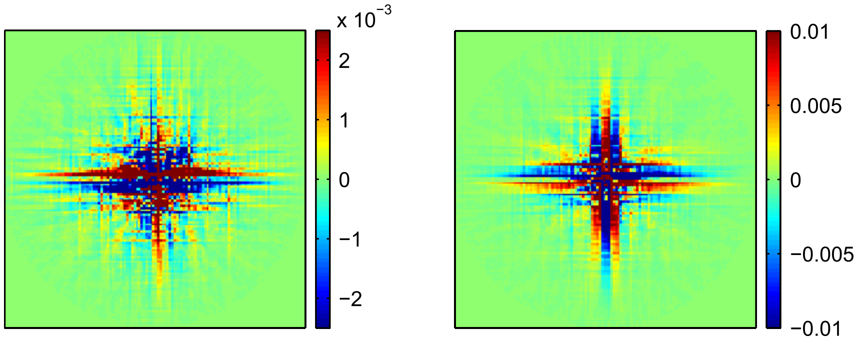
$$\mathbf{y} = \mathbf{Hf} + \epsilon = \mathbf{H\Psi x} + \epsilon, \tag{8}$$

where $\epsilon$ is an unknown noise vector (deterministic or random) and in the second equality we substituted Eq. (7). A key result in CS is stated in terms of the restricted isometry property (RIP). The matrix $\mathbf{\Phi} = \mathbf{H\Psi}$ is said to satisfy the RIP of order $S$ if there exists a constant $\delta_s \in (0, 1)$ such that

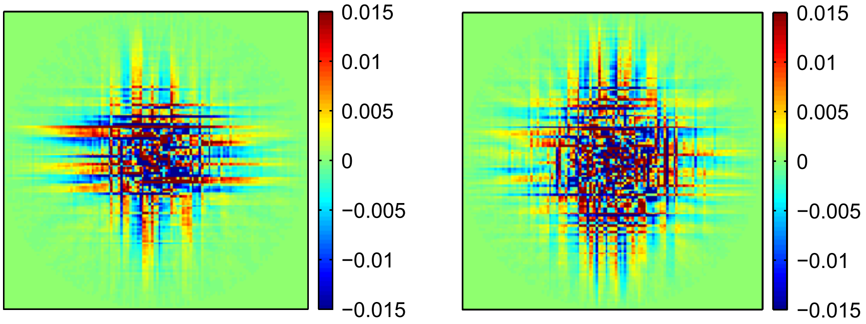
$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \le \|\mathbf{\Phi x}\|_2^2 \le (1 + \delta_s)\|\mathbf{x}\|_2^2, \tag{9}$$

for all vectors $\mathbf{x}$ having at most $S$ nonzero entries ($S$ sparse). Known examples of matrices that are guaranteed to have this property with high probability are random matrices with entries drawn independently from Gaussian or Bernoulli distributions. However, in practical physical problems measurements have predefined structure, e.g., in tomography each measurement corresponds to a line integral, and the random matrices described above are not very useful. Also, the task of verifying that a given (structured) matrix satisfies the RIP has exponential complexity, since one needs to verify that each combination of up to $S$ columns of $\mathbf{\Phi}$ satisfies the inequality in Eq. (9) (assuming we know $S$).

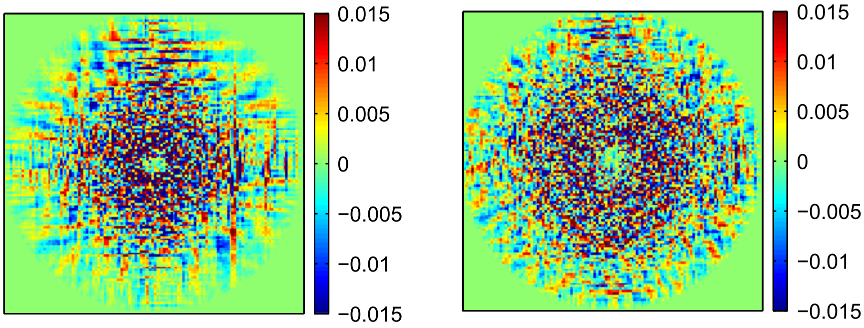
To address this issue, alternative approaches have been proposed in [34–36] that are based on the notion of the *coherence* between the measurements and the representation basis. A measurement ensemble given by **H** in Eq. (8) is said to be "incoherent" with respect to the representation $\mathbf{\Psi}$ in Eq. (7) if the rows of **H** have a small inner product with the columns of $\mathbf{\Psi}$. The main result in [34] assures that incoherent



(a) UD - last (b) UD - 10th before last



(c) UD - 50th before last (d) UD - 100th before last



(e) UD - 500th before last (f) UD - 1000th before last

Fig. 6. Right singular vectors with lowest singular values in the case of uniform detector (UD) subsampling by a factor of 16. Note that a circular domain was used.

measurements subsampled at random can be inverted with overwhelming probability. This definition of coherence is restricted to the case where both **H** and **Ψ** are unitary matrices. Unfortunately, as discussed in the introduction, our tomographic measurements do not comply with this requirement, so this definition of incoherence cannot be used here.

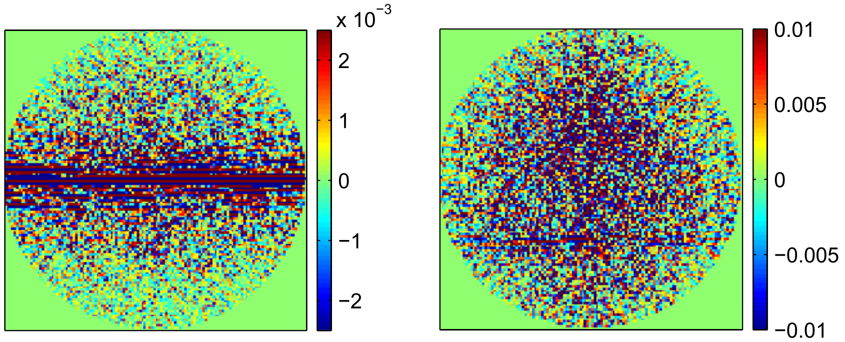
Another line of work [41] uses a different definition for coherence, which we shall employ in our study, since it allows arbitrary measurements and also arbitrary dictionary representations of the object (not necessarily an orthonormal basis). The measurements are said to be "incoherent" if any two different basis functions lead to different measurements. In order to calculate this coherence, we define the transform point spread function as in [42,43], given by

$$\mathrm{TPSF}(i,j) = (\mathbf{\Psi}^T \mathbf{H}^T \mathbf{H} \mathbf{\Psi})_{i,j}, \tag{10}$$

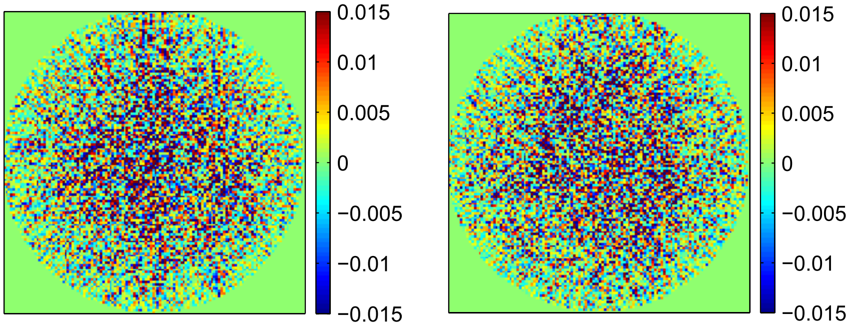
where $()_{i,j}$ denotes the element at the $i$th row and $j$th column and **H** was defined in Eq. (4). The coherence is defined as

$$\mu \triangleq \max_{i\neq j} |\mathrm{TPSF}(i,j)| / \sqrt{\mathrm{TPSF}(i,i)\mathrm{TPSF}(j,j)}. \tag{11}$$

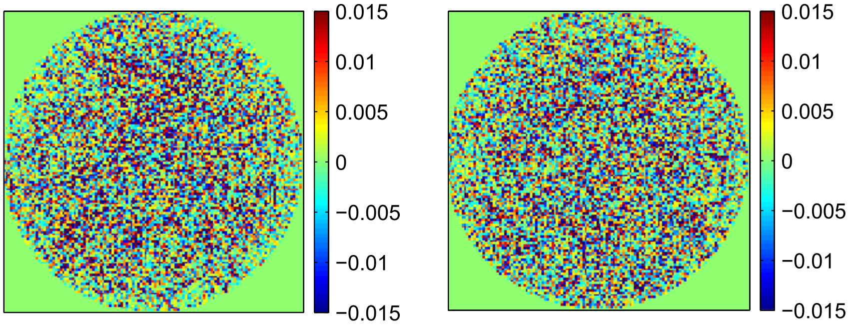
According to this definition, coherence is the maximal inner product between two measurements generated by different basis images. The coherence is also related to the RIP as described in [44]. As mentioned in the introduction, the system must maximally separate basis images in the measurement range (sinogram) space in order to distinguish between different basis images. This separation is quantified here by the incoherence, or lack of correlation between measurements corresponding to different basis images, so a low coherence is desirable. The connection between coherence and the null space property mentioned in Section 4.B is discussed in [41], and here we shall briefly mention it to point out potential limitations of this metric. The null space vectors $\{\boldsymbol{\gamma} : \mathbf{\Phi}\boldsymbol{\gamma} = 0\}$; have nonzero entries at locations corresponding to linearly dependent columns in **Φ** (by definition of linear dependence). Accordingly we have $\|\boldsymbol{\gamma}\|_0 \geq \mathrm{Spark}\{\mathbf{\Phi}\}$;, where the spark is the minimal number of linearly dependent columns. The



(a) DRD - last (b) DRD - 10th before last



(c) DRD - 50th before last (d) DRD - 100th before last



(e) DRD - 500th before last (f) DRD - 1000th before last

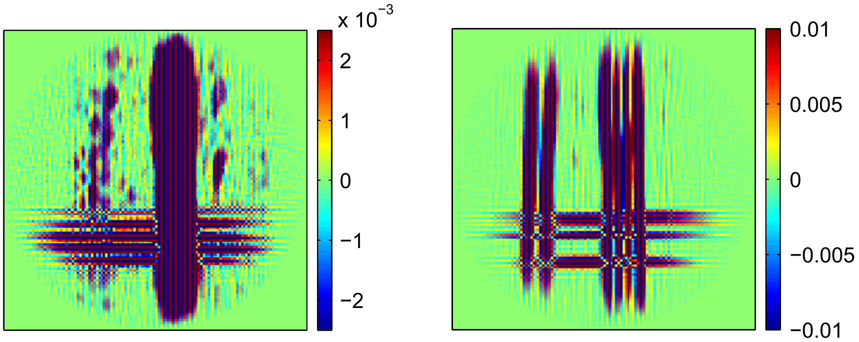
Fig. 7. Right singular vectors with lowest singular values in the case of dynamic random detector (DRD) subsampling by a factor of 16. Note that a circular domain was used.

null space property in Eq. (6) is satisfied if $\mathrm{Spark}\{\boldsymbol{\Phi}\} \geq 2S$. Computing the spark is not feasible, since it requires considering all possible combinations of columns. The bound $\mathrm{Spark}\{\boldsymbol{\Phi}\} \geq 1 + 1/\mu$ can be used instead, where $\mu$ is the coherence defined in Eq. (11). This illustrates the pessimistic nature of the incoherence metric, since it forms a lower bound on the spark.

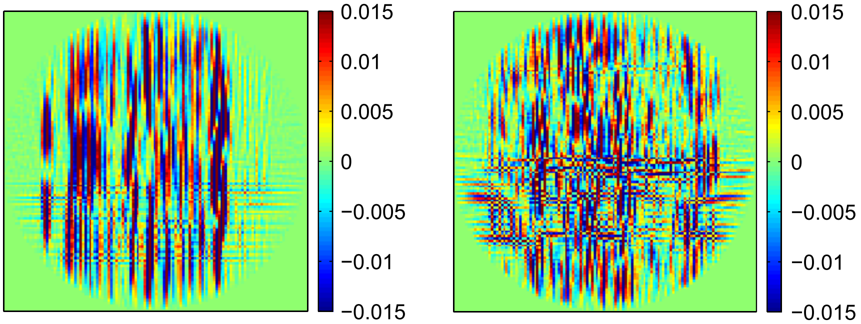
### A. Results

The coherence for different subsampling strategies with respect to the Haar wavelet basis has been calculated according to the definition in Eq. (11). The columns of the matrix $\boldsymbol{\Psi}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\Psi}$ have been normalized so that the sum of the squared entries in each column equals 1. The off-diagonal element of the resulting matrix with the maximal absolute value is equal to the coherence. Here we use a circular image domain to be consistent with the previous analysis. We remove any wavelet basis images that are outside this domain and also disregard any pairs of wavelets that are identical inside the image domain. The random selection of views/detectors for each strategy has been repeated 50 times, and the mean and standard deviation of the coherence for this ensemble are computed.

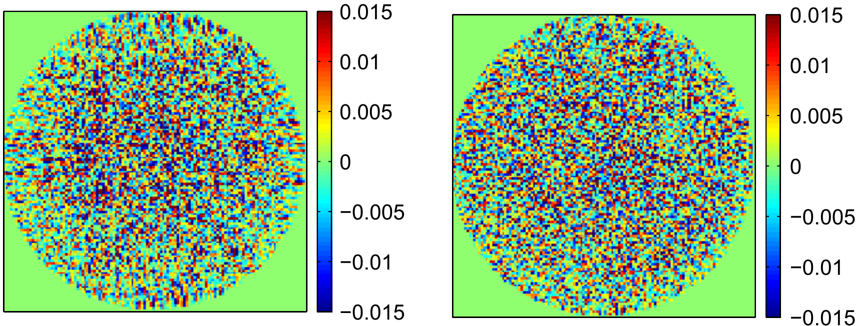
The results for different random subsampling strategies are shown in Table 2. It can be seen that the mean performance of DRD sampling is comparable to RV sampling up to ×16 subsampling but with 2 times lower standard deviation, so DRD is less sensitive than RV to the particular choice of measurements. SRD sampling is clearly a bad choice, especially for high subsampling rates, which agrees with the SVD analysis in Section 4. A simple way to visualize the different approaches is to picture the measurement (sinogram) space as a 2D matrix where rows correspond to detectors and columns correspond to views. The SRD corresponds to randomly removing rows, while the RV approach corresponds to randomly removing columns. The DRD approach, on the other



(a) UV - last (b) UV - 10th before last



(c) UV - 50th before last (d) UV - 100th before last

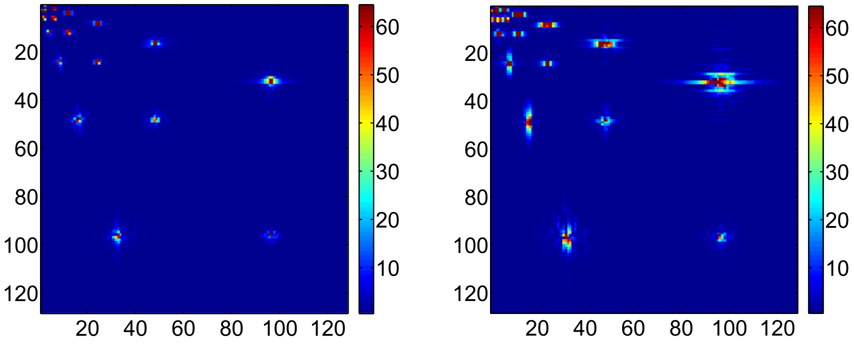


(e) UV - 500th before last (f) UV - 1000th before last

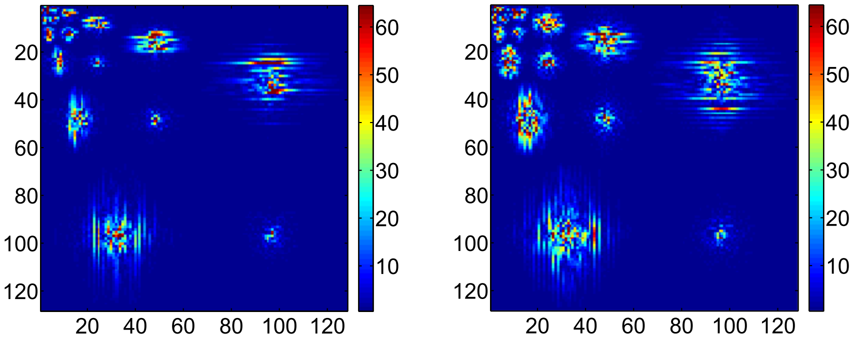
Fig. 8. Right singular vectors with lowest singular values in the case of uniform view-angle (UV) subsampling by a factor of 16. Note that a circular domain was used.

hand, corresponds to randomly removing individual pixels, so it is more spread out in sinogram space. One might naively expect that SRD and RV should be comparable, but the results clearly show the advantage of RV. The above intuition relies on notions of object smoothness and implies some form of interpolation. However, we are considering subsampling rates for which standard interpolation techniques fail, and we assume sparseness in the wavelet domain instead of smoothness, which implies piecewise constant objects when employing Haar wavelets or piecewise smooth objects when employing Daubechies wavelets. According to the same intuition, one might also expect that DRD would perform better than RV, and in this case, this is indeed seen by the differences in standard deviation in Table 2.

The wavelet pairs or basis images that lead to the highest coherence (correlation of measurements) vary according to a few typical cases, depending on the strategy used and the specific realization of the selected measurements when random subsampling is employed. In one case that occurs frequently, one basis image is the scaling function (constant across all image space) and the second basis image is one of the second-level wavelets occupying a $64 \times 64$ quadrant of the image. We shall refer to this case as type I. In the second case, the basis images are neighboring finest-scale wavelets around the center of the image occupying a $2 \times 2$ square, or alternatively, parent–child wavelet pairs with the child in the finest scale. We shall refer to this case as type II. For DRD and SRD most occurrences (random realizations) are of type II, but as the subsampling factor increases considerably (say, 16 and up), type I becomes more frequent and eventually dominates. In contrast, for RV subsampling, type I is always dominant, with a few more occurrences of type II when going to higher subsampling rates. The correlation between wavelet pairs of type II is expected based on classical sampling theory: if one chooses to sample below the minimal rate for a given minimal resolution, details of the smallest size might not be resolved in measurement space. The correlation between wavelet pairs of type I, which involve coarse-level wavelets, seems to



(a) UD - last (b) UD - 10th before last



(c) UD - 50th before last (d) UD - 100th before last

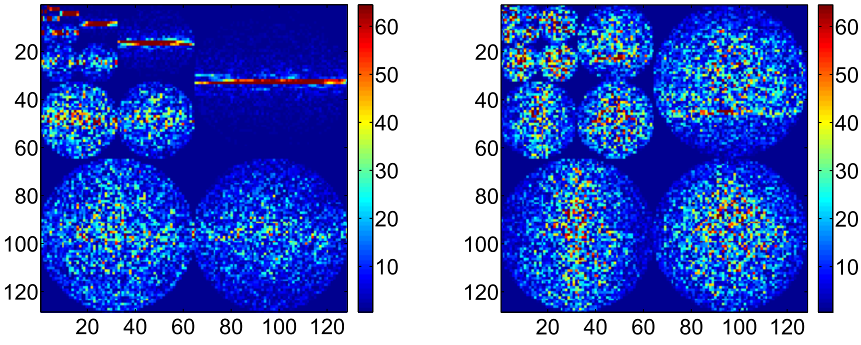
Fig. 9. Multiresolution image decomposition using Haar wavelets for the UD right singular vectors shown in Fig. 6. Each block contains the magnitudes of wavelet coefficients corresponding to a different spatial scale, with the horizontal scale decreasing from left to right and the vertical scale decreasing from top to bottom, e.g., the block near the bottom-right corner contains the wavelet coefficients for the finest scale in both directions (the support of these wavelets is $2 \times 2$ pixels). Within a block, each entry corresponds to a different spatial translation of the wavelet. The magnitude of the wavelet coefficients is presented in logarithmic scale. These plots were generated using the MATLAB command "plotwavelet2" and with a rescaling value of 100 [40].

contradict those arguments. This is due to the fact that wavelets have positive and negative values and various cancellation effects occur.

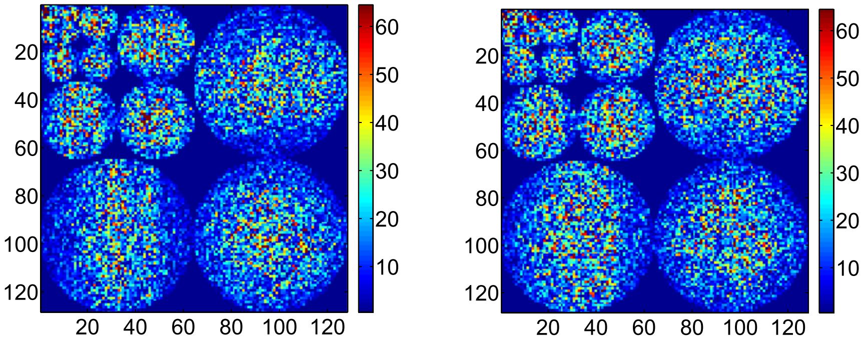
Table 3 shows the coherence for different uniform subsampling strategies. It is seen that UV sampling has lower coherence than RV sampling, and also lower coherence than DRD. The most correlated wavelet pair for UV sampling is of type I defined above. The next most correlated pair is a"parent–child" pair in the coarsest scale ($128 \times 128$ and $64 \times 64$ basis images, respectively). This situation does not change when we go to high subsampling factors, unlike in RD sampling (DRD, SRD), where essentially most correlated wavelets were of the finest scale (type II). It is interesting to note in Table 3 that the coherence for UV does not change much as the object is increasingly undersampled and even slightly decreases. This is consistent with the fact that the most correlated wavelets are always at the coarsest level (type I) with large spatial support. Clearly this demonstrates some sort of pitfall in the coherence metric, and one should be careful in interpreting these results. For a UD with low undersampling rates we obtain a similar result; however, for higher undersampling rates the most correlated wavelets become type II and coherence increases rapidly to very high values, which supports the conclusions made in Section 4.

Generally speaking, as incoherence increases (coherence decreases) it becomes possible to distinguish between objects that are less sparse, i.e., that have more nonzero coefficients in their representation. Wavelet transforms of natural images will typically have most of their zeros or small-valued coefficients in the fine-scale levels (high frequencies). Therefore, from a CS perspective, it is better to have higher coherence for fine-scale wavelets than to have high coherence for coarse-scale wavelets, since the coarse-scale content will be less sparse and its reconstruction will be more prone to fail for higher coherence. This fact is not reflected in the coherence metric used here. Instead of considering typical situations, it rather reflects a "worst-case" scenario (for more details see, for example, [41,45]). Therefore, it is actually encouraging to see that in DRD the fine-scale wavelets are the ones responsible for the highest coherence, as opposed to UV, where they are the coarsest-scale wavelets. Indeed, in practice, when comparing DRD with UV, we have seen several cases in which there is a clear advantage to DRD, as described in the sequel.

In summary, coherence, like any other metric, has its own merits and pitfalls, and this is why we have employed several different measures of performance. In the case of random sampling (detector and view), it proved to be useful and consistent with the SVD analysis and the performance seen in practice for several objects (to be reported in Sections 6 and 7). However, for uniform sampling (detector and view), it led to some confusing results such as the slight decrease of coherence as the subsampling rate was increased, and for UV subsampling this was persistent even at high subsampling rates. We also found a disconnect with the actual performance of UV during reconstructions. We report these surprising results in order to promote the development of

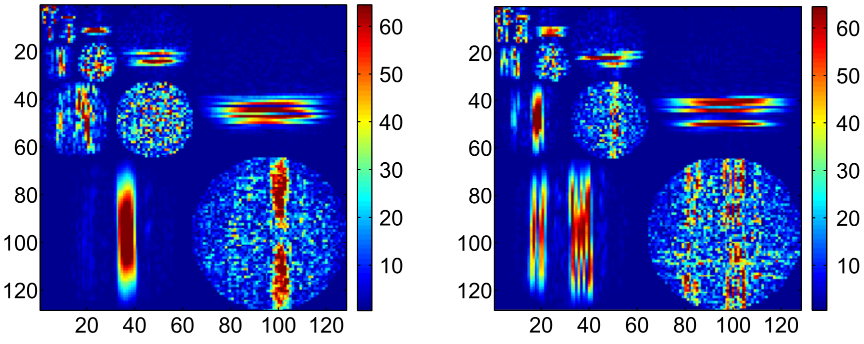


(a) DRD - last (b) DRD - 10th before last

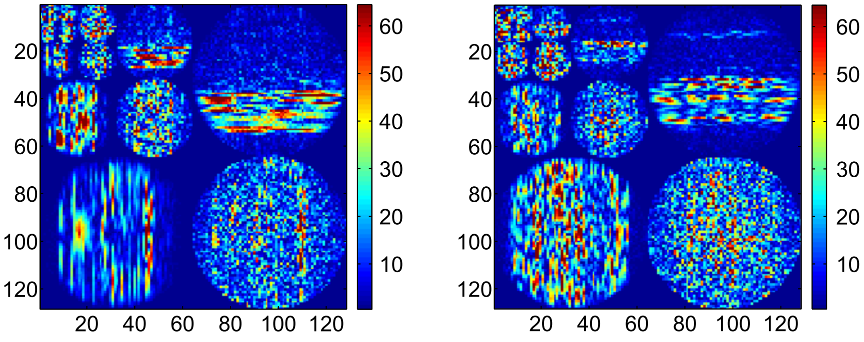


(c) DRD - 50th before last (d) DRD - 100th before last

Fig. 10. Multiresolution image decomposition using Haar wavelets for the DRD right singular vectors shown in Fig. 7. See further details in the caption of Fig. 9.



(a) UV - last (b) UV - 10th before last



(c) UV - 50th before last (d) UV - 100th before last

Fig. 11. Multiresolution image decomposition using Haar wavelets for the UV right singular vectors shown in Fig. 8. See further details in the caption of Fig. 9.

**Table 2. Mutual Coherence of Different Random Sampling Strategies with Respect to the Haar Wavelet Basis and for Different Undersampling Ratios**[a]

| | DRD | | SDRD | | SRD | | RV | |
|---|---|---|---|---|---|---|---|---|
| Downsampling | Mean | Std | Mean | Std | Mean | Std | Mean | Std |
| ×2 | 0.3295 | 0.0011 | 0.3292 | 0.0036 | 0.3457 | 0.0319 | 0.3295 | 0.0015 |
| ×4 | 0.3305 | 0.0017 | 0.3281 | 0.0064 | 0.4649 | 0.1344 | 0.3317 | 0.0030 |
| ×8 | 0.3324 | 0.0024 | 0.3307 | 0.0119 | 0.6361 | 0.1743 | 0.3322 | 0.0050 |
| ×16 | 0.3341 | 0.0043 | 0.3634 | 0.0344 | 0.8195 | 0.1513 | 0.3460 | 0.0234 |
| ×32 | 0.3953 | 0.0215 | 0.4794 | 0.1028 | 0.9236 | 0.1028 | 0.3793 | 0.0564 |

[a] DRD, dynamic random detector; SDRD, semidynamic random detector; SRD, static random detector; RV, random view. For each approach, the random ray selection is repeated 50 times and the mean and standard deviation (std) of the coherence for this ensemble is shown. The coherence for complete measurements (no subsampling) is 0.3285. The image is of resolution 128 × 128.

**Table 3. Mutual Coherence of Different Uniform Sampling Strategies with Respect to the Haar Wavelet Basis and for Different Undersampling Ratios**[a]

| Downsampling | Detectors (UD) | Views (UV) |
|---|---|---|
| ×2 | 0.328444 | 0.328329 |
| ×4 | 0.328436 | 0.328322 |
| ×8 | 0.398522 | 0.328311 |
| ×16 | 0.7118 | 0.3281 |
| ×32 | 0.9212 | 0.3286 |

[a] The coherence for complete measurements (no subsampling) is 0.3285. The image is of resolution 128 × 128. UD, uniform detector sampling; UV, uniform view sampling.

better metrics. Lastly, we note that coherence merely provides a lower bound on the spark [see discussion after Eq. (11)], which is the natural metric but is computationally unfeasible.

## 6. RECONSTRUCTION USING SYNTHETIC CT DATA

### A. Methods

So far we have assumed in Sections 4 and 5 that the line integrals are measured directly. Here, we shall consider a more detailed physical model for the measurements, which is used in practice in x-ray CT and which will also be used in Section 7 for experimental data. The mean number of photons reaching a detector after a ray has passed the object is given by Beer's law [46]:

$$I_i = I_{i0}\exp\left(-\int_{L_i} f(\mathbf{x})\mathrm{d}z\right), \tag{12}$$

where $i$ denotes a source–detector pair connected by ray $L_i$, $I_{i0}$ is the mean number of photons that would reach the $i$th detector without the object, and $f \geq 0$ is the distribution of the linear attenuation constant in $\mathrm{mm}^{-1}$ inside the object under test. Denoting $d_i$ and $d_{i0}$ as the $i$th detector reading with and without the object, respectively, we define the log-transformed and calibrated measurements as

$$y_i = \ln\frac{d_{i0}}{d_i} = \ln d_{i0} - \ln d_i. \tag{13}$$

For a monoenergetic source, the detected signal $d_i$ is generally assumed to be Poisson distributed with the mean and variance equal to $I_i$ given by Eq. (12) [46]. For large values of $I_i$, the log-transformed measurement $y_i$ can be shown to have a Gaussian distribution [47]. After discretizing the line integrals and $f$ as described in Eq. (4), the Gaussian noise model reads

$$\mathbf{y}\sim\mathbf{Hf}+\epsilon, \qquad \epsilon\sim\mathcal{N}(0,\Sigma), \tag{14}$$

where the $i$th element in $\mathbf{y}$ is given by Eq. (13) and $\epsilon$ is a normally distributed noise with a zero mean and a covariance matrix given by

$$\Sigma = \mathrm{diag}[I_0^{-1}\exp(\mathbf{Hf})]. \tag{15}$$

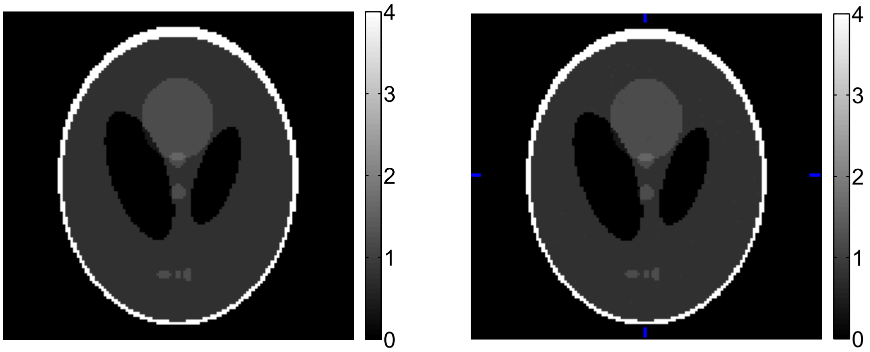
Note that for simplicity we assume $I_{i0} = I_0$ for all $i$. In the following numerical simulations we use this model with $I_0 = 10^5$ to generate synthetic data. In the reconstruction phase, we are given the measurements $\mathbf{y}$, and we estimate the attenuation of x-ray radiation per unit length in the object given by $\mathbf{f}$, which distinguishes between different materials.

The reconstruction algorithm employed here is the relevance vector machine (RVM), which was first introduced in the context of machine learning [48] and later adopted for compressive sensing in [49]. In the RVM, the inversion of compressive measurements is done from a Bayesian perspective [50]. Specifically, we have a prior belief that $\mathbf{f}$ should be sparse in some basis, i.e., in the representation $\mathbf{f} = \mathbf{\Psi x}$, many $\mathbf{x}$ have small values [see Eq. (7)]. This belief is represented by a probability density function for the weights $\mathbf{x}$ that assigns high probability to small weights but still allows a few large weights. Instead of just providing a single estimate for $\mathbf{x}$, a full posterior probability distribution function is computed using Bayes' rule. The posterior distribution represents the current state of knowledge, combining the prior knowledge with the information gained by measurements, and describes how likely different possible estimates for $\mathbf{x}$ are. Typical choices for point estimates are the mean of the posterior distribution for $\mathbf{x}$ and the value of $\mathbf{x}$ for which the posterior is maximal (maximum a *posteriori* = MAP). In addition to a point estimate for $\mathbf{x}$, the Bayesian approach also provides error bars that may be used to give a sense of confidence in the estimated value, and they may also be used to guide the optimal design of additional CS measurements, implemented with the goal of reducing the uncertainty about $\mathbf{x}$ or $\mathbf{f}$ [49]. It is worth emphasizing that the RVM performs a nonlinear reconstruction of the object by imposing a sparse solution and thus utilizes the underlying signal model. The MAP estimates for the object bear similarity to standard nonlinear CS algorithms [49], where in addition to minimizing the distance between the measured and modeled signal, there is also a sparsity-promoting penalty that is determined by the prior probability distribution.

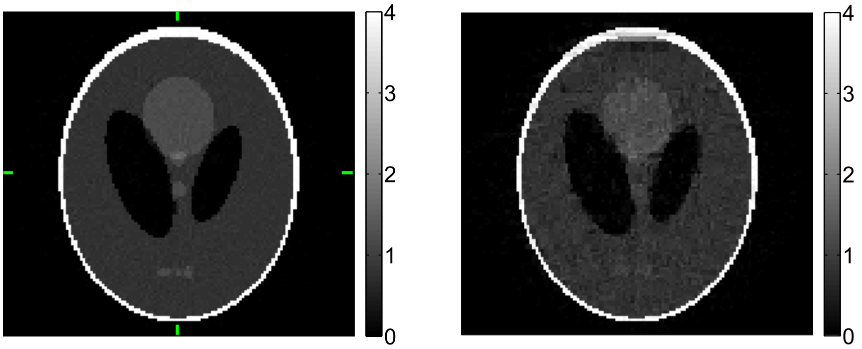
Here, we shall use a simple version of the RVM with some minor modifications to adapt it to x-ray transmission tomography. The noise covariance matrix in [48] is $\Sigma = \sigma^2 \mathbf{I}$, with $\mathbf{I}$ being the identity matrix. In order to convert the model in Eq. (15) to the standard RVM model, one has to perform two modifications. We follow the approach in [17] and estimate the object-dependent covariance matrix $\Sigma$ in Eq. (15) by replacing $\mathbf{Hf}$ with the measurements $\mathbf{y}$. In all our experiments we found this simple estimate to work very well compared to using the exact nonlinear model with $\Sigma$ given by Eq. (15). Next, we apply a noise-whitening procedure by replacing $\mathbf{y}$ and $\mathbf{H}$ with $\tilde{\mathbf{y}} = \Sigma^{-1/2}\mathbf{y}$ and $\tilde{\mathbf{H}} = \Sigma^{-1/2}\mathbf{H}$, respectively, reducing the covariance matrix for $\tilde{\mathbf{y}}$ to the identity matrix. A full treatment of the RVM for x-ray transmission tomography will be presented in a subsequent paper, which will include the option of using the original noise model in Eq. (15) without estimating the noise covariance matrix first.

### B. Results

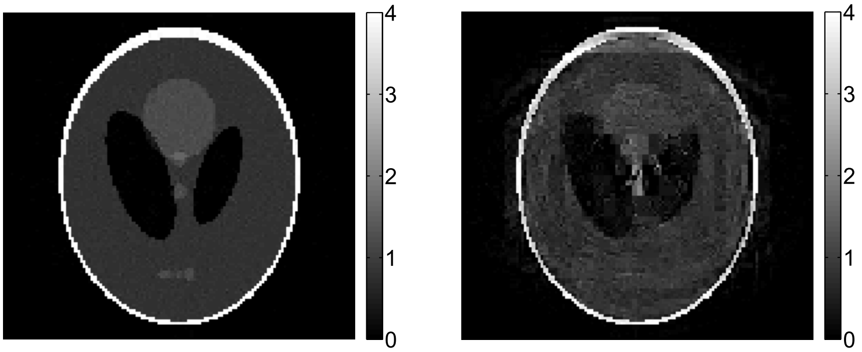
Figure 12 shows reconstructions of the relative attenuation maps for a 128 × 128 Shepp–Logan phantom [Fig. 12(a)] using the RVM algorithm and for different random sampling strategies. To determine convergence during RVM iterations, we verified that successive iterations satisfy the condition $\|\mathbf{f}^{(t+1)} - \mathbf{f}^{(t)}\|_2 / \|\mathbf{f}^{(t)}\|_2 \leq 0.01$, where the superscript denotes the iteration number. In this case four iterations satisfied the above criterion for all strategies. Relative attenuation



(a) Truth (b) Full data



(c) DRD (d) RV



(e) SDRD (f) SRD

Fig. 12. (a) Truth, (b)–(d) reconstructed images using RVM with four iterations. Values represent the attenuation per unit length divided by the attenuation per unit length of water. (b) Using all measurements; (c)–(f) measurements downsampled by a factor of 32 using the following sampling strategies: (c) dynamic random detector (DRD), (d) random view (RV), (e) semidynamic random detector (SDRD), (f) static random detector (SRD). Data have been generated according to Eqs. (14) and (15) with $I_0 = 10^5$. Colored hatch marks in (b) and (c) indicate the cross sections shown in Figs. 14(a) and 14(b), respectively (with matching colors).

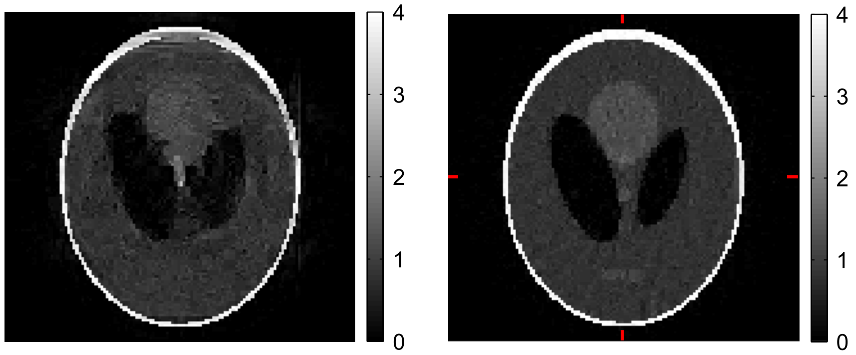
**Table 4. Root Mean-Squared Error (RMSE) for RVM Reconstructions Averaged over 50 Different Random Selections from the Complete Dataset and for Different Sampling Strategies**[a]

| Downsampling | DRD | SDRD | SRD | RV |
|---|---|---|---|---|
| ×4 | $8.4 \times 10^{-3}$ | $8.4 \times 10^{-3}$ | $9.6 \times 10^{-3}$ | $9 \times 10^{-3}$ |
| ×8 | $1.87 \times 10^{-2}$ | $1.87 \times 10^{-2}$ | $3.77 \times 10^{-2}$ | $2.56 \times 10^{-2}$ |
| ×16 | $1.89 \times 10^{-2}$ | $1.9 \times 10^{-2}$ | $8.3 \times 10^{-2}$ | $4.32 \times 10^{-2}$ |
| ×32 | $3.63 \times 10^{-2}$ | $3.69 \times 10^{-2}$ | $2.918 \times 10^{-1}$ | $1.243 \times 10^{-1}$ |

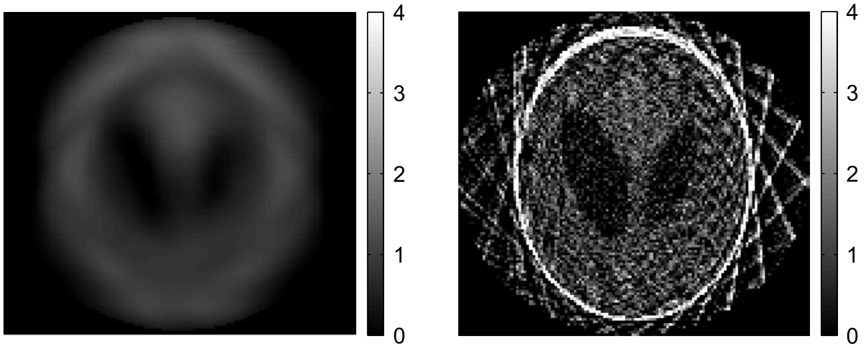
[a] The RMSE has been normalized by the $\ell_2$ norm of the truth. DRD, dynamic random detector; SDRD, semidynamic random detector; SRD, static random detector; RV, random view. The source intensity parameter is $I_0 = 10^5$. For each set of measurements only four iterations in RVM were used. Since noise does not vary considerably between different measurements, a single dataset has been generated according to the model in Eqs. (14) and (15), and the same dataset has been used for all strategies. We use the $128 \times 128$ Shepp–Logan phantom of Fig. 12(a). The $\ell_2$ error when using the complete set of measurements (no subsampling) is $6.5 \times 10^{-3}$.

values limited to the window [0.5, 2]; are presented in Appendix B, where it is easier to see the differences between the different approaches. The measurements are synthesized from the model in Eqs. (14) and (15) with $I_0 = 10^5$. The reconstruction using the complete synthesized measurements is shown in Fig. 12(b), which is almost indistinguishable from the truth in Fig. 12(a). Figures 12(c)–12(f) show reconstructions when using different random sampling strategies with downsampling by a factor of 32. It is clearly seen (see also Appendix B) that SRD sampling [Fig. 12(f)] and RV sampling [Fig. 12(d)] have the worst results, with an advantage to RV. DRD sampling [Fig. 12(c)] yields the best result among all explored strategies. It should also be noted that the image quality with DRD [Fig. 12(c)] is not degraded much relative to the reconstructed image using the complete set of measurements [Fig. 12(b)], and that almost the same quality can be achieved with SDRD sampling [Fig. 12(e)], using only 10 masks in this case. Table 4 shows the root mean-squared error (RMSE) averaged over 50 different random measurement selections and verifies the previous statements regardless of the particular random realization.

Figure 13 shows reconstructions of relative attenuation maps when uniform sampling strategies are used. Relative attenuation values limited to the window [0.5, 2]; are presented in Appendix B, where it is easier to see the differences between different sampling strategies. It can be seen that UV sampling is preferred over RV sampling, which is consistent with the observations made in Sections 4 and 5. When comparing the best view-angle sampling strategy, UV [Fig. 13(b)], to the best detector sampling strategy, DRD [Fig. 12(c)], it is evident that the latter yields a better image in this case; see also Appendix B. Figure 14 presents a comparison of these two approaches for 1D cuts of the images. Figures 13(c) and 13(d) show the reconstructions

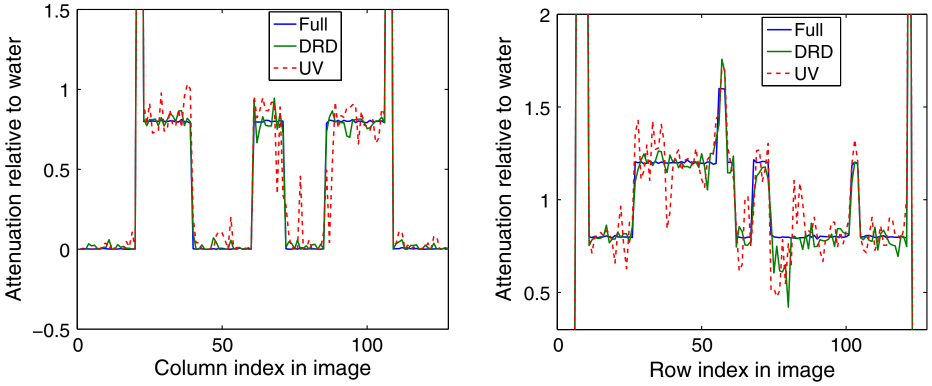


(a) RVM - uniform detector (UD) (b) RVM - uniform view (UV)



(c) FBP - uniform detectors (UD) (d) FBP - uniform views (UV)

Fig. 13. Reconstructed images for uniform undersampling by a factor of 32. (a), (b) RVM after four iterations; (c), (d) filtered backprojection (FBP). (a), (c) uniform detector (UD) sampling; (b), (d) uniform view (UV) sampling. The data have been generated according to the model in Eqs. (14) and (15) with $I_0 = 10^5$. Cross sections of the image in (b) along the horizontal and vertical lines are shown in Figs. 14(a) and 14(b), respectively (with matching colors).



(a) Horizontal Cut (b) Vertical Cut

Fig. 14. 1D Cuts of the images in Fig. 12(b) (blue) for complete data, Fig. 12(c) (green) for dynamic random detector sampling (DRD), and Fig. 13(b) (red) for uniform view (UV) sampling. These cuts are indicated in the images by matching colored hatch marks. (a) Horizontal cut at the middle of the image and (b) vertical cut at the middle of the image.

for UD and UV subsampling, respectively, using the filtered backprojection algorithm [1]. This is merely done to illustrate just how much the object is undersampled, and we do not claim our results are the best possible by using filtered backprojection (e.g., one can attempt to interpolate the missing rays).

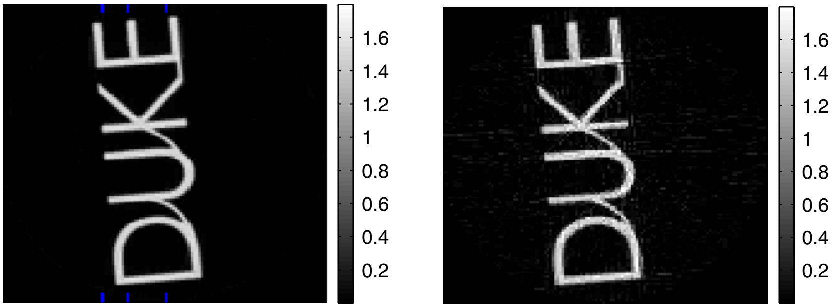
## 7. EXPERIMENTAL CT DATA

### A. Physical Phantom Dataset

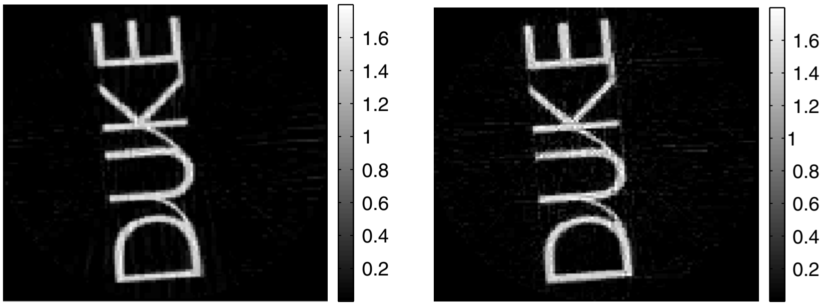
Here we present reconstructions for a dataset acquired with an experimental CT system at Duke University. Details of the experimental system are provided in Appendix C. In these experiments we image the "DUKE" geometry summarized in that appendix. In this case we have a good estimate of the ground truth to help us evaluate the quality of the reconstruction (as opposed to Section 7.B, where a human patient is being scanned and the truth is unknown). As mentioned already, the system geometry is identical to the one used in the simulations in Section 6 and enables us to fairly compare simulation and practice.

We have already established in Section 4–6 that the two most competitive strategies are UV sampling and DRD sampling, and we shall focus on them here. Continuing the approach used in Section 6, we would like to make sure that we have a fair starting point for comparison according to classical sampling theory. However, when comparing the system specifications (see Appendix C) with the required resolution detailed in Section 3, we find that detector resolution is oversampled by a factor of 1.5 and the view angle is undersampled by a factor of 2. One way to make a comparison between detector and view-angle sampling is to uniformly undersample detectors by a factor of 3 so that both views and detectors are ×2 undersampled relative to the minimal theoretical sampling rates, and any additional subsampling would use this as a starting point. In this way, the total number of measurements (projections) would be the same after the additional subsampling with respect to either view angles or detectors. However, as demonstrated previously, *uniform* detector undersampling was one of the worst sampling strategies, whereas for view angles, uniform undersampling was shown to be the best choice. Therefore, comparing detector and view-angle subsampling under these conditions is not fair. The reconstructed images for UV and DRD are shown in Fig. 15 for ×8 undersampling, which was the highest rate without introducing significant artifacts. Despite the unfair comparison, it can be seen that DRD and UV are comparable. In addition, it is seen that the use of just 10 masks (SDRD) is comparable to using different masks for all views (DRD). In all cases, undersampling was done by disregarding available measurements before any processing was done. A study concerning the undersampling of detectors using an actual mask or a beam stop array will be reported in a future publication. The reconstruction was done using the same method discussed in Section 6. The criterion for determining convergence was the same as in Section 6.B, for which eight iterations were sufficient for all strategies.

We perform another comparison where we use the full available data as a starting point (with detectors oversampled by a factor of 1.5 and views undersampled by a factor of 2), and we make sure that both view angles and detectors are approximately undersampled by the same factor relative to classical theory. More specifically, if we downsample the data by, say, a factor of $S/2$ for UV, then for DRD we downsample by a factor of $3S/2$. For DRD, several detector intervals would be greater than $1/S$ th of the minimal interval required by classical theory (with high probability), while for UV the view intervals would be exactly $1/S$th of the minimal view interval, which gives UV an advantage over DRD. In addition, UV sampling uses the full set of detectors, which are oversampled by a factor of 1.5, while DRD uses a set of views that is undersampled by 2. In this comparison we shall show that the proposed method (DRD) outperforms UV, despite these disadvantages. In this kind of comparison the total number of measurements (projections) will be smaller for detector sampling, as it needs to be subsampled by a higher factor. Figures 16(a) and 16(b) compare the reconstructions for ×16 undersampling, where the result for DRD sampling has fewer artifacts than UV sampling. In short, DRD yields a higher quality image with three times fewer measurements. A major improvement is obtained by combining these two strategies. Figure 16(c) shows the results for subsampling by a factor of 4 with respect to both detectors and view angles. The advantage over UV in terms of image quality is clearly seen in Fig. 17, which shows 1D cuts of the reconstructed image. Because of the way we perform the detector subsampling, the number of measurements is still three times less

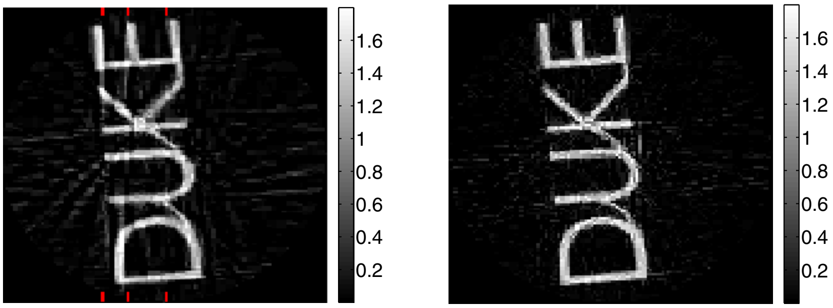


(a) All Data (b) DRD

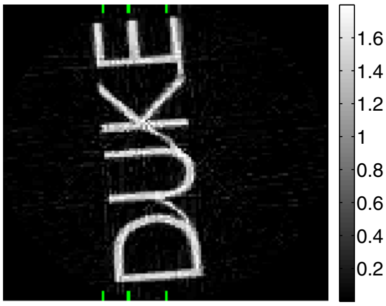


(c) UV (d) SDRD

Fig. 15. Reconstructed images using RVM for experimental data. Values represent the attenuation per unit length divided by the attenuation per unit length of water and are limited to the window $[0, 1.8]$. Undersampling ratios are relative to the minimal sampling rate dictated by classical theory. (a) The object is ×2 uniformly undersampled with respect to both detectors and views; (b)–(d) the available data are downsampled by an additional factor of 4 with respect to detectors or views (×8 undersampled for either detectors or views and ×2 undersampled for either views or detectors, respectively). (b) Dynamic random detector sampling (DRD), (c) uniform view (UV) sampling; (d) semidynamic random detector (SDRD) sampling. Only eight iterations of RVM have been used to generate all images. All strategies have the same number of measurements (line integrals). Colored hatch marks in (a) indicate the cross sections shown in Fig. 17.

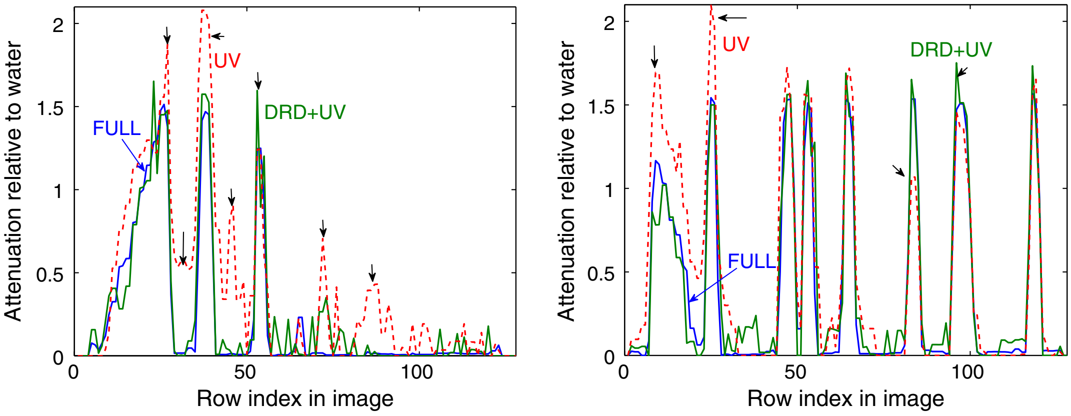


(a) UV ×16 (b) DRD ×16

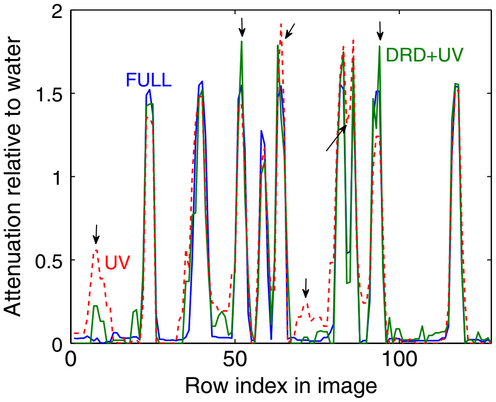


(c) DRD ×4 and UV ×4

Fig. 16. Reconstructed images using RVM for experimental data. Values represent the attenuation per unit length divided by the attenuation per unit length of water. Undersampling ratios are relative to the minimal sampling rates dictated by classical theory. (a) View angles are undersampled by a factor of 16, and all available detectors are used, which are oversampled by 1.5; (b) detectors are ×16 undersampled using the DRD strategy, and views are ×2 undersampled due to limited data; (c) in the combined approach detectors are ×4 undersampled using the DRD strategy, and view angles are ×4 undersampled uniformly (object is undersampled by a total factor of 16). Unlike in Fig. 15, detector selection is done from all available detectors, which are ×1.5 oversampled, and to obtain the same undersampling relative to classical theory, the detectors were downsampled ×3 more than views, so in (b) and (c) there are ×3 fewer measurements (line integrals) than in (a). All other specifications are the same as in Fig. 15.



(a) Left Column (b) Middle Column



(c) Right Column

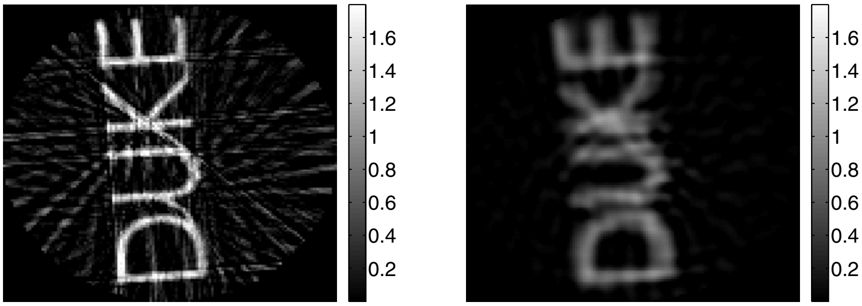
Fig. 17. Linear attenuation coefficients (relative to water) along the columns indicated by matching colored hatch marks in the images in Figs. 15 and 16. Blue represents Fig. 15(a) for full data, red represents Fig. 16(a) for UV sampling, and green represents Fig. 16(c) for combined DRD-UV. Arrows point to regions where significant differences between methods exist.

than for UV. To demonstrate just how much the object is undersampled, we present in Fig. 18 the reconstructions using the filtered backprojection algorithm for UD and UV subsampling.

## B. Clinical Dataset

With approval from the Washington University Institutional Review Board (IRB), patient data were acquired on a Siemens Sensation 16 scanner at St. Louis Children's Hospital using a spiral scanning mode. Further details are available in Appendix D.

The computations run on this clinical dataset were consistent with those reported here for the simulated dataset and the Duke dataset. In particular, SRD and UD sampling, i.e., the use of the same detectors for every view, produced images with noticeable spiral artifacts. The best performance for detector sampling was obtained using DRD sampling, where the detectors used are changed for each view. The use of UV sampling led to better performance than RV sampling. As mentioned before, undersampling was done by disregarding available measurements before any processing was done, and no masks were used during the actual measurements.



(a) FBP - Uniform view (UV) (b) FBP - Uniform detector (UD)

Fig. 18. Reconstructed images for experimental data using the filtered backprojection (FBP) algorithm. (a) View angles are undersampled uniformly by a factor of 16, and detectors are undersampled uniformly by a factor of 2 (relative to classical sampling theory) and (b) detectors are undersampled uniformly by a factor of 16, and view angles are undersampled uniformly by a factor of 2 (relative to classical sampling theory).

Shown below in Figs. 19–21 are examples of using DRD and UV sampling on the clinical data. The complete results will be presented elsewhere. Each of these three figures has three subfigures showing the axial, coronal, and sagittal cuts through the center of the volume. Figure 19 shows the reconstructed images obtained by using the entire dataset without subsampling. The reconstruction was done using our regularized alternating minimization (AM) algorithm, which computes the MAP estimate for the attenuation function using the Beer's law model [51]. For the regularization

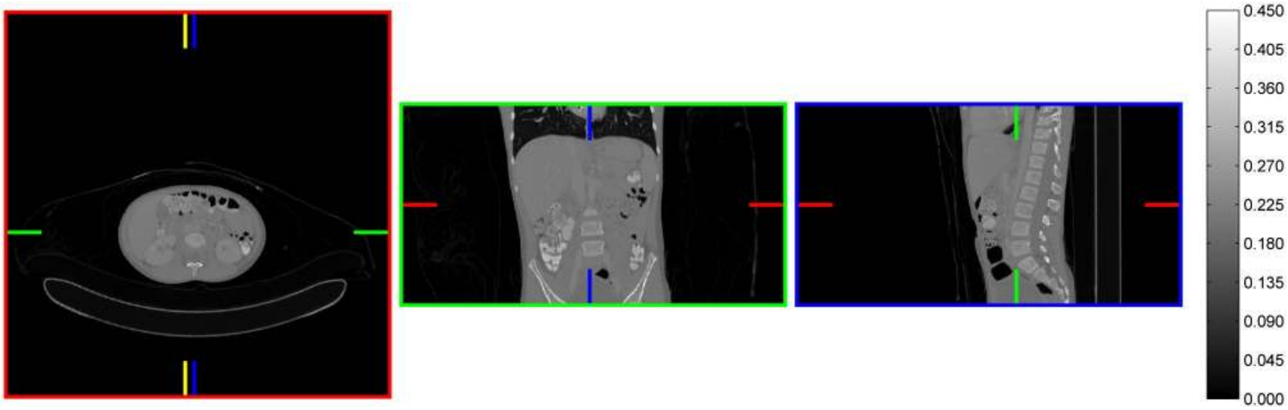


Fig. 19. Cross sections through the reconstruction of pediatric patient data using all the data, after 50 iterations of the AM algorithm, using 145 ordered subsets (to speed convergence) and a neighborhood penalty with $\beta = 3200$. The axial (red frame), coronal (green frame), and sagittal (blue frame) views are shown from left to right. The color-coded hatch marks within the enclosing frames indicate the planes from which the orthogonal views are taken. Linear attenuation coefficients from 0 (air) to 0.45 $\mathrm{cm}^{-1}$ (approximately bone) are shown in the range from black to white, respectively. The image volume is $512 \times 512 \times 135$ voxels, with voxel dimensions of 1 mm × 1 mm × 2 mm. A cut along the column indicated by yellow hatch marks in the axial view is shown in Fig. 22.

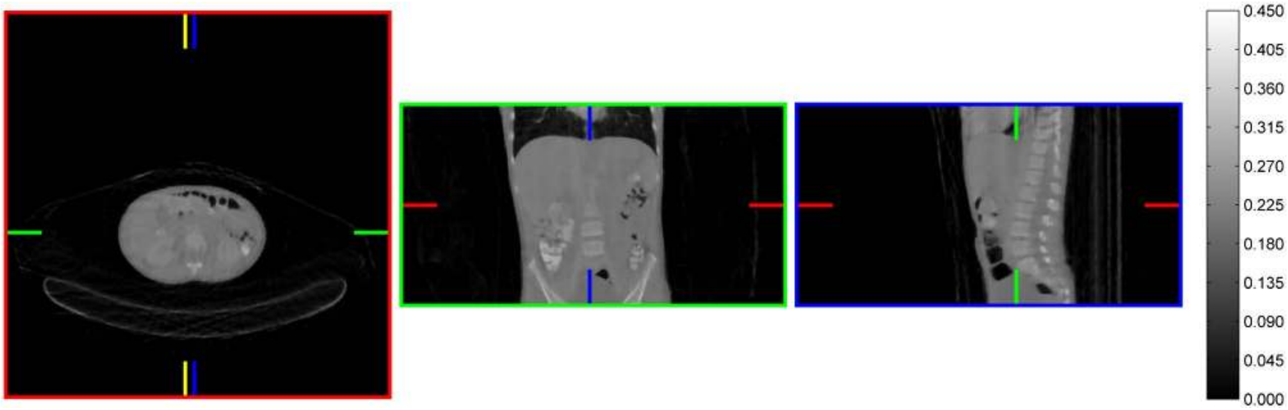


Fig. 20. Cross sections through the reconstructions of patient data using 1/29 of the views, uniformly spaced in angle of rotation around the patient, but all the detectors, and using 10 ordered subsets. All other specifications are the same as in Fig. 19.

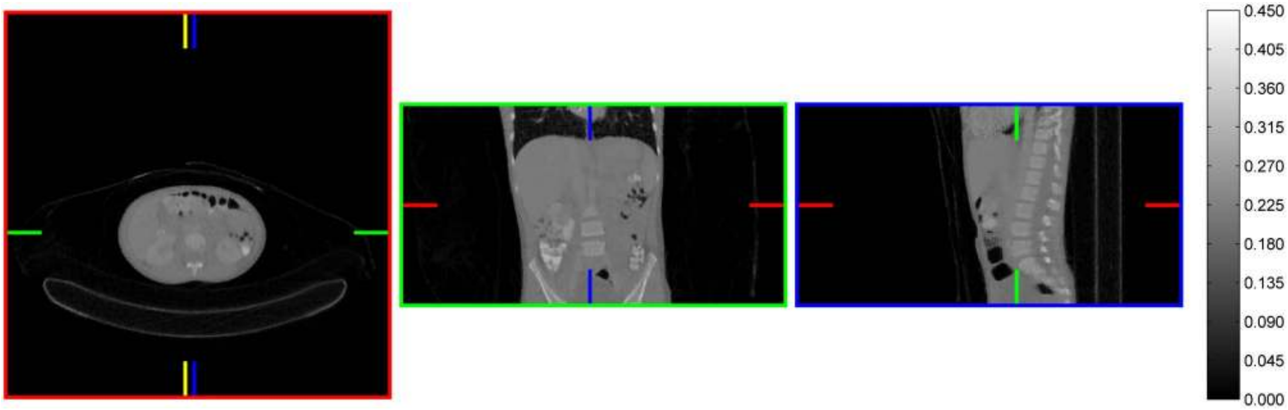


Fig. 21. Cross sections through the reconstructions of patient data using all of the views, but only 1/29 of the detector measurements, and using 10 ordered subsets. All other specifications are the same as in Fig. 19.

we used a Huber-class neighborhood penalty on the voxels. The penalty is of the form

$$\Psi(f) = \sum_{i} \sum_{j \in \mathcal{N}_i} w_{ij} \psi(f_i - f_j), \tag{16}$$

where $\mathcal{N}_i$ is the neighborhood for the $i$th voxel, $w_{ij}$ is a weight depending upon (inversely proportional to) the physical distance between the $i$th and $j$th voxels, $f_i$ is the linear attenuation constant at the $i$th voxel (as before), and $\psi$ is chosen to be locally quadratic for small values and linear for large values, such as

$$\psi(t) = \frac{1}{\delta^2} (|\delta t| - \ln(1 + |\delta t|)). \tag{17}$$

The penalty is weighted by a factor $\beta$ to control the trade-off between the data fit term (the log-likelihood in our case) and the penalty. For this example, we chose $\delta = 5000$ and $\beta = 3200$. This corresponds to a penalty that is close to a TV penalty. In order to validate this choice, we compared it to the maximum likelihood estimate, which is the MAP estimate in the limit where the penalty goes to zero (not shown here for brevity).

Figure 20 shows the result for the regularized AM algorithm using UV sampling with 1/29 of the views. Figure 21 shows the images for the same regularization parameters using data obtained by DRD subsampling with 1/29 of the data. The parameters for the regularization are the same as in Fig. 19. However, given that there are 1/29 the number of terms in the data fit, this effectively places a higher weight on the neighborhood penalty. For DRD sampling, the images remain close to the images obtained by using all of the data. By comparing DRD (Fig. 21) to UV (Fig. 20), one can see there are clearly fewer artifacts using DRD, and fine details (high-frequency content) are seen more clearly. To see these differences more clearly, we have presented in Fig. 22 the linear attenuation coefficients along lines marked in the images. The profiles shown in Fig. 22 demonstrate the fact that images reconstructed from DRD data exhibit fewer artifacts and suffer less from loss of contrast than images reconstructed from UV data.

Treating the regularized reconstruction of full data as the "gold standard," the sum of squared differences between it and other reconstructions are shown in Fig. 23. Unlike the previous experiments, UV sampling exhibited a smaller error than DRD sampling when one-tenth or more of the data were used. This may be understood by examining the sampling characteristics of the Siemens Sensation 16 CT scanner from which the data were acquired (see details in Appendix D). The sum of squared differences was computed only for voxels within the patient. The pediatric patient was small compared to an adult and positioned so that the maximum distance from the center of rotation occupied by the patient was measured to be approximately 125 mm. Therefore, the domain of interest including the patient was smaller than the one for which the system was designed and based on which the sampling rates were selected. Calculating the sampling rates according to the second row of Table 1 for this small object with $\rho = 125$ mm together with $r = 570$ mm for the Sensation 16 scanner reveals that this system oversamples the region where the patient is located with respect to view intervals. To see this, note that a system designed for this smaller region would require (based on Table 1) $\Delta\beta/\Delta\alpha = 5.56$, whereas the actual ratio $\Delta\beta/\Delta\alpha$ for the geometry and sampling of the Sensation 16 is 4. Since according to Table 1 only $\Delta\beta$ depends on $\rho$ (the size of the object), this means that $\Delta\beta$ is smaller than it needs to be (note that the working resolution, denoted as $\delta$ in Table 1, is the same for both cases, since the imaged domain and the number of voxels are the same). It is therefore not surprising that UV sampling performs better than DRD sampling for moderate downsampling rates, as is the case on the left portion of Fig. 23. However, as the fraction of data that is kept is reduced further, DRD performs better than UV and has a lower error, which is in accord with the conclusions of the previous experiment.

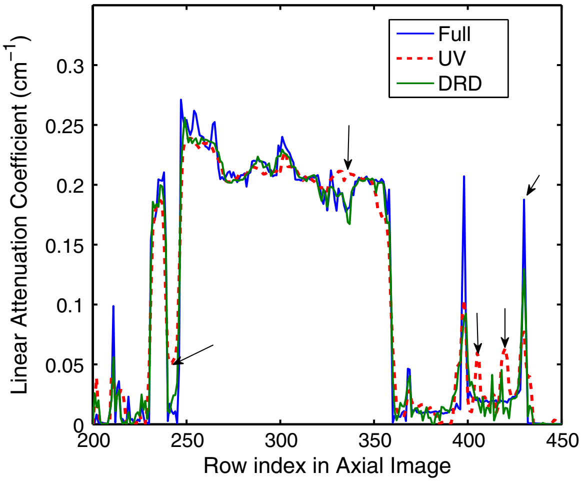


Fig. 22. Linear attenuation coefficients along the columns indicated by yellow hatch marks in Fig. 19 for full data (blue), Fig. 20 for UV sampling (red), and Fig. 21 for dynamic random detector (DRD) sampling (green), plotted on the same axis. Arrows indicate portions of the graph in which UV sampling led to substantial reconstruction errors. Some of these errors were in regions of sudden transition, indicating that some high-frequency content was lost.

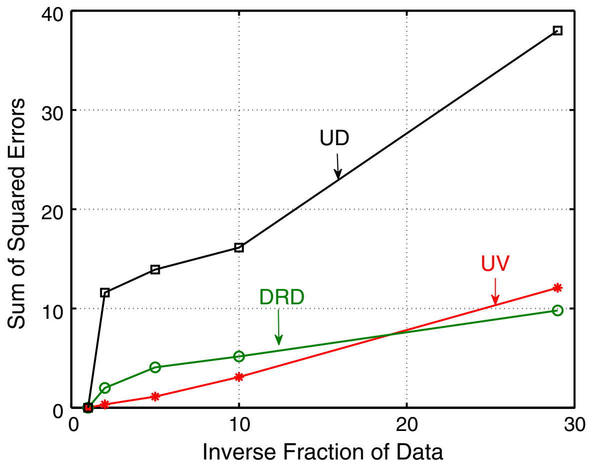


Fig. 23. Sum of squared differences between reconstructions using various sampling strategies and our gold standard, which used full data and regularization (shown in Fig. 19). The sum extends only over voxels within the $512 \times 512 \times 135$ image volume, which were inside the patient. The units are mm$^{-2}$. DRD, dynamic random detector; UV, uniform view; UD, uniform detector.

## 8. CONCLUSIONS

We have investigated sampling strategies for tomography that go beyond classical sampling theory and enable one to undersample the field of view while retaining the essential information about the object. Inspired by the field of CS, our approach is based on the understanding that many real objects are compressible in some known representation and that the number of degrees of freedom that defines an object is much lower than the number of pixels in the image, so fewer measurements should be sufficient if one incorporates this prior information during object reconstruction. We have proposed a new approach based on nonstructured (random) detector subsampling, whereas previous works have only addressed view-angle subsampling. The performance of different sampling strategies has been compared to the proposed approach using object-independent performance measures such as SVD and mutual coherence, and also based on reconstruction results for specific objects with synthetic and real data. In most parts of this paper we considered a fan-beam geometry and represented the imaged object using a Haar wavelet basis, but the framework we presented is far more general and can be applied to other geometries and for other sparse representations of the object. We have also demonstrated how the proposed approach is equally successful for spiral CT, using a different algorithm for the reconstruction and utilizing the fact that the gradient magnitude image is sparse.

We have shown that despite the differences between practical tomographic systems and CS theory, we have reached similar conclusions to what has been previously demonstrated in the CS literature, with a few important differences. Our conclusion is that distinguishing between different compressible objects in measurement space requires the sampling to be as unstructured as possible in *sinogram* space, i.e., what we termed DRD. In this approach, the selection of detectors changes quasi-randomly from one source position (view angle) to the other, and one can imagine this by randomly removing points in the sampled sinogram space (the 2D space where the horizontal axis corresponds to views and the vertical axis corresponds to detectors). This conclusion has been based on an SVD analysis and another study based on CS theory that employs coherence as a figure of merit. Interestingly, we have also shown that even if no prior information about the object is assumed, DRD is still the subsampling strategy least sensitive to noise.

Reconstructions based on simulated data clearly demonstrate the superiority of DRD subsampling over other subsampling strategies for a given number of measurements. We also presented reconstructions for real data acquired in an experimental x-ray laboratory at Duke University and from a commercial Siemens scanner for human imaging located at St. Louis Children's Hospital. In the former we used a fan-beam geometry to scan a 2D cross section of an object, while the latter is a spiral CT geometry that performs a 3D scan of a patient. These studies are very important in order to validate the practical usefulness of the proposed approach, since considerable model mismatch can occur. We have shown that the proposed approach sustains its advantages for real data and for two different geometries. The image quality obtained with the proposed approach was superior to the one achieved by view-angle subsampling with the same number of measurements, and in some cases even with fewer measurements. In addition, in some cases, combining detector and view-angle subsampling outperformed each of these strategies when used separately.

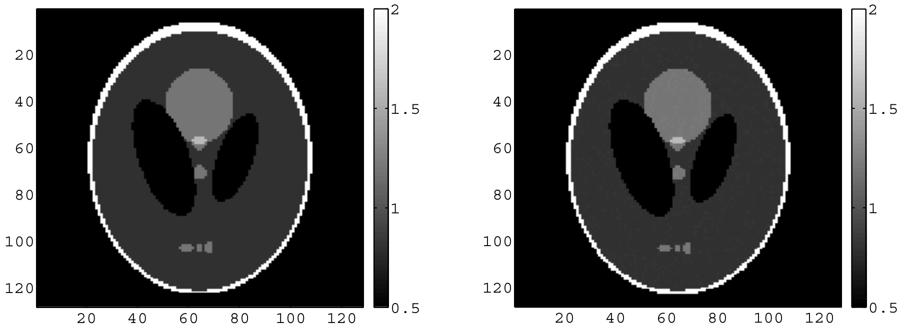
Nonuniform detector subsampling offers many advantages that are not taken into account in this work. It allows for the design freedom to combine detectors dedicated to measuring scattered radiation, which provides additional information about the object in some applications and enables estimation of the noise level in others. When detector subsampling is implemented by reducing the number of detectors, it leads to lower detector costs or, alternatively, allows integration of detectors for scattered radiation, which is of interest in several applications. In medical systems, it could be implemented using ray-blocking masks at the source side, leading to a reduced radiation dose to patients. Blocking rays can potentially reduce scattered radiation while keeping the same energy for the transmitted rays, resulting in a higher signal-to-noise ratio than the one achieved by lowering exposure times or the energy of the source. Also, the proposed detector sampling could be leveraged in the design of snapshot compressive tomography, where the object is simultaneously illuminated by several sources while a coded aperture performs detector subsampling, such that each source will only measure a single ray. All of these will be the subjects of future work.

We would like to comment on the connection between our work and prior art in [7,9,10,14–17], also discussed in Section 1. First, it is important to note that our work focuses on *comparing sampling strategies* and not on image reconstruction techniques. More specifically, we consider *how* to undersample the object even if a complete set of measurements can be taken, with the potential advantages mentioned above. Traditionally, prior art considers incomplete data due to *physical constraints* (as opposed to introducing them by design), and the focus is on a proposed image reconstruction technique to cope with the missing data. Regardless of the cause of missing data, some prior knowledge about the object needs to be incorporated into the reconstruction technique in order to provide regularization to the otherwise ill-posed problem. When comparing different sampling strategies, we have used the same image reconstruction technique in order to isolate the effect of the sampling strategy itself. Furthermore, we have used two different image reconstruction techniques to demonstrate that our results are independent of the specific technique. Several recent studies [7,14,17] consider the possibility of deliberately undersampling the object with respect to view angles but do not distinguish between designed and physically constrained undersampling, simply because there is no need for new system design to perform view-angle subsampling. In contrast, the methods of detector subsampling we consider do lead to new system designs. In [7,14,17] the authors consider UV subsampling in divergent beam geometries (cone beam, fan beam), and we have used this approach to compare to the proposed method. In some cases, such as in [10], there are physical constraints due to which the *number* of views is limited. Our results in Figs. 16 and 17 suggest the possibility of using RD subsampling in combination with few views; however, a more careful study is required. A different class of problems arises when the *range* of view angles is limited [5–7] and is known to be extremely ill-posed even before any sampling is considered [52]. This class of problems is beyond the scope of our work.

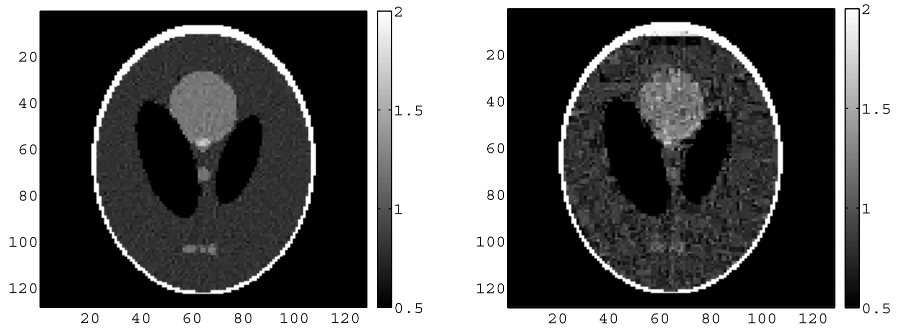
Lastly, we would like to note some remaining questions for future research. Our approach throughout this work was to implicitly assume a given object complexity (sparsity level), a given undersampling rate, and a given noise level, for which we have tried to determine the best subsampling strategy. According to CS theory, the sparsity level of the object (how compressible it is in some given representation) controls the maximal undersampling rate that can be used without considerable loss of detail. This suggests a study of the dependence between the sparsity level and the maximal sampling rate. Closely related to this is a question regarding the SDRD strategy, for which only a few random masks or detector patterns were used instead of different patterns (masks) for every view. It remains to be studied how the required number of masks (or detector patterns) changes with the object's complexity. To see the connection between these two questions, we note that by decreasing the number of RD patterns (masks), the rays are less spread out in the image domain and some regions of the object will be sampled at a lower rate. Therefore, in accordance with CS theory we expect the sparsity level to control the number of masks that are sufficient. Regarding the influence of noise, we have seen in Section 4.A that the singular value analysis suggests that our conclusions might change in the very low-noise regime. Therefore, another desired study is that of performance versus noise level. Despite this fact, it is important to note that the noise level in both simulations and experiments is standard for CT and is the same for all sampling strategies.

## APPENDIX A: OUTLINE OF THE DERIVATION OF SAMPLING CONDITIONS

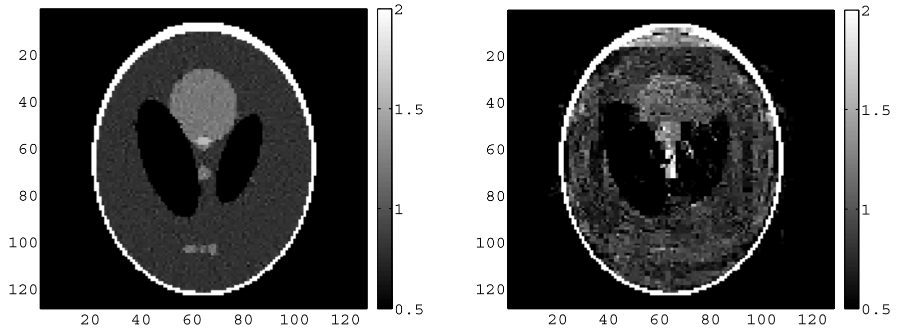
Next we quickly outline the derivation of the above sampling conditions and make the above statements more precise. In



(a) Truth (b) Full data



(c) DRD (d) RV



(e) SDRD (f) SRD

Fig. 24. Values represent the attenuation per unit length divided by the attenuation per unit length of water. (a) Truth; (b)–(f) reconstructed images via RVM: (b) using all measurements, (c)–(f) after measurements have been downsampled by ×32. (c) Dynamic random detector sampling (DRD), (d) random view sampling (RV), (e) SDRD sampling, (f) static random detector sampling (SRD). Only four iterations of RVM have been used. The data have been generated according to the model in Eqs. (14) and (15) with $I_0 = 10^5$. Showing values in $[0.5, 2]$;.

doing so, we shall try to keep the details down to a minimum, and for further reading we recommend [2]. The theoretical analysis starts with the definition of an essentially band-limited function $f(\mathbf{x})$ that satisfies

$$\int_{|\boldsymbol{\xi}|>b} |\hat{f}(\boldsymbol{\xi})|^2 \mathbf{d}\boldsymbol{\xi} \leq \epsilon_0, \tag{A1}$$

where $\hat{f}$ is the 2D Fourier transform of $f(\mathbf{x})$ and $\epsilon_0$ is assumed to be small. We shall denote the set of line integrals over $f$ by the function $g(\boldsymbol{\gamma})$ given by either the Radon or fan-beam transform in Eq. (1) or (2), respectively. The function $g$ is periodic with respect to the variables in $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T$ which are generic line parameters, e.g., $\boldsymbol{\gamma} = (\beta, \alpha)^T$ for fan beams (see Fig. 1). Next, we denote the set of samples of $g$ on a grid by

$$S \triangleq \{\boldsymbol{\gamma} = \mathbf{Wl}, \quad \mathbf{l} \in \mathbb{Z}^2\}, \tag{A2}$$

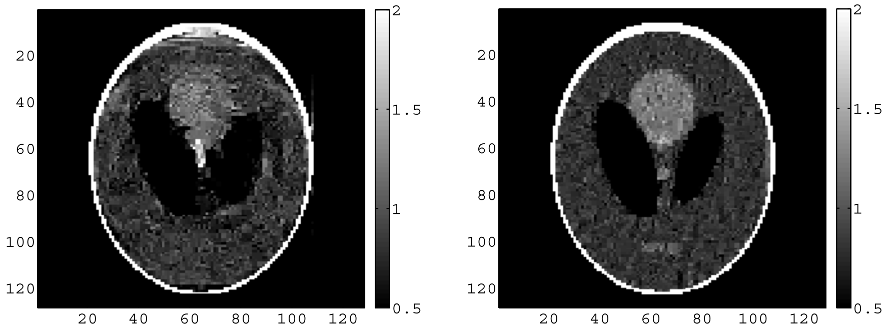
where $\mathbf{W}$ is a real and nonsingular $2 \times 2$ matrix that is defined by the scanning geometry and the sampling rates for $\boldsymbol{\gamma}$. To find sufficient conditions in terms of the sampling rates for $\boldsymbol{\gamma}$, one constructs a set $\mathbf{K}$ in the 2D $\boldsymbol{\gamma}$ space that is tailored to the specific scanning geometry. Utilizing the Petersen–Middleton theorem [53] for periodic functions, it can be proven that for a lattice $\mathbf{W}$ such that the sets $\{\mathbf{K} + 2\pi(\mathbf{W}^{-1})^T\mathbf{l}, \mathbf{l} \in \mathbb{Z}^2\}$; are mutually disjoint and for any $f$ such that $g(\mathbf{S}) = 0$, we have [2]

$$\|\mathbf{R}f\|_{L_\infty} \leq C_1 e^{-\lambda b}\|f\|_{L_1} + C_2\epsilon_0, \tag{A3}$$

where $C_1, C_2, \lambda$ are some positive constants, $\|\cdot\|_{L_1}, \|\cdot\|_{L_\infty}$ denote the norms in the corresponding Lebesgue spaces, and $b$ and $\epsilon_0$ are defined in Eq. (A1). The sampling rates for $\boldsymbol{\gamma}$; are then determined by the conditions for which the sets $\{\mathbf{K} + 2\pi(\mathbf{W}^{-1})^T\mathbf{l}, \mathbf{l} \in \mathbb{Z}^2\}$; are mutually disjoint. In plain English, Eq. (A3) states that if $f$ yields zero line integrals for the samples satisfying the conditions above, then the Radon transform of $f$ cannot have an arbitrarily large norm and is bounded by terms proportional to $\epsilon_0$ (the deviation from band-limited functions) and the $L_1$ norm of $f$. In that sense, choosing small $C_{1,2}$ restricts the null space of the operator $f \to g(\mathbf{Wl})$ to negligible functions. This assures us that a function can be reconstructed reliably (stably), i.e., a small error in measurements does not correspond to an arbitrarily large error in the reconstructed function (as measured by the $L_\infty$ norm of the Radon transform). The constants $C_1, C_2, \lambda$ and the set $\mathbf{K}$ for parallel-beam geometry can be found in [2,54] and for the fan-beam geometry with curved and flat detectors in [39] and [55], respectively. It is important to note that in this approach the sampling conditions are derived by construction (namely, the choice of the set $\mathbf{K}$) and so are only sufficient conditions, not necessary ones. Indeed, there are well-known examples where more efficient sampling conditions have been found, such as the interlaced parallel beam geometry [54] and its analog in fan-beam geometry [39]. This line of work does not use any assumptions on $f$ such as sparsity.

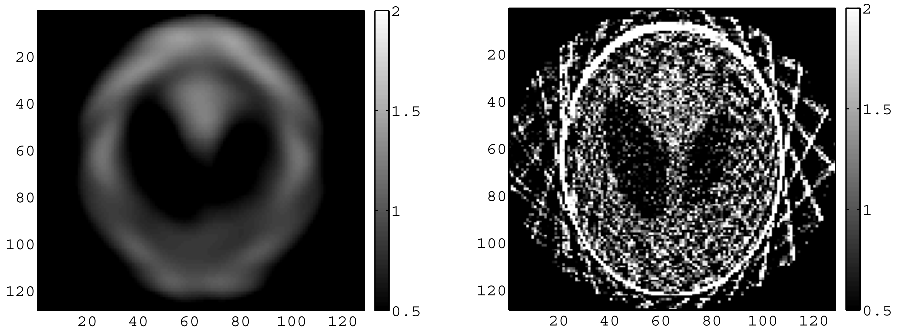
## APPENDIX B: RECONSTRUCTION RESULTS FOR SYNTHETIC DATA

In order to see more clearly the differences between the different sampling strategies, we present the reconstructions of Fig. 9 and Fig. 10 in the main paper with the attenuation values



(a) RVM for Uniform detector (UD)

(b) RVM for Uniform view (UV)



(c) FBP for Uniform detector (UD)

(d) FBP for Uniform view (UV)

Fig. 25. Reconstructed images for uniform undersampling ×32. (a), (b) Using RVM after four iterations; (c), (d) using filtered backprojection (FBP). (a),(c) UD sampling; (b), (d) UV sampling. Only four iterations in RVM have been used. The data have been generated according to the model in Eqs. (14) and (15) with $I_0 = 10^5$. *The attenuation values are limited to the range* $[0.5, 2]$, where the differences are more apparent.

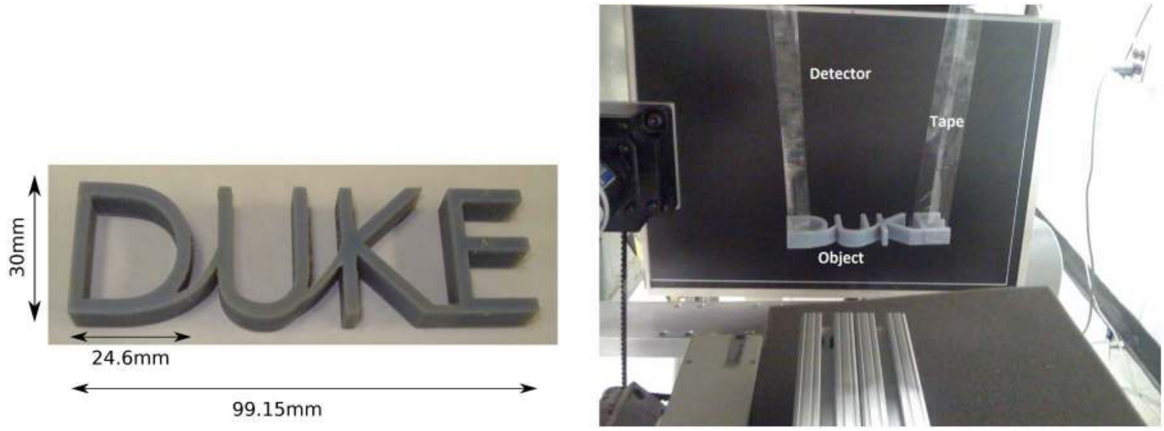


Fig. 26. Left: the scanned object, 3D letters “DUKE.” Right: positioning of object. We used tape, and other means, to suspend objects in the air as well as to minimize scatter compared to other potential holders.

limited to the window [0.5, 2];, and these are shown in Figs. 24 and 25, respectively.

## APPENDIX C: EXPERIMENTAL SETUP FOR THE PHYSICAL PHANTOM DATASET

The data have been acquired at the Duke Multi-Modality Imaging Laboratory (MMIL) [56]. The object under test consists of the 3D letters “DUKE” made from the VeroBlue material (see Fig. 26) used in 3D printers, which according to our experiments has a linear attenuation constant 1.458 times that of water. The x-ray source specifications are 60 kVp, 50 mA, and 25 ms. A 0.55 mm Cerium filter was used to strongly filter the bremsstrahlen (continuous spectrum x-rays) and create a pseudo-monoenergetic source spectrum. The flux-energy curve of the source with the Cerium filter was calculated with XSPECT software [57] and is shown in Fig. 27. The detector has a pixel pitch of 127 μm and a fill factor of 0.75 (pixel pitch divided by pixel size). Detector measurements were software binned into $2 \times 2$ pixels to increase the number of photon counts per pixel. By integrating the flux-energy curve and accounting for detector fill factor, we estimated a total of $I_0 = 2 \times 10^3$ photons per detector pixel. The view angle resolution was 1°. (In order to achieve the flux requirements necessary for proper photon statistics and model assumptions, x-ray tube heating constraints limited the number of acquisitions while maintaining a consistent experiment to about 1° per frame). Additional details are shown in Fig. 28.

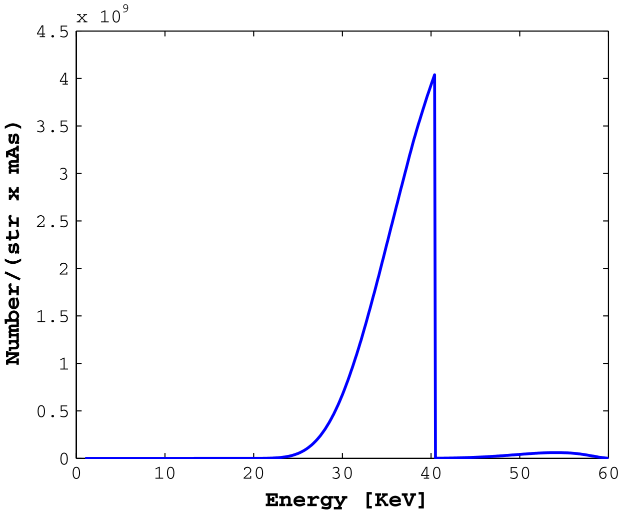


Fig. 27. Simulation of the flux-energy curve for the x-ray source with the 0.55 mm Cerium filter used in the experiment, obtained by XSPECT software.

## APPENDIX D: EXPERIMENTAL SETUP FOR THE CLINICAL DATASET

With approval from the Washington University IRB, patient data were acquired on a Siemens Sensation 16 scanner at St. Louis Children’s Hospital using a standard abdominal imaging protocol with contrast agent. The Sensation 16 is a third-generation multidetector-row CT scanner and was operated in spiral scanning mode. The x-ray source orbits around the isocenter of the system every 0.5 s at a distance of 570 mm. The number of uniformly spaced views per rotation is 1160. The source is collimated to a width of 1.5 cm at the isocenter of the system. The detectors are arranged in 16 rows of 672 detectors, each on a cylindrical arc with a radius of 1040 mm and centered at the source. Each detector subtends an arc of 0.00135413 rad, so the entire fan subtends approximately 52°. The center of the detector array was offset by 1/4 of a

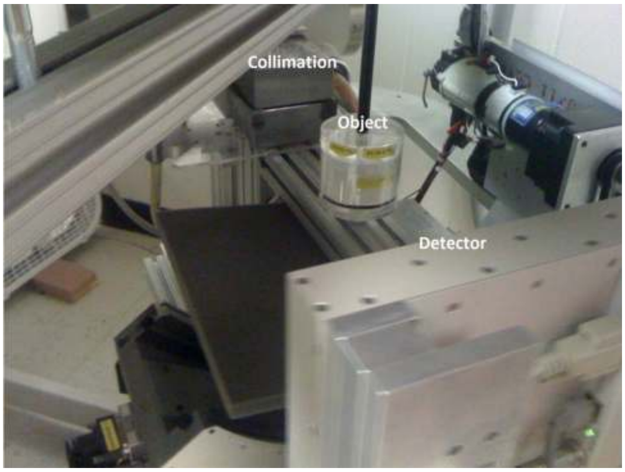


Fig. 28. We collimated the source into a fan beam using two lead blocks that were both made flat and parallel with a fly cutter. The length of collimator blocks helped ensure the beam had a very limited divergence.

detector to improve sampling. The voltage and current of the x-ray tube were 120 kV and 150 mAs, respectively. The patient bed traveled 24 mm per rotation of the gantry at a uniform speed.

## ACKNOWLEDGMENTS

This work was supported by the Department of Homeland Security, Science and Technology Directorate, Explosives Division, through contract HSHQDC-11-C-0083. We would like to thank Prof. Martin P. Tornai for providing us access to the Duke Multi-Modality Imaging Lab (MMIL) where some of the experiments were conducted. Finally, we would like to thank the anonymous reviewers for providing helpful comments about the paper.

## REFERENCES

1. M. Slaney and A. Kak, *Principles of Computerized Tomographic Imaging* (SIAM, 1988).
2. F. Natterer, *The Mathematics of Computerized Tomography* (Wiley, 1986).
3. E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. Pure Appl. Math. **59**, 1207–1223 (2006).
4. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).
5. R. Rangayyan, A. P. Dhawan, and R. Gordon, "Algorithms for limited-view computed-tomography—an annotated bibliography and a challenge," Appl. Opt. **24**, 4000–4012 (1985).
6. A. K. Louis, "Incomplete data problems in x-ray computerized tomography," Numer. Math. **48**, 251–262 (1986).
7. E. Y. Sidky, C. M. Kao, and X. H. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," J. X-Ray Sci. Technol. **14**, 119–139 (2006).
8. M. H. Li, H. Q. Yang, and H. Kudo, "An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections," Phys. Med. Biol. **47**, 2599–2609 (2002).
9. J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," Med. Phys. **34**, 4476–4483 (2007).
10. E. Hansis, D. Schafer, O. Dossel, and M. Grass, "Evaluation of iterative sparse object reconstruction from few projections for 3-D rotational coronary angiography," IEEE Trans. Med. Imaging **27**, 1548–1555 (2008).
11. D. C. Youla and H. Webb, "Image restoration by the method of convex projections: part 1 theory," IEEE Trans. Med. Imaging **1**, 81–94 (1982).
12. M. I. Sezan and H. Stark, "Image restoration by the method of convex projections: part 2 applications and numerical results," IEEE Trans. Med. Imaging **1**, 95–101 (1982).
13. E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**, 489–509 (2006).
14. E. Y. Sidky and X. C. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).
15. G. H. Chen, J. Tang, and S. H. Leng, "Prior image constrained compressed sensing: a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," Med. Phys. **35**, 660–663 (2008).
16. J. W. Stayman, W. Zbijewski, Y. Otake, A. Uneri, S. Schafer, J. Lee, J. L. Prince, and J. H. Siewerdsen, "Penalized-likelihood reconstruction for sparse data acquisitions with unregistered prior images and compressed sensing penalties," Proc. SPIE **7961**, 79611L (2011).
17. K. Choi, J. Wang, L. Zhu, T. S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," Med. Phys. **37**, 5113–5125 (2010).
18. A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," Opt. Express **17**, 6368–6388 (2009).
19. J. Hahn, S. Lim, K. Choi, R. Horisaki, and D. J. Brady, "Video-rate compressive holographic microscopic tomography," Opt. Express **19**, 7289–7298 (2011).
20. K. P. MacCabe, A. D. Holmgren, M. P. Tornai, and D. J. Brady, "Snapshot 2D tomography via coded aperture x-ray scatter imaging," Appl. Opt. **52**, 4582–4589 (2013).
21. J. Greenberg, K. Krishnamurthy, and D. Brady, "Compressive single-pixel snapshot x-ray diffraction imaging," Opt. Lett. **39**, 111–114 (2014).
22. D. Brady and D. Marks, "Coding for compressive focal tomography," Appl. Opt. **50**, 4436–4449 (2011).
23. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," Opt. Express **21**, 10526–10545 (2013).
24. D. J. Brady, N. Pitsianis, and X. Sun, "Reference structure tomography," J. Opt. Soc. Am. A **21**, 1140–1147 (2004).
25. K. Choi and D. J. Brady, "Coded aperture computed tomography," Proc. SPIE **7468**, 74680B (2009).
26. R. Ning, X. Tang, and D. L. Conover, "X-ray scatter suppression algorithm for cone-beam volume CT," Proc. SPIE **4685**, 774–781 (2002).
27. R. Ning, X. Tang, and D. Conover, "X-ray scatter correction algorithm for cone beam CT imaging," Med. Phys. **31**, 1195–1202 (2004).
28. G. Harding, J. Kosanetzky, and U. Neitzel, "X-ray diffraction computed tomography," Med. Phys. **14**, 515–525 (1987).
29. H. Strecker, "Automatic detection of explosives in airline baggage using elastic x-ray scatter," Medicamundi **42**, 30–33 (1998).
30. C. Cozzini, S. Olesinski, and G. Harding, "Modeling scattering for security applications: a multiple beam x-ray diffraction imaging system," in *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)* (IEEE, 2012), pp. 74–77.
31. F. Smith, *Industrial Applications of X-Ray Diffraction* (CRC Press, 1999).
32. S. Pani, E. Cook, J. Horrocks, L. George, S. Hardwick, and R. Speller, "Modelling an energy-dispersive x-ray diffraction system for drug detection," IEEE Trans. Nucl. Sci. **56**, 1238–1241 (2009).
33. J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances* (SPIE, 2009).
34. E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Probl. **23**, 969–985 (2007).
35. F. Krahmer and R. Ward, "Beyond incoherence: stable and robust sampling strategies for compressive imaging," arXiv:1210.2380 (2012).
36. B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing," arXiv:1302.0561 (2013).
37. H. Peng and H. Stark, "Direct fourier reconstruction in fan-beam tomography," IEEE Trans. Med. Imaging **6**, 209–219 (1987).
38. G.-H. Chen, S. Leng, and C. A. Mistretta, "A novel extension of the parallel-beam projection-slice theorem to divergent fan-beam and cone-beam projections," Med. Phys. **32**, 654–665 (2005).
39. F. Natterer, "Sampling in fan beam tomography," SIAM J. Appl. Math. **53**, 358–380 (1993).
40. http://www.mathworks.com/matlabcentral/fileexchange/27375-plot-wavelet-image-2d-decomposition/content/plotwavelet2.m.
41. D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$ minimization," Proc. Natl. Acad. Sci. USA **100**, 2197–2202 (2003).
42. M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," IEEE Signal Process. Mag. **25**(2), 72–82 (2008).
43. M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," Magn. Reson. Med. **58**, 1182–1195 (2007).
44. M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications* (Cambridge University, 2012), Chap. 1.

45. J. A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” IEEE Trans. Inf. Theory **52**, 1030–1051 (2006).
46. M. Sonka and J. M. Fitzpatrick, *Handbook of Medical Imaging*, Vol. **2** of Medical Image Processing and Analysis (SPIE, 2000).
47. G. T. Herman, *Image Reconstruction from Projections* (Academic, 1980).
48. M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” J. Mach. Learn. Res. **1**, 211–244 (2001).
49. S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” IEEE Trans. Signal Process. **56**, 2346–2356 (2008).
50. J. M. Bernardo and A. F. Smith, *Bayesian Theory*, Vol. **405** of Wiley Series in Probability and Statistics (Wiley, 2009).
51. J. A. O’Sullivan and J. Benac, “Alternating minimization algorithms for transmission tomography,” IEEE Trans. Med. Imaging **26**, 283–297 (2007).
52. M. E. Davison, “The ill-conditioned nature of the limited angle tomography problem,” SIAM J. Appl. Math. **43**, 428–448 (1983).
53. D. P. Petersen and D. Middleton, “Sampling and reconstruction of wave-number-limited functions in N-dimensional euclidean spaces,” Inf. Control **5**, 279–323 (1962).
54. A. G. Lindgren and P. A. Rattey, “The inverse discrete Radon transform with applications to tomographic imaging using projection data,” Adv. Electron. Electron Phys. **56**, 359–410 (1981).
55. S. H. Izen, “Sampling in flat detector fan beam tomography,” SIAM J. Appl. Math. **72**, 61–84 (2012).
56. D. J. Crotty, R. L. McKinley, and M. P. Tornai, “Experimental spectral measurements of heavy k-edge filtered beams for x-ray computed mammotomography,” Phys. Med. Biol. **52**, 603–616 (2007).
57. C. W. Dodge, *A Rapid Method for the Simulation of Filtered X-Ray Spectra in Diagnostic Imaging Systems* (ProQuest, 2008).