SCIENTIFIC REPORTS

OPEN

Sparse recovery of undersampled intensity patterns for coherent diffraction imaging at high X-ray energies

S. Maddali[1], I. Calvo-Almazan[1], J. Almer[2], P. Kenesei[2], J.-S. Park[2], R. Harder[2], Y. Nashed[3,4] & S. O. Hruszkewycz[1]

Coherent X-ray photons with energies higher than 50 keV offer new possibilities for imaging nanoscale lattice distortions in bulk crystalline materials using Bragg peak phase retrieval methods. However, the compression of reciprocal space at high energies typically results in poorly resolved fringes on an area detector, rendering the diffraction data unsuitable for the three-dimensional reconstruction of compact crystals. To address this problem, we propose a method by which to recover fine fringe detail in the scattered intensity. This recovery is achieved in two steps: multiple undersampled measurements are made by in-plane sub-pixel motion of the area detector, then this data set is passed to a sparsity-based numerical solver that recovers fringe detail suitable for standard Bragg coherent diffraction imaging (BCDI) reconstruction methods of compact single crystals. The key insight of this paper is that sparsity in a BCDI data set can be enforced by recognising that the signal in the detector, though poorly resolved, is band-limited. This requires fewer in-plane detector translations for complete signal recovery, while adhering to information theory limits. We use simulated BCDI data sets to demonstrate the approach, outline our sparse recovery strategy, and comment on future opportunities.

Coherent X-ray diffraction imaging (CDI) methods applied to Bragg peaks have emerged as a powerful tool in materials science for characterising lattice distortion fields and defects in crystalline nanostructures[1–4]. Such experiments are currently feasible at photon energies in the lower end of the hard X-ray spectrum (9 keV to 16 keV), beyond which third-generation synchrotrons have very limited coherent flux. However, ongoing and planned construction of next-generation synchrotrons will bring about greatly increased coherence at beam energies greater than 50 keV, making high-energy Bragg CDI feasible. This capability will enable nanoscale structural characterization of material volumes in environments accessible only with highly penetrating X-rays. For example, a critical new research area will be to probe deformation states within the individual grains of macroscopic volumes, which find wide application as structural and functional materials. Such measurements will complement and enhance models gleaned from existing non-destructive imaging techniques such as high-energy diffraction microscopy (HEDM)[5,6] and diffraction contrast tomography[7] by providing structure and strain resolution within grains. As a step in this direction, this paper addresses prominent issues that arise when imaging individual nano- or micro-scale crystals with high-energy coherent X-rays. In the context of structural materials, the compact crystal is a proxy for a single grain in a polycrystal. Otherwise, it can represent an isolated nanocrystal in a dense medium that requires the penetrative power of high energy X-rays (e.g. high temperature catalysis, or materials growth), opening yet more potential avenues of study[8]. Our treatment is focused on simulations of a BCDI measurement, in which a single compact crystal, illuminated uniformly by coherent high-energy X-rays, is rotated through the Bragg condition. Consequently a three-dimensional volume of reciprocal space is measured with an area detector in a sequence of parallel layers (Fig. 1(a)).

[1]Materials Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA. [2]X-ray Sciences Division, Argonne National Laboratory, Lemont, IL, 60439, USA. [3]Mathematics & Computer Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA. [4]Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL, 60208, USA. Correspondence and requests for materials should be addressed to S.M. (email: smaddali@anl.gov)
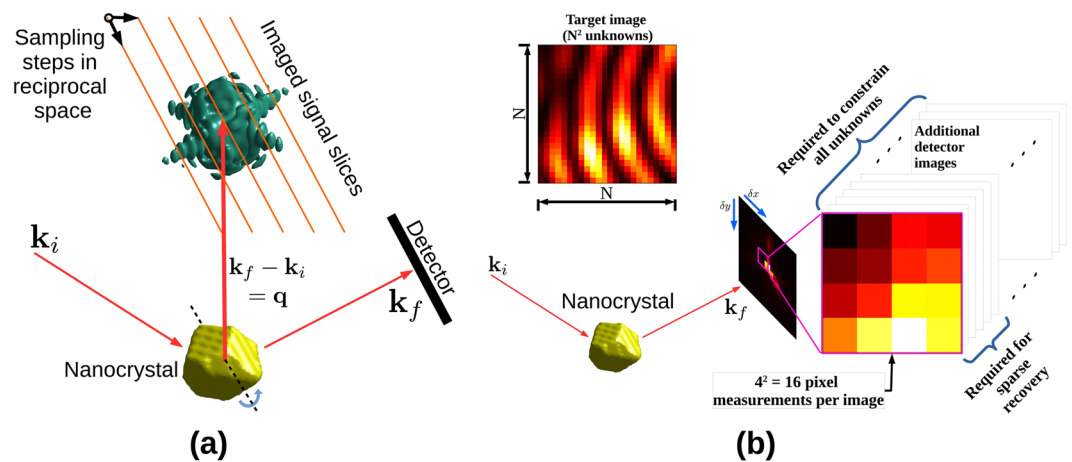
Figure 1. (**a**) Bragg CDI experimental geometry, where $\mathbf{k}_i$ and $\mathbf{k}_f$ are the incident and scattered wave vectors. Three-dimensional reciprocal space is sampled in finite steps as shown. Two of the three steps are in the detector plane (one of which is perpendicular to the figure plane, denoted by '∘') and the third dictates the migration of the plane. This set of sampling vectors is fixed by the crystallographic orientation of the scatterer with respect to the beam, and the manner of its rotation. (**b**) Acquisition of more data by detector translations perpendicular to $\mathbf{k}_f$ in case of coarse detector resolution. Application of sparse recovery techniques greatly reduces the number of additional measurements required.
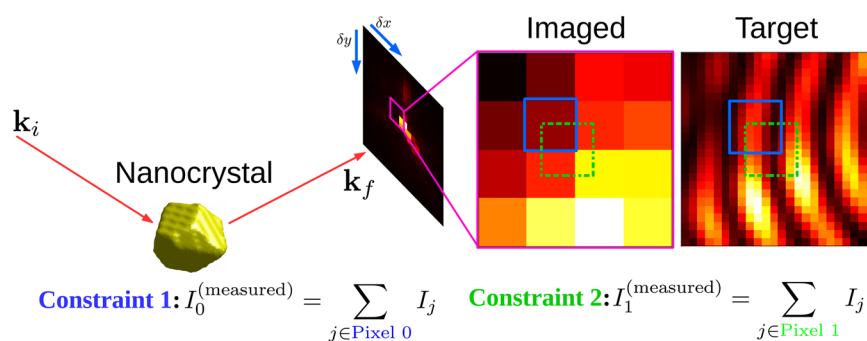
Figure 2. The target of the recovery algorithm is a sufficiently well-resolved approximation to the continuous scattered field. With each sub-pixel translation of the detector, new constraints on the pixel values are obtained.

In a BCDI experiment, increasing the beam energy from a value currently suitable for CDI (about 9 keV) to one necessary for HEDM ($\geq$50 keV) compresses reciprocal space by at least 80% in each dimension. Under these conditions, the solid angle subtended by each detector pixel covers a substantial portion of the diffraction pattern. As a result signal features (diffraction fringes) critical to inversion of the diffraction pattern are no longer resolved by a typical pixelated X-ray area detector. This is demonstrated in Fig. 1(b) which shows a coherent diffraction intensity pattern depicted with well-resolved and under-resolved fringes, mimicking measurements made at lower and higher X-ray energies respectively. The subsequent loss of diffraction feature visibility at higher energies renders conventional phase retrieval ineffective.

For the recovery of fine feature detail in the intensity pattern, additional independent measurements (upsamples) of the wave field intensity are therefore necessary. In this work, we consider the case where these additional measurements come from translating the area detector perpendicular to the exit beam in sub-pixel steps, as has been explored previously[9]. Each pixel measurement in such a data set imposes a constraint on a particular region of the scattered intensity (Fig. 2). In theory, the fine detail at a desired sub-pixel resolution could be recovered if one were to acquire an adequate number of these constraints, taken from sufficiently small detector offsets. The difficulties in fulfilling this requirement are described in detail in Section 1 of the Supplementary Material. In this paper we utilise compressed sensing[10,11] to demonstrate signal recovery with significantly fewer of such pixel constraints than there are fine pixels in the desired image. The required number of constraints is dictated by the information content when the desired image is expressed in Fourier space. The reduced number of constraints (or measurements, in compressed sensing parlance) is made possible by the fact that the BCDI intensity pattern of a compact single crystal is necessarily band-limited. In the compressed sensing approach employed here, one needs to make approximately as many coarse pixel measurements of the poorly-resolved signal as there are non-zero components in the Fourier representation of the well-resolved signal. The specific representation used in our method is the cosine basis. Recovery of this condensed representation of the signal is perfectly suited

to mathematical optimization algorithms that specialize in sparse arrays. Throughout this paper, we refer to the upsampled intensity patterns resulting from this sparse recovery scheme as the 'recovered' images or patterns.

This paper is organised in the following manner: Beginning with simplifying assumptions, we describe our BCDI simulations at a range of energies from 9 keV through to 54 keV, the latter corresponding to a high-energy experiment conceivable at a high-brightness synchrotron. This includes establishing a suitable ground truth intensity pattern with well-resolved intensity fringes consistent with currently feasible measurements, against which all recovered intensity patterns are benchmarked. We describe the detector sub-pixel translation method and the additional constraints provided by the intrinsic band-limit of the intensity pattern, followed by the compressed sensing algorithm employed to recover each slice of the 3D data set independently. We demonstrate the recovery of qualitatively different intensity patterns, with simulations that represent different beam energies and varying sub-pixel detector translations. Finally, we show three-dimensional reconstructions of a simulated nanocrystal obtained by applying conventional BCDI phase retrieval to simulated high-energy data sets recovered with our approach. We describe how our method can emulate a variety of physical modifications to a BCDI experiment that may be difficult to realise experimentally. We close with comments on new experiments that this method can potentially enable at next-generation synchrotrons.

High-energy BCDI Simulations

BCDI measurements query the wave intensity in the Fraunhofer regime[12] in which the scattered wave front is the Fourier transform (FT) of the illuminated compact crystal. As the scatterer is rotated through the Bragg condition the intensity pattern is measured, slice by slice, on the area detector. For a crystal of size ~300 nm illuminated with hard X-rays of energy ~9 keV, the required range of this rotation is sufficiently small (about 0.7°) that the successive slices are approximated as parallel in reciprocal space[4].

We simulate the compact crystal as the set of grid points inside a faceted volume, at the centre of a three-dimensional complex array. The interior points are complex-valued with magnitude 1 and a spatially varying phase that mimics a strain field. The exterior points have magnitude 0. The corresponding diffraction signal is obtained via the three-dimensional FT. In the 3D array of this FT, two dimensions represent pixel coordinates of the area detector while the third represents successive images acquired by rotating the crystal through the Bragg condition. In our simulations, the crystal resides inside a $22 \times 24 \times 22$ box within the simulation array of size $128 \times 128 \times 70$, denoting 70 angular steps with a 128×128 pixel detector. The ground truth intensity pattern is taken to be the squared modulus of the FT of this array. We nominally associate a beam energy of 9 keV to this ground truth. A data set of this kind could conceivably be collected at an existing BCDI facility.

Simulation of an overbinned diffraction signal at higher energies is now straightforward: for each 2D slice in the ground truth, blocks of pixels are summed to a single intensity value (Fig. 2). This binning operation mimics a high-energy BCDI experiment since photons that would have spread over a larger solid angle at lower x-ray energies now aggregate into fewer pixels at higher energies because of the compression of reciprocal space. Equivalently, one may imagine the overbinning to arise from a 9 keV measurement with proportionately larger pixels. The ratio of pixel sizes ('pixel binning factor' or PBF) is equal to the ratio of the beam energies. For example, the diffraction features at 9 keV contained in every 6×6 block of pixels is squeezed into a single pixel when the beam energy is 54 keV. In this case, PBF = 6.

In this manner binning effectively reduces the feature visibility of a BCDI intensity measurement (Fig. 1(b)), which we redress through additional information collected from sub-pixel detector translations. A similar method for ptychography has been proposed by Batey *et al.*[13]. In general the finite size of the physical pixels makes it impossible to sufficiently constrain the intensity features of the ground truth that are lost to binning, regardless of how finely the detector is translated in its plane. This assertion is rigorously proved in Section 1 of the Supplementary Material. As mentioned earlier, our approach instead is to demonstrate sub-pixel-scale feature recovery with fewer measurements, which is enabled by the sparsity of the Fourier representation of the scattered intensity. This is the focus of Section 3 in which we discuss our binned data sets in the context of a compressed sensing measurement and motivate the use of sparse recovery techniques.

We point out that in a real-world BCDI experiment at energies >9 keV, a given detector would span a larger q-space aperture that would not be entirely covered at 9 keV. In our simulations we consider a fixed q-space aperture corresponding to the 128×128 detector space of the ground truth. We focus on the recovery of fine features in the original ground truth simulation from binned data sets that subtend this q range. We justify this by pointing out that in typical BCDI experiments at 9 keV the edge pixels of the detector capture relatively low intensities compared to the central pixels (by a few orders of magnitude). The scattered intensities outside this aperture contribute negligibly to the Fourier representation of the original scatterer.

Sparse recovery: Mathematical Details

We seek a method to reverse the binning process described in Section 2 and obtain the original ground-truth diffraction pattern that features well-resolved fringes. In our method, recovery of the fine detail from a limited set of binned pixel measurements hinges on representing each two-dimensional slice in the discrete cosine basis, which we utilise as a numerically convenient variant of the FT. The two-dimensional discrete cosine transform (DCT) is defined for an $N \times N$-sized image A_{ij} as the linear transformation[14]:

$$\left[\mathrm{DCT}\left(A_{ij}\right)\right]_{mn} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} A_{ij}\mathbb{C}(i;\,m)\mathbb{C}(j;\,n)$$

(1)

where $\mathbb{C}(p;\,q) = \cos[(p+1/2)q\pi/N]$. This transform is instrumental in many digital media compression formats such as JPEG and MP3[15,16]. Figure 3(b) and (c) compare the FT and DCT representations of the diffraction pattern in 3(a). We demonstrate in Section 4 that the results of the sparse recovery algorithm do not depend on the
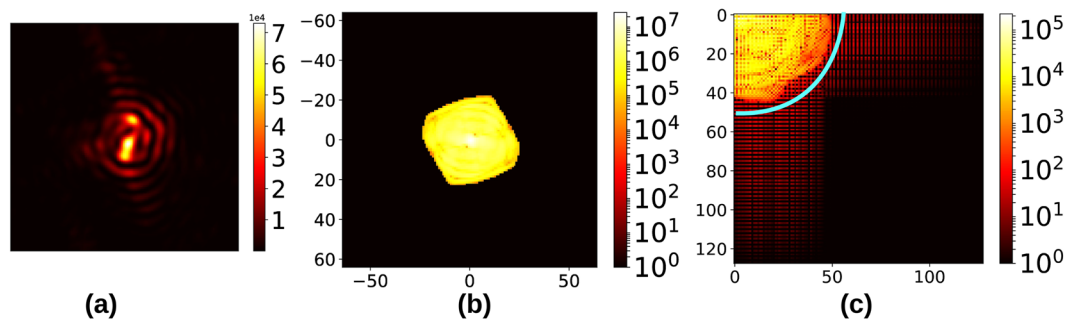
Figure 3. (**a**) Diffraction from a faceted object, (**b**) Magnitude of its FT, (**c**) Magnitude of its 2-D DCT. In the cosine basis, the spectral components inside the marked region contribute the most to (**b**). The axes of (**b**) and (**c**) have been labeled to reflect the correct information content in the respective Fourier/cosine components, with the usual row-major matrix convention of coordinates.

smallest DCT components. This gives it a certain degree of robustness in the presence of detector noise. Section 3 of the Supplementary Material addresses the issue of sparse recovery from noisy data.

The three-dimensional diffraction pattern from a finite crystal is inherently a band-limited function, since its FT is the auto-correlation of the scattering factor of a finite crystal (the Patterson function, Fig. 3(b)). From this it follows that each two-dimensional slice of the diffraction pattern is also band-limited (this is proved rigorously in Section 2 of the Supplementary Material). Since each slice contains essentially the same signal information as its two-dimensional DCT, one would require approximately as many pixel measurements as the number of non-zero components in this latter representation in order to fully determine it. This well-known principle from information theory underlies all compressed sensing/sparse recovery algorithms. This approach is of immense importance in scenarios where a scant number of physical measurements seemingly fail to mathematically constrain a system. After sparse recovery, the target intensity pattern we ultimately seek is obtained simply by inverting this transform.

To avoid information loss in our ground truth simulation, the Patterson function of the simulated crystal has to be fully captured within the three dimensional simulation array. In other words, the buffer width (populated with zero-valued pixels) around the scatterer should be at least the span of the scatterer in each dimension (this is equivalent to the Nyquist sampling criterion[17]). The simulation sizes described in Section 2 go well above this minimum requirement to ensure that the Patterson function is not only fully captured, but also sparse in the array.

We define the sparse recovery problem in terms of (1) The unknown (target) diffraction slice **I** of the desired fine-pixel resolution, which is sparse in the DCT representation: $\mathbf{I} \equiv B\mathbf{x}$ (where x is sparse and the columns of matrix B are the inverse DCT basis vectors) and (2) a set of measurements made on this signal, represented by a linear operation A resulting in measured (binned) values $\mathbf{I}^{(\text{measured})} = A\mathbf{I} = A(B\mathbf{x})$. We illustrate these operations using the example intensity pattern shown in Fig. 2 that has a relatively small number of pixels. In this example, the coarse pixelation of a fine grid of size $28 \times 28 = 784$ pixels results in a total of $4 \times 4 = 16$ measurements, and therefore A is a 16×784 matrix. Additional measurements are obtained by translating the detector in the plane perpendicular to the exit beam as shown. Each detector offset provides a new set of pixel measurements (fewer than 16 per detector translation, since we ignore the pixels that fall outside the original q range of interest). Each pixel measurement corresponds to a row of the matrix A. As we have proved in Section 1 of the Supplementary Material, it is not possible to obtain 784 independent rows of the matrix A through detector shifts, and we address this problem with compressed sensing. All compressed sensing techniques solve the system of equations $AB\mathbf{x} = \mathbf{I}^{(\text{measured})}$ by enforcing that x be sparse.

While there exist various algorithms for sparse recovery[18–20], we adopt the LASSO regression method[21] common in machine learning applications:

$$\mathbf{x}_{\text{optimal}} = \underset{\mathbf{x}}{\text{argmin}}\left\{ \left|AB\mathbf{x} - \mathbf{I}^{(\text{measured})}\right|^2 + \alpha|\mathbf{x}| \right\}$$

(2)

for some small $\alpha > 0$ (set to 2×10^{-4} throughout this paper). The $|\mathbf{x}|$ penalty imposition on the objective function in Equation (2) explicitly enforces sparsity on the unknown x (ℓ_1-optimisation). Briefly, the optimisation converges to that solution x which has the fewest non-zero components and simultaneously satisfies the set of constraints $AB\mathbf{x} = \mathbf{I}^{(\text{measured})}$. The target image is recovered by inverting the sparsifying transform ($\mathbf{I} = B\mathbf{x}_{\text{optimal}}$). We point out that this recovery scheme does not strictly enforce the constraint of non-negativity on **I**. This results in recovered images with a few negative-valued pixels, which we simply threshold to zero (Fig. 4(g,h)). We show in Section 4 that this thresholding still results in a working approximation of the continuous signal, from a phase retrieval point of view.

With regard to the number of independent coarse pixel measurements M, compressed sensing theory provides a more precise success threshold: for an image of size $N \times N$ pixels, $M \geq K\log(N^2/K)$ where K is the number of significant (nonzero) components in the sparse representation[11]. In general, results of the recovery when M is well above this threshold are negligibly different from the original image itself.
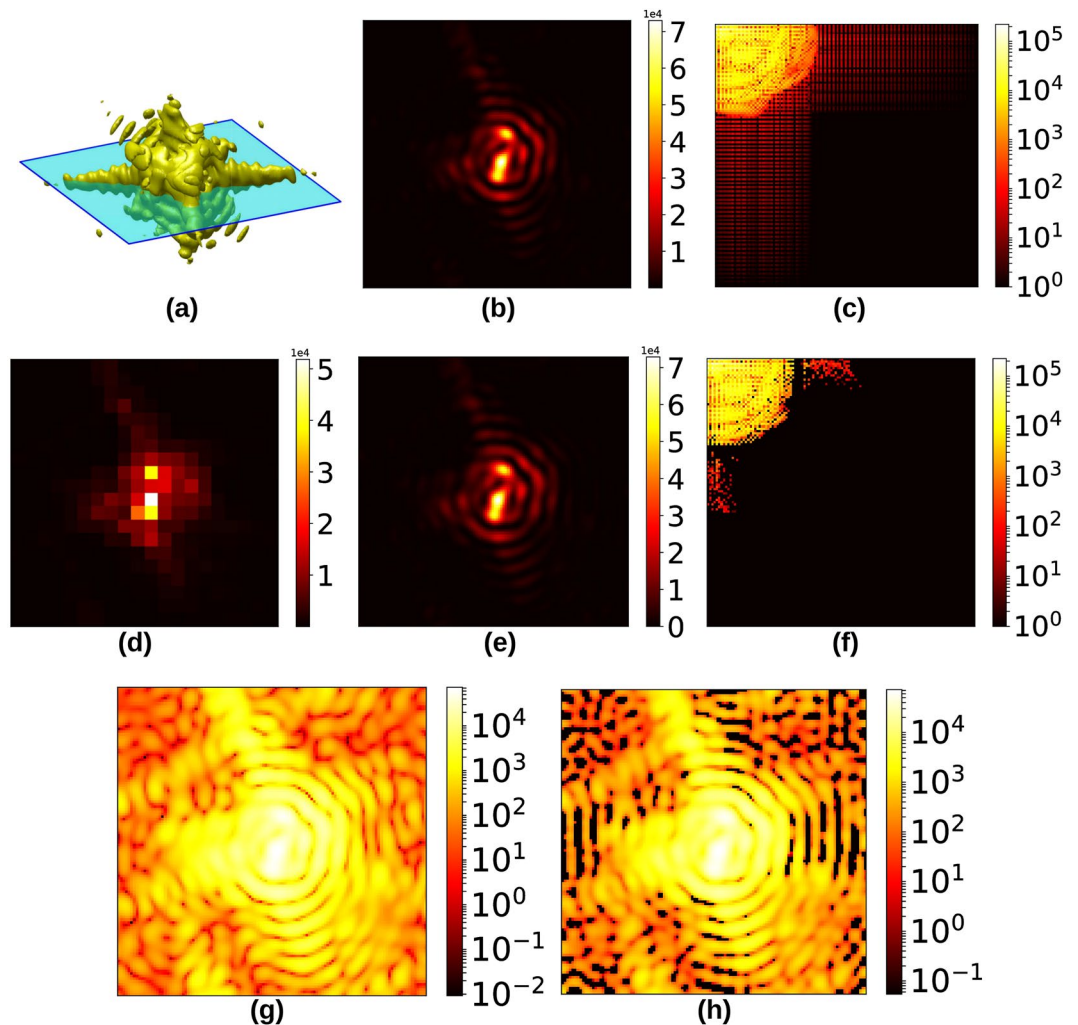
Figure 4. (**a**) Slice of 3-D diffraction signal, (**b**) Slice imaged on detector, nominally 9 keV, (**c**) DCT spectrum magnitude of 9 keV signal, (**d**) Equivalent signal in (b) simulated at 54 keV, (**e**) Recovered image with pixelation equivalent to 9 keV signal, (**f**) DCT spectrum magnitude of recovered signal, (**g**) Benchmark image on a log scale, and (**h**) Recovered image on a log scale, showing the regions where the negative pixels were thresholded to zero.

We note that K depends on the basis of the sparse representation (in our case, the discrete cosine basis). Given the simulated particle described in Section 2, a crude estimate of K for the DCT can be made in the following manner. Since the central slice of the particle is contained within a 22×22-pixel box at the centre of a two-dimensional array, any projection of the Patterson function parallel to this plane is contained in a 44×44-pixel box (from the projection-slice theorem[22]). Since our simulated object has facets, this projection may be simply modeled as a 'diamond'-shaped area whose vertices lie on the midpoints of the sides of the 44×44-pixel square. The 'diamond' has an area of $44^2/2 = 968$ pixels. Therefore the estimated number of non-zero components in this projection (which is the Fourier transform of a single detector measurement) is approximately $2 \times 968 = 1936$, after accounting for comlpex numbers. Given the close relation between the Fourier and cosine transforms, it is reasonable to assume a similar number of components in the cosine basis as well. In other words, $K \simeq 1936$.

Our estimate of $\simeq 1500$ comes from simply counting the number of significant components in Fig. 3(c) after suitable thresholding. In a real experiment it is impossible to estimate K without making a measurement.

Recovery Results

Figure 4 shows a visual comparison between a benchmark ground truth diffraction slice (4(b)), and the corresponding fine-pixel diffraction slice recovered from a set of simulated intensity patterns at 54 keV obtained with sub-pixel detector translation (4(e)). The coarsely binned diffraction slice has size 20×20 pixels (4(d)), while the recovered diffraction slice has size 120×120 (fine) pixels, corresponding to a PBF of 6.

For a better quantitative picture, we examine the fidelity of two different recovered diffraction patterns to their respective 9 keV ground truth benchmark slices, as a function of the beam energy and degree of upsampling. The
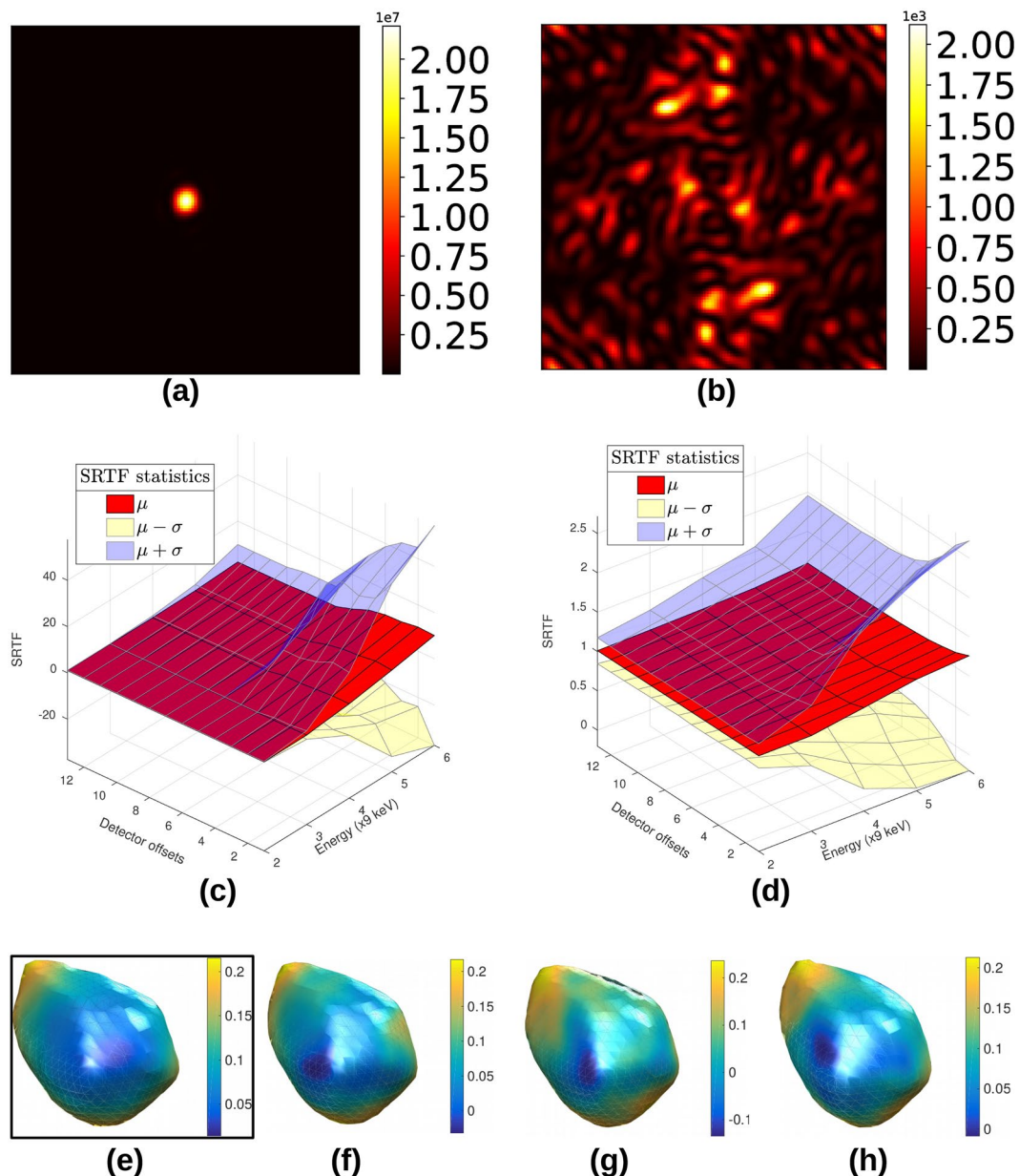
Figure 5. (**a**) Central slice of a 3-D diffraction pattern, (**b**) edge slice of the same 3-D diffraction pattern, (**c**) SRTF trend for central slice, (**d**) SRTF trend for edge slice, (**e**) original synthetic particle with surface phase variation, (**f–h**) reconstructed particle and surface phase variation (in radians) corresponding to energies 36 keV, 45 keV and 54 keV respectively. The phase retrieval recipe used was: solvent flipping (400 iter.) → hybrid input-output ($\beta = 0.8$, 240 iter.) → solvent flipping (400 iter.) → error reduction (100 iter.)[24], with shrinkwrapping every 25 iterations. Section 3 of the Supplementary Material describes the SRTF trend in the presence of simulated Poisson noise.

benchmark images represent qualitatively different intensity distributions: the "on-Bragg" central slice (Fig. 5(a)) has one strong, highly localised peak while the intensity distribution in the "off-Bragg" terminal slice (Fig. 5(b)) is weaker and more spread out. The fidelity is quantified by the sparse recovery transfer function (SRTF), which we define for a single recovered image as

$$\mathrm{SRTF}(i, j) = \sqrt{\frac{I_{\mathrm{recovered}}(i, j)}{I_{\mathrm{groundtruth}}(i, j)}}$$

(3)

where the indices (i, j) run over the pixels of the finely pixelated diffraction pattern. The SRTF is analogous to the phase retrieval transfer function common in phase retrieval literature[23]. For a perfect recovery, SRTF = 1 for all pixels. Our results are expressed in terms of the mean μ of the SRTF and the standard deviation spread $\mu \pm \sigma$ around the mean, evaluated over the recovered (upsampled) images.
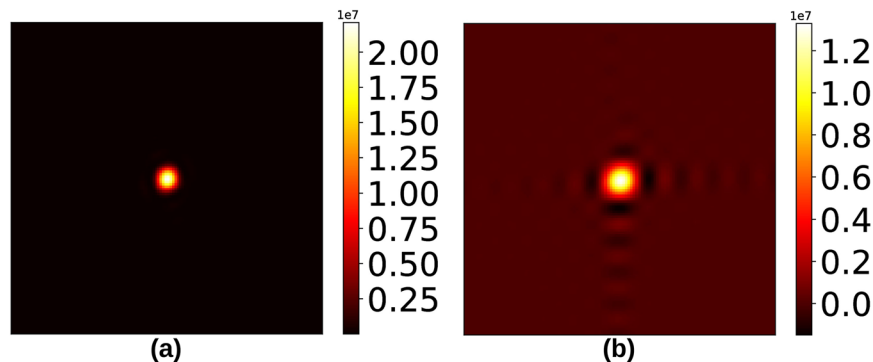
Figure 6. (**a**) Central "on-Bragg" benchmark image, (**b**) Bad recovery at 54 keV from a single detector position, or equivalently $20 \times 20 = 400$ measurements *i.e.* no detector displacement.

Figure 5(c) and (d) indicate a steady improvement in the SRTF with more detector offsets. Sufficient sub-pixel translations of the detector ensure that the SRTF of the recovered image remains in the vicinity of the ideal value of 1. With higher beam energy, there is a greater variance in the SRTF. This is due to the appearance of artefacts from an insufficiently constrained optimisation, as can most prominently be seen in Fig. 6(a) and (b). In Fig. 5(c) and (d), on the other hand, the slight deviation of the SRTF from 1 at high PBF stems from the small penalty $\alpha|\mathbf{x}|$ on the objective function in Equation (2).

Finally in Fig. 5(f–h) we show the results of the standard BCDI phase retrieval applied to diffraction intensity patterns upsampled from simulations of 36, 45 and 54 keV data sets (PBF = 4, 5, 6) with sub-pixel translations. Each upsampled intensity pattern data set was recovered from the original detector position and 12 additional detector offsets along the pixel diagonals. The reconstructions are in agreement with the original structure 5(e). Section 1 of the Supplementary Material describes the rationale for offsets along both pixel diagonals, as opposed to purely horizontal or vertical offsets.

Discussion

We have outlined a framework by which fine features in high-energy coherent diffraction intensity patterns from individual compact crystals can be recovered using compressed sensing and sparse recovery. The success of these methods is often justified in literature by demonstrating recovery with fewer measurements than the size of the actual signal array. Thus far we have also described our methodology in terms of these coarse 'pixel measurements' when quantifying information content in a signal. In this section we translate this phrasing into quantities that are directly related to the design of CDI experiments. Specifically, we show how our sparse recovery technique can be exploited to emulate smaller pixel sizes or larger sample-detector distances. We thereby demonstrate the advantage of putting the burden of signal detection on numerical algorithms, as opposed to substantially modifying an existing experimental setup or constructing massive experimental enclosures to enable large detector distances at large Bragg angles.

Firstly, Table 1a shows the number of pixel measurement constraints in our simulations as a function of the X-ray energy multiplier and the number of detector positions. The maximum number of pixel measurements at each PBF is computed using Equation 2 of the Supplementary Material. The two regions of interest are: (1) the numbers in ()-brackets which fall below the theoretical limit of $M = K\log(N^2/K)$, and (2) the numbers in []-brackets which show that further detector offsets along the diagonals do not give additional binning constraints due to the grid periodicity. Here, we aimed to recover the ground truth within a 120×120 pixel region, such that $N = 120$. The number of unknown sub-pixels to recover is therefore $N^2 = 14400$. Also, $K \simeq 1500$ for the simulated crystal (*i.e.* the number of significant components in the DCT of the central diffraction slice, Fig. 4(f)). The numbers in region (1) result from an inadequate number of detector positions. In this regime, the acquired signal information is truly deficient and sparse recovery is not possible. Region (2) denotes the limitations of the chosen detector translation strategy (in this case, diagonal offsets only). For a given crystal size and detector pixel size, appropriate choices of sample-detector distance and detector offset strategy ensure the existence of the intermediate region between 1 and 2 where compressed sensing can successfully recover the diffracted intensity pattern.

The connection between M and the experimental parameters is made if we imagine that M binning constraints can also be obtained from a *single* detector image of $\sqrt{M} \times \sqrt{M}$ pixels that queries the same region of interest in reciprocal space. Stated differently, the set of offset detector images with a larger pixel size is conceptually equivalent to a single detector image with a smaller pixel size. From an information theoretical point of view, the two sets of constraints are equivalent descriptions of the scattered wave intensity. This immediately suggests a reinterpretation of Table 1a in terms of experimental parameters.

For example, at a beam energy of 54 keV, we have PBF = 6 relative to the same measurement at 9 keV. The original 120×120-pixel q-space region of interest in the ground truth simulation is now squeezed into a 20×20-pixel region in the centre of the detector due to reciprocal space compression. The effective coarse pixel size over the original region of interest is simply $\Delta x = 1/20 = 0.05$ units. But with 7 detector positions at this energy, Table 1a shows that we have access to a total of 2566 pixel measurements, effectively emulating a $\sqrt{2566} \times \sqrt{2566}$ pixel grid over the same region of interest. The effective pixel size is now scaled by a factor of

7

		2	3	4	5	6
		(a) PBF (energy multiplier)				
Number of detector positions	13	[14043]	[7684]	[6787]	[4808]	[4371]
	12	[14043]	[7684]	[6787]	[4808]	4371
	11	[14043]	[7684]	[6787]	[4808]	4010
	10	[14043]	[7684]	[6787]	[4808]	3649
	9	[14043]	[7684]	[6787]	4808	3288
	8	[14043]	[7684]	6787	4279	2927
	7	[14043]	[7684]	5946	3750	2566
	6	[14043]	[7684]	5105	3221	2205
	5	[14043]	7684	4264	2692	1844
	4	14043	6163	3423	2163	1483
	3	10562	4642	2582	1634	(1122)
	2	7081	3121	1741	(1105)	(761)
	1	3600	1600	(900)	(576)	(400)
		(b) Beam energy ($\times 9\,\mathrm{keV}$)				
Number of detector positions	13	[1.975]	[2.191]	[2.746]	[2.889]	[3.306]
	12	[1.975]	[2.191]	[2.746]	[2.889]	3.306
	11	[1.975]	[2.191]	[2.746]	[2.889]	3.166
	10	[1.975]	[2.191]	[2.746]	[2.889]	3.02
	9	[1.975]	[2.191]	[2.746]	2.889	2.867
	8	[1.975]	[2.191]	2.746	2.726	2.705
	7	[1.975]	[2.191]	2.57	2.552	2.533
	6	[1.975]	[2.191]	2.382	2.365	2.348
	5	[1.975]	2.191	2.177	2.162	2.147
	4	1.975	1.963	1.95	1.938	1.925
	3	1.713	1.703	1.694	1.684	(1.675)
	2	1.402	1.397	1.391	(1.385)	(1.379)
	1	1	1	(1)	(1)	(1)

Table 1. (**a**) Number of unique binning constraints M resulting from diagonal detector offsets alone, as a function of PBF. Values of M below the information theoretic limit are shown in ()-brackets. The values in ()-brackets indicate when further detector offsets do not contribute any new binning constraints. (**b**) Effective multiplier for sample-detector distance and its variation with beam energy and number of detector offsets. The parenthesis codes are the same as for (a) For reference, in an earlier BCDI experiment at 9 keV, a 300 nm crystal has been imaged at a sample detector distance of 0.63 m and a pixel size of $55\,\mu$ m[4].

20$\diagup\sqrt{2566}$, or equivalently, we are able to resolve spatial frequencies better by a factor of $\sqrt{2566}\diagup20 = 2.533$, which we label f.

In the Fraunhofer approximation, the quantity f can also be interpreted as an effective increase in sample-detector distance, given a fixed pixel size. The spatial frequency step is related to the pixel size Δx and sample-detector distance z by $\Delta k \propto \Delta x/z$. A scaled pixel size of $\Delta x/f$ implies an effective sample-detector distance of fz. Table 1b shows Table 1a reinterpreted in terms of the distance multiplier fz.

With a fixed beam energy, the limit of f is reached when all possible binning constraints are accounted for and further detector offsets do not give new constraints. We have shown in Section 1 of the Supplementary Material that in this case, the number of pixel measurements that fall fully within the q-space region of interest is $M = [N + (\mathrm{PBF} - 1)(N - 1)]^2$. Figure 7 shows the maximum values of f under this condition, for a wider range of beam energies. We see that maximum upsampling nearly undoes the effect of pixel binning at high energies, through a proportionate increase of f, effectively emulating the sample-detector distance that would otherwise be needed to resolve intensity fringes. The difference is accounted for by the fact that pixels that are not fully contained within the original q-range are not used as measurement constraints.

Summary

We have described a signal recovery technique for single-crystal BCDI data sets acquired at X-ray energies typically suited to HEDM applications (>50 keV). Our methodology relies on a modification of the conventional BCDI setup in which additional data is acquired by translating the detector perpendicular to the diffracted beam. We have described a fundamental incompleteness in the data collected by this method of upsampling, which calls for the incorporation of some form of extra information about the signal. With this necessity we have motivated the use of techniques that recover not the signal directly, but sparse representations of it. We have based our methodology on a rigorous proof of the existence of such a representation for all coherent scattering from a fully-illuminated compact single crystal. We have quantified image recovery by this compressed sensing process
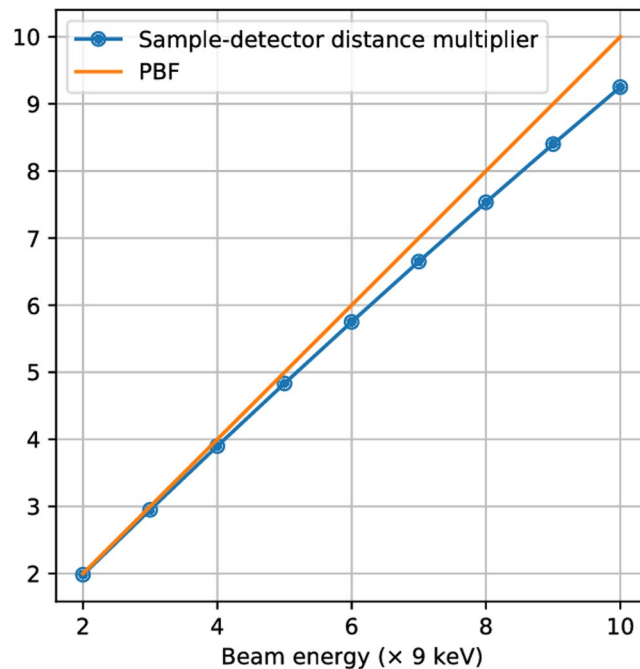
Figure 7. Behaviour of maximum value of sample-detector distance multiplier *f* for different beam energies. For a fixed pixel size, the distance multiplier asymptotically approaches the PBF in the limit of large reciprocal space aperture (a wider detector).

and shown subsequent phase retrieval on the recovered data sets that are in agreement with the original simulated crystal.

We note that the upsampling of coarsely pixelated data in some transformed form is in general a valuable alternative to direct measurement for many high-energy diffraction experiments. The subsequent signal recovery is readily achieved by identifying an appropriate sparse basis for a given measurement.

Information theory- based analysis of detector-space upsampling shows us that smaller pixels and larger sample-detector distances can be emulated, with minimal changes to the experimental setup. This capability has the potential to significantly influence the design of space-constrained experiments at high energy coherent scattering beamlines. When incorporated into existing near- and far-field HEDM workflows, high-energy BCDI could be the nanoscale component of a generalised experiment for imaging and characterisation of polycrystal-line samples at multiple length scales. Coupled with the development of new computational methodologies, such multiscale characterisation capabilities could go a long way in validating existing and new models of materials physics, as well as informing the creation and processing of engineering materials.

Data availability. The ground truth simulation and the Python code for sparse recovery are available upon reasonable request. Please contact the corresponding author.

References
1. Robinson, I. K., Vartanyants, I. A., Williams, G. J., Pfeifer, M. A. & Pitney, J. A. Reconstruction of the shapes of gold nanocrystals using coherent x-ray diffraction. *Phys. Rev. Lett.* **87**, 195505, https://doi.org/10.1103/PhysRevLett.87.195505 (2001).
2. Robinson, I. & Harder, R. Coherent x-ray diffraction imaging of strain at the nanoscale. *Nat Mater* **8**, 291–298, https://doi.org/10.1038/nmat2400 (2009).
3. Ulvestad, A. *et al*. Topological defect dynamics in operando battery nanoparticles. *Science* **348**, 1344–1347 (2015).
4. Cha, W. *et al*. Three dimensional variable-wavelength x-ray bragg coherent diffraction imaging. *Phys. Rev. Lett.* **117**, 225501, https://doi.org/10.1103/PhysRevLett.117.225501 (2016).
5. Suter, R. M. *et al*. Probing microstructure dynamics with x-ray diffraction microscopy. *Journal of Engineering Materials and Technology* **130**, 021007–021007-5, https://doi.org/10.1115/1.2840965 (2008).
6. Bernier, J. V., Barton, N. R., Lienert, U. & Miller, M. P. Far-field high-energy diffraction microscopy: a tool for intergranular orientation and strain analysis. *The Journal of Strain Analysis for Engineering Design* **46**, 527–547, https://doi.org/10.1177/0309324711405761 (2011).
7. Ludwig, W., Schmidt, S., Lauridsen, E. M. & Poulsen, H. F. X-ray diffraction contrast tomography: a novel technique for three-dimensional grain mapping of polycrystals. I. Direct beam case. *Journal of Applied Crystallography* **41**, 302–309, https://doi.org/10.1107/S0021889808001684 (2008).
8. Liu, Q. *et al*. Quantifying the nucleation and growth kinetics of microwave nanochemistry enabled by *in situ* high-energy x-ray scattering. *Nano letters* **16**, 715–720 (2015).
9. Chushkin, Y. & Zontone, F. Upsampling speckle patterns for coherent X-ray diffraction imaging. *Journal of Applied Crystallography* **46**, 319–323, https://doi.org/10.1107/S0021889813003117 (2013).
10. Candes, E. J., Romberg, J. & Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* **52**, 489–509 (2006).
11. Donoho, D. L. Compressed sensing. *IEEE Transactions on Information Theory* **52**, 1289–1306 (2006).

12. Goodman, J. *Introduction to Fourier Optics*. McGraw-Hill physical and quantum electronics series, https://books.google.com/books?id=ow5xs_Rtt9AC (W. H. Freeman, 2005).
13. Batey, D. J. *et al*. Reciprocal-space up-sampling from real-space oversampling in x-ray ptychography. *Phys. Rev. A* **89**, 043812, https://doi.org/10.1103/PhysRevA.89.043812 (2014).
14. Makhoul, J. A fast cosine transform in one and two dimensions. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**, 27–34 (1980).
15. Wallace, G. K. The jpeg still picture compression standard. *Commun. ACM* **34**, 30–44, https://doi.org/10.1145/103085.103089 (1991).
16. Musmann, H. G. Genesis of the mp3 audio coding standard. *IEEE Transactions on Consumer Electronics* **52**, 1043–1049 (2006).
17. Shannon, C. E. Communication in the presence of noise. *Proceedings of the IRE* **37**, 10–21 (1949).
18. Suykens, J. A., Lukas, L. & Vandewalle, J. Sparse approximation using least squares support vector machines. In *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, vol. 2, 757–760 (IEEE, 2000).
19. Huang, J., Ma, S. & Zhang, C.-H. Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica* 1603–1618 (2008).
20. Candès, E. J., Eldar, Y. C., Needell, D. & Randall, P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis* **31**, 59–73, http://www.sciencedirect.com/science/article/pii/S1063520310001156 (2011).
21. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B* (*Methodological*) 267–288 (1996).
22. Bracewell, R. N. Numerical transforms. *Science* **248**, 697–704, http://science.sciencemag.org/content/248/4956/697 (1990).
23. Chapman, H. N. *et al*. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *J. Opt. Soc. Am. A* **23**, 1179–1200, http://josaa.osa.org/abstract.cfm?URI=josaa-23-5-1179 (2006).
24. Marchesini, S. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments* **78**, 011301, https://doi.org/10.1063/1.2403783 (2007).

Acknowledgements

Author Contributions

The analysis was developed by S.M. and implemented by S.M. with help from Y.N. and I.C.A. S.M. wrote the manuscript with contributions from S.H., Y.N., R.H. and J.S.P. All authors contributed to the refining of the concepts presented.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-23040-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.