On asymptotic structure in compressed sensing

Bogdan Roman *, **Ben Adcock** † **and Anders Hansen** *

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK, and †Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

This paper demonstrates how new principles of compressed sensing, namely asymptotic incoherence, asymptotic sparsity and multilevel sampling, can be utilised to better understand underlying phenomena in practical compressed sensing and improve results in real-world applications. The contribution of the paper is fourfold: First, it explains how the sampling strategy depends not only on the signal sparsity but also on its structure, and shows how to design effective sampling strategies utilising this. Second, it demonstrates that the optimal sampling strategy and the efficiency of compressed sensing also depends on the resolution of the problem, and shows how this phenomenon markedly affects compressed sensing results and how to exploit it. Third, as the new framework also fits analog (infinite dimensional) models that govern many inverse problems in practice, the paper describes how it can be used to yield substantial improvements. Fourth, by using multilevel sampling, which exploits the structure of the signal, the paper explains how one can outperform random Gaussian/Bernoulli sampling even when the classical l^1 recovery algorithm is replaced by modified algorithms which aim to exploit structure such as model based or Bayesian compressed sensing or approximate message passaging. This final observation raises the question whether universality is desirable even when such matrices are applicable. Examples of practical applications investigated in this paper include Magnetic Resonance Imaging (MRI), Electron Microscopy (EM), Compressive Imaging (CI) and Fluorescence Microscopy (FM). For the latter, a new compressed sensing approach is also presented.

Compressed sensing (CS), introduced by Candès, Romberg & Tao (1) and Donoho (2), states that under appropriate conditions one can overcome the Nyquist sampling barrier and recover signals using far fewer samples than dictated by the classical Shannon theory. This has important implications in many practical applications which caused CS to be intensely researched in the past decade.

CS problems can be divided into two types. **Type I** are problems where the physical device imposes the sampling operator, but allows some limited freedom to design the sampling strategy. This category is vast, with examples including Magnetic Resonance Imaging (MRI), Electron Microscopy (EM), Computerized Tomography, Seismic Tomography and Radio Interferometry. **Type II** are problems where the sensing mechanism offers freedom to design both the sampling operator and the strategy. Examples include Fluorescence Microscopy (FM) and Compressive Imaging (CI) (e.g. single pixel and lensless cameras). In these two examples, many practical setups still impose some restrictions regarding the sampling operator, e.g. measurements must typically be binary.

Traditional CS is based on three pillars: *sparsity* (there are s important coefficients in the vector to be recovered, however, the location is arbitrary), *incoherence* (the values in the measurements matrix should be uniformly spread out) and *uniform random subsampling*.

For Type I problems the issue is that the above pillars are often lacking. As we will argue, many Type I problems are coherent due to the physics or because they are infinite-dimensional. The traditional CS framework is simply not applicable. However, CS was used successfully in many Type I problems, though with very different sampling techniques than uniform random subsampling, which lack a mathematical justification.

For Type II problems the traditional CS theory is applicable, e.g. in CI one can use random Bernoulli matrices. The issue is that the use of complete randomness does not allow one to exploit the structure

of the signal to be recovered from a sampling point of view. As we argue, real world signals are not sparse, but asymptotically sparse in frames such as wavelets(3, 4) or their *-let cousins such as curvelets (5), contourlets (6) or shearlets (7). In particular, the asymptotic sparsity is highly structured.

New CS principles. To bridge the gap between theory and practice, the authors introduced a new CS theory (8) that replaces the traditional CS pillars with three new CS principles: *asymptotic incoherence*, *asymptotic sparsity* and *multilevel sampling*. The new theory and principles reveal that the optimal sampling strategy and benefits of CS depend on two factors: the structure of the signal and the resolution. This suggests a new understanding of the underlying phenomena and of how to improve CS results in practical applications, which are main topics of this paper. At the same time, this paper demonstrates how the new CS principles go hand in hand even with applications where traditional CS is applicable, and that substantial improvements and flexibility can be obtained.

Traditional Compressed Sensing

A traditional CS setup is as follows. The aim is to recover a signal f from an incomplete (subsampled) set of measurements y. Here, f is represented as a vector in $\mathbb{C}^N$ and is assumed to be s-sparse in some orthonormal basis $\Phi \in \mathbb{C}^{N \times N}$ (e.g. wavelets) called *sparsity* basis. This means that its vector of coefficients $x = \Phi f$ has at most s nonzero entries. Let $\Psi \in \mathbb{C}^{N \times N}$ be an orthonormal basis, called *sensing* or *sampling* basis, and write $U = \Psi\Phi^* = (u_{ij})$, which is an isometry. The coherence of U is

$$\mu(U) = \max_{i,j} |u_{ij}|^2 \in [1/N, 1]. \qquad \textbf{[1]}$$

and U is said to be perfectly incoherent if $\mu(U) = 1/N$.

Let the *subsampling pattern* be the set $\Omega \subseteq \{1, \ldots, N\}$ of cardinality m with its elements chosen uniformly at random. Owing to a result by Candès & Plan (9) and Adcock & Hansen (10), if we have access to the subset of measurements $y = P_\Omega \Psi f$ then f can be
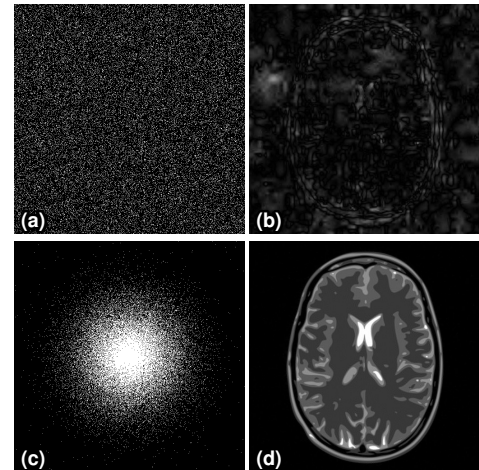

Fig. 1. (a) 12.5% uniform random subsampling scheme, (b) CS reconstruction from uniform subsampling, (c) 12.5% multilevel subsampling scheme, (d) CS reconstruction from multilevel subsampling.

recovered from y exactly with probability at least $1 - \epsilon$ if

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot (1 + \log(1/\epsilon)) \cdot \log(N), \qquad \textbf{[2]}$$

where $P_\Omega \in \{0, 1\}^{N \times N}$ is the diagonal projection matrix with the jth entry 1 if $j \in \Omega$ and 0 otherwise, and the notation $a \gtrsim b$ means that $a \geq C b$ where $C > 0$ is some constant independent of a and b. Then, f is recovered by solving

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \quad \text{subject to} \quad \|y - P_\Omega U z\| \leq \eta. \qquad \textbf{[3]}$$

where η is chosen according to the noise level (0 if noiseless). The key estimate [Eq. 2] shows that the number of measurements m required is, up to a log factor, on the order of the sparsity s, provided

CS recovery **CS flipped recovery** **Subsampling map**

Fluorescence Microscopy (FM), $U = \Psi_{\text{Had}} \Phi_{\text{dwt}}^*$ 512×512 @ 6.25%

Compressive Imaging (CI), $U = \Psi_{\text{Had}} \Phi_{\text{dwt}}^*$ 256×256 @ 12.5%

MRI, $U = \Psi_{\text{dft}} \Phi_{\text{dwt}}^*$ 1024×1024 @ 20%

Tomography, Electron Microscopy (FM), $U = \Psi_{\text{dft}} \Phi_{\text{dwt}}^*$ 512×512 @ 12.5%

Radio Interferometry, $U = \Psi_{\text{dft}} \Phi_{\text{dwt}}^*$ 512×512 @ 15%

Fig. 2. Flip test. Recovery from direct versus flipped wavelet coefficients showing that the RIP does not hold in these cases. The percentage shown is the subsampled fraction of Fourier/Hadamard coefficients.

the coherence $\mu(U) = \mathcal{O}(1/N)$. This is the case, for example, when U is the DFT, which was studied in some of the first CS papers (1).

The real world is often coherent. Consider the MRI CS setup, i.e. $U = \Psi_{\text{dft}} \Phi_{\text{dwt}}^* \in \mathbb{C}^{N \times N}$, where Ψ_{dft} and Φ_{dwt} are the discrete Fourier and wavelet transforms (sampling and sparsity bases) respectively. The coherence here is

$$\mu(U) = \mathcal{O}(1), \quad N \to \infty,$$

for any wavelet basis, so this problem has the worst possible coherence. The traditional CS bound [Eq. 2] states that all samples are needed in this case (i.e. full sampling, $m = N$), even though the orignal signal is typically highly sparse in wavelets. This lack of incoherence means that uniform random subsampling leads to a very poor recovery. This is known in MRI and is illustrated in Fig. 1.

The root cause of this lack of incoherence is the discretization of what is intrinsically an infinite-dimensional problem into a finite-dimensional one. In short, U converges to an infinite matrix (8) and since the incoherence is the supremum of its entries, there exists some N for which a *coherence barrier* is hit, resulting in the worst case for a CS recovery. This is not restricted to MRI. Any such discretization of an infinite-dimensional problem will suffer the same fate, including MRI, tomography, microscopy, seismology, radio interferometry etc. Changing Ψ may provide marginal benefits, if any, since the coherence barrier always occurs at some N.

Sparsity, flip test and the absence of RIP. Traditional CS states that the sampling strategy is completely independent of the location of the nonzero coefficients of an s-sparse vector x, i.e. with the s nonzero coefficients at arbitrary locations. The *flip test* allows one to evaluate whether this holds in practice. Let $x \in \mathbb{C}^N$ be a vector, and $U \in \mathbb{C}^{N \times N}$ a measurement matrix. We then sample according to some pattern $\Omega \subseteq \{1, \ldots, N\}$ with $|\Omega| = m$ and solve [Eq. 3] for x, i.e. $\min \|z\|_1$ s.t $P_\Omega U z = P_\Omega U x$ to obtain a reconstruction $z = \alpha$. Now we flip x to obtain a vector x' with reverse entries, $x_i' = x_{N-i}, i = 1, \ldots, N$ and solve [Eq. 3] for x' using the same U and Ω, i.e. $\min \|z\|_1$ s.t. $P_\Omega U z = P_\Omega U x'$. Assuming z to be a solution, then by flipping z we obtain a second reconstruction α' of the original vector x, where $\alpha_i' = z_{N-i}$.

Assume Ω is a sampling pattern for recovering x using α. If sparsity alone dictates the reconstruction quality, then α' must yield the same reconstruction quality (since x' has the same sparsity as x, being merely a permutation of x). Is this true in practice?

Fig. 2 investigates this for several applications using $U = \Psi_{\text{dft}} \Phi_{\text{dwt}}^*$ or $U = \Psi_{\text{Had}} \Phi_{\text{dwt}}^*$, where $\Psi_{\text{dft}}, \Psi_{\text{Had}}, \Phi_{\text{dwt}}$ are the discrete Hadamard, Fourier and wavelet transforms respectively. As is evident, the flipped recovery α' is substantially worse than its un-flipped version α. This confirms that sparsity alone does not dictate the reconstruction quality. Furthermore, note that $P_\Omega U$ cannot satisfy an RIP for realistic values of N, m and s. Had this been the case, both vectors would have been recovered with the same error, and this is in direct contradiction with the results of the flip test.

It is worth noting that the same phenomenon exists for total variation (TV) minimization. Briefly, the CS TV problem is $\min_{z \in \mathbb{C}^n} \|z\|_{TV}$ s.t. $\|y - P_\Omega \Psi z\| \leq \eta$, where the TV norm $\|x\|_{TV}$ in case of images is the ℓ^1 norm of the image gradient, $\|x\|_{TV} = \sum_{i,j} \|\nabla x(i,j)\|_2$ with $\nabla x(i,j) = \{D_1 x(i,j), D_2 x(i,j)\}$, $D_1 x(i,j) = x(i+1,j) - x(i,j)$, $D_2 x(i,j) = x(i,j+1) - x(i,j)$. Fig. 3 shows an experiment where we we chose an image $x \in [0,1]^{N \times N}$ and then built an image x' from the gradient of x so that $\{\|\nabla x'(i,j)\|_2\}$ is a permutation of $\{\|\nabla x(i,j)\|_2\}$ for which $x' \in [0,1]^{N \times N}$. Thus, the two images have the same "TV sparsity" and the same TV norm. It is evident how the reconstruction errors differ substantially for the two images when using the same sampling pattern, confirming that sparsity structure matters for TV recovery as well.
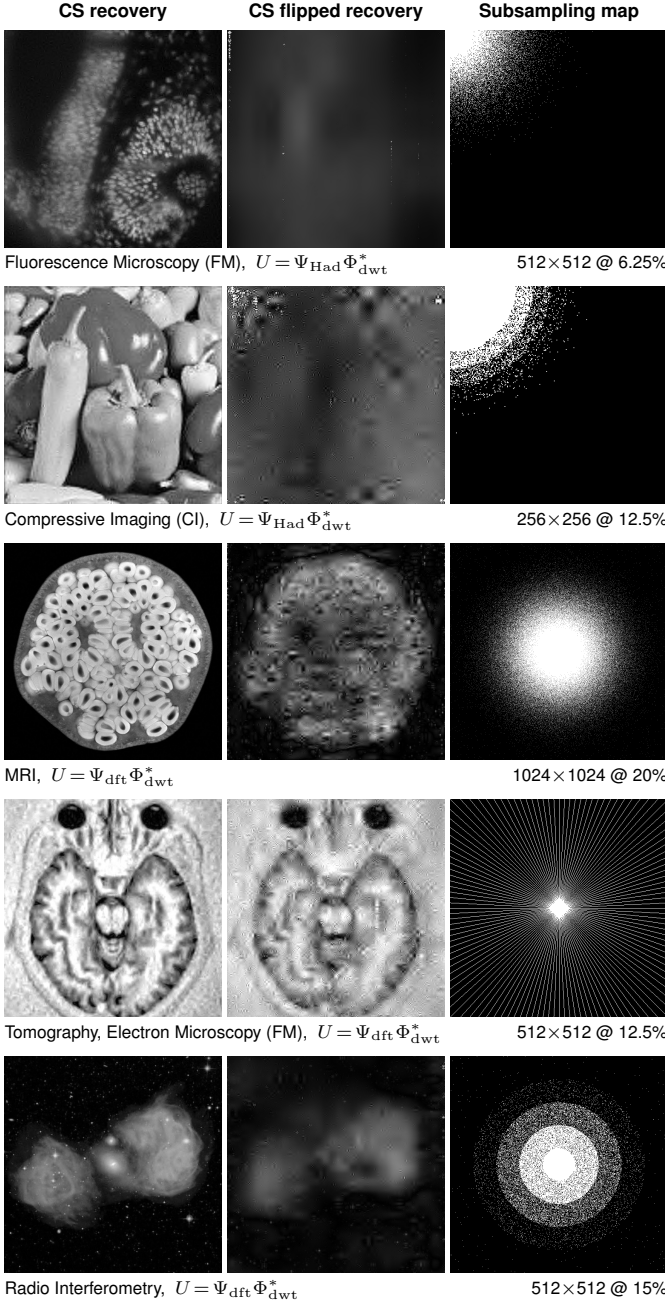
New Compressed Sensing Principles

The previous discussion on traditional CS calls for a more general approach. We consider the generalization of the traditional principles of sparsity, incoherence, uniform random subsampling into *asymptotic sparsity*, *asymptotic incoherence* and *multilevel subsampling* (8).

Asymptotic sparsity. We saw that signal structure is essential, but what structure describes such sparse signals? Let us consider a wavelet basis $\{\varphi_n\}_{n \in \mathbb{N}}$. Recall that there exists a decomposition of $\mathbb{N}$ into finite subsets according to different wavelet scales, i.e. $\mathbb{N} = \bigcup_{k \in \mathbb{N}} \mathcal{M}_k$, where $\mathcal{M}_k = \{M_{k-1} + 1, \ldots, M_k\}$ is the set of indices corresponding to the kth scale, with $0 = M_0 < M_1 < M_2 < \ldots$. Let $x \in l^2(\mathbb{N})$ be the coefficients of a function f in this basis. Suppose that $\epsilon \in (0, 1]$ is given, and define the *local sparsities*

$$s_k = s_k(\epsilon) = \min\left\{ L : \left\| \sum_{i \in \mathcal{M}_{k,L}} x_i \varphi_i \right\| \geq \epsilon \left\| \sum_{i \in \mathcal{M}_k} x_i \varphi_i \right\| \right\}, \quad [\textbf{4}]$$

where $\mathcal{M}_{k,L} \subseteq \mathcal{M}_k$ is the set of indices of the largest L coefficients at the kth scale, i.e. $|x_l| \geq |x_j|, \forall l \in \mathcal{M}_{k,L}, \forall j \in \mathcal{M}_k \setminus \mathcal{M}_{k,L}$. In order words, s_k is the effective sparsity of the wavelet coefficients of f at the kth scale.

Sparsity of x means that for a given large scale $r \in \mathbb{N}$, the ratio $s/M_r \ll 1$, where $M = M_r$ and $s = \sum_{k=1}^{r} s_k$ is the total sparsity of x. However, Fig. 4 shows that besides being sparse, practical signals have more structure, namely *asymptotic sparsity*, i.e.

$$s_k(\epsilon)/(M_k - M_{k-1}) \to 0, \quad [\textbf{5}]$$

rapidly as $k \to \infty$, $\forall \epsilon \in (0, 1]$: they are far sparser at fine scales (large k) than at coarse scales (small k). This also holds for other function systems such as curvelets (5), contourlets (6) or shearlets (7).

Given the structure of modern function systems such as wavelets and their generalizations, we propose the notion of sparsity in levels:

Definition 1. *Let $x \in \mathbb{C}$. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that x is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \ldots, r$, the sparsity band*

$$\Delta_k := \text{supp}(x) \cap \{M_{k-1} + 1, \ldots, M_k\},$$

satisfies $|\Delta_k| \leq s_k$. We denote the set of $(\mathbf{s}, \mathbf{M})$-sparse vectors by $\Sigma_{\mathbf{s}, \mathbf{M}}$.

Asymptotic incoherence. In contrast with random matrices (e.g. Gaussian or Bernoulli), many sampling and sparsifying operators typically found in practice yield fully coherent problems, such as the Fourier with wavelets case discussed earlier. Fig. 5 shows the absolute values of the entries of the matrix U for three examples. Although there are large values of U in all three case (since U is coherent as per [Eq. 1]), these are isolated to a leading submatrix. Values get asymptotically smaller once we move away from this region.

Definition 2. *Let $\{U_N\}$ be a sequence of isometries with $U_N \in \mathbb{C}^N$. $\{U_N\}$ is asymptotically incoherent if $\mu(P_K^\perp U_N)$, $\mu(U_N P_K^\perp) \to 0$, when $K \to \infty$, with $N/K = c, \forall c \geq 1$. Here P_K is the projection onto $\text{span}\{e_j : j = 1, \ldots, K\}$, where $\{e_j\}$ is the $\mathbb{C}^N$ canonical basis, and $P_K^\perp$ is its orthogonal complement.*

In brief, U is asymptotically incoherent if the coherences of the matrices formed by removing either the first K rows or columns of U are small. As Fig. 5 shows, Fourier/wavelets, discrete cosine/wavelets and Hadamard/wavelets are examples of asymptotically incoherent problems.

Multilevel sampling. Asymptotic incoherence calls for a different strategy than uniform random sampling. High coherence in the first few rows of U means that important information about the signal to
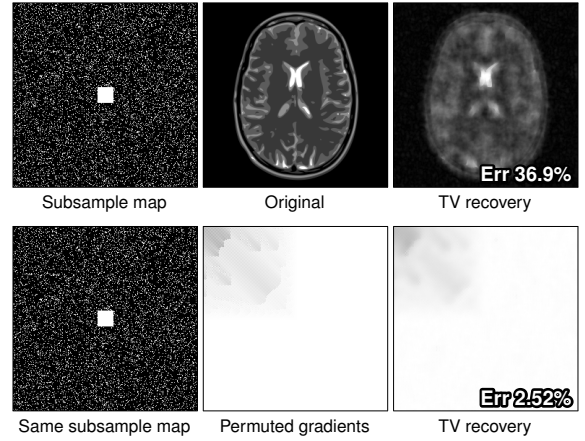

Fig. 3. TV flip test. TV recovery at 256×256 from 8192 DFT samples (12.5% subsampling). The *Permuted gradients* image was built from the gradient vectors of the *Original* image, having the same TV norm and gradient sparsity, differing only in the ordering and sign of the gradient vectors. The large error difference confirms that sparsity structure matters for TV recovery as well.
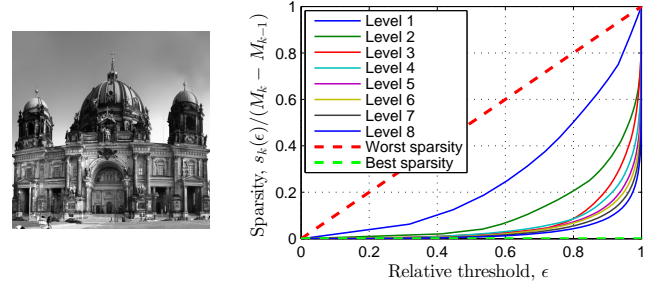

Fig. 4. Sparsity of Daubechies-8 coefficients of an image. The levels correspond to wavelet scales and $s_k(\epsilon)$ is given by [Eq. 4]. Each curve shows the relative sparsity at level k as a function of ϵ. The decreasing nature of the curves for increasing k confirms asymptotic sparsity [Eq. 5].
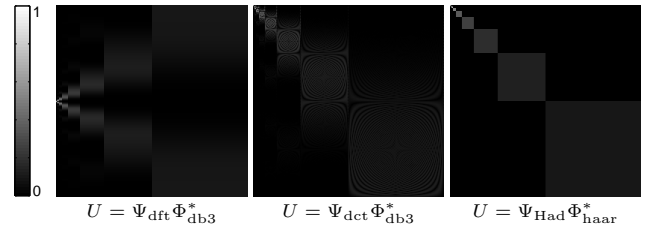

Fig. 5. Visualizing incoherence. The absolute values of the matrix U.

be recovered is likely to be contained in the corresponding measurements, and thus we should fully sample these rows. Once outside this region, as coherence starts decreasing, we can subsample gradually.

Definition 3. *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that $\Omega_k \subseteq \{N_{k-1} + 1, \ldots, N_k\}, |\Omega_k| = m_k$, are chosen uniformly at random, where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N}, \mathbf{m}} = \bigcup_{k=1}^{r} \Omega_k$ as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme (using r levels).*

Briefly, for a vector x, the sampling amount m_k needed in each sampling band Ω_k is determined by the sparsity of x in the corresponding sparsity band Δ_k and the asymptotic coherence $\mu(P_{N_k}^\perp U)$.

Designing a multilevel sampling scheme

Let U be an isometry. The $(k,l)^{\text{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ is given by

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)},$$

where $k,l = 1, \ldots, r$ and P_b^a is the projection matrix corresponding to indices $\{a+1, \ldots, b\}$. The kth relative sparsity is

$$S_k = S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{z \in \Theta} \|P_{N_k}^{N_{k-1}} U z\|^2,$$

where $\Theta = \{z : \|z\|_\infty \leq 1, |\text{supp}(P_{M_l}^{M_{l-1}} z)| = s_l, l = 1, \ldots, r\}$.

Sampling in levels. From (8) we know that to recover an $(\mathbf{s}, \mathbf{M})$-sparse vector $x \in \mathbb{C}^N$ from multilevel sampled measurements $P_\Omega U x$, it suffices that the number of samples m_k in each level satisfies

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \left(\sum_{l=1}^r \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l \right) \cdot \log(N), \qquad \textbf{[6]}$$

where $m_k \gtrsim \hat{m}_k \cdot \log(N)$, and $\hat{m}_k$ is such that

$$1 \gtrsim \sum_{k=1}^r \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k, \qquad \textbf{[7]}$$

$\forall l = 1, \ldots, r$ and $\forall \tilde{s}_1, \ldots, \tilde{s}_r \in (0, \infty)$ for which $\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r$ and $\tilde{s}_k \leq S_k(\mathbf{N}, \mathbf{M}, \mathbf{s})$.

The bounds [Eq. 6] and [Eq. 7] are key. As opposed to the traditional CS bound [Eq. 2], which relates the total amount of subsampling m to the global coherence and the global sparsity, these new bounds relate the local sampling amounts m_k to the local coherences $\mu_{\mathbf{N},\mathbf{M}}(k,l)$ and local and relative sparsities s_k and S_k. A direct result is that the theorem agrees with the flip test shown earlier: the optimal sampling strategy must indeed depend on the signal structure. Another important note is that the bounds [Eq. 6] and [Eq. 7] are sharp in the sense that they reduce to the information-theoretic limits in a number of important cases. Furthermore, in the case of Fourier sampling with wavelet sparsity, they provide near-optimal recovery guarantees using the infinite-dimensional generalization of the theorem. For further details and proofs, the reader is referred to (8).

We devised a flexible *all-round* multilevel sampling scheme. Assuming the coefficients $f \in \mathbb{C}^{N \times N}$ of a sampling orthobasis, e.g. DFT, our multilevel sampling scheme divides f into n regions delimited by $n-1$ equispaced concentric circles plus the full square. Normalizing the support of f to $[-1,1]^2$, the circles have radius r_k with $k = 0, \ldots, n-1$, which are given by $r_0 = m$ and $r_k = k \cdot \frac{1-m}{n-1}$ for $k > 0$, where $m \in (0,1)$ is a parameter. In each of the n regions, the fraction of coefficients sampled with uniform probability is

$$p_k = \exp\left(-(b\,k/n)^a \right), \qquad \textbf{[8]}$$

where $k = 0, \ldots, n$ and $a > 0$ and $b > 0$ are parameters. The total fraction of subsampled coefficients is $p = \sum_k p_k A_k$, where A_k is the normalized area of the kth region. Since $p_0 = 1$ and $p_k > p_{k+1}$, the first region will sample all coefficients and the remaining regions will sample a decreasing fraction. An example is shown later in 9.

Effects and benefits of the new principles

Having reviewed the theory, we now discuss the important effects and benefits of asymptotic incoherence and asymptotic sparsity, and of exploiting them via multilevel sampling. We show how they allow one to improve the CS recovery, and take practical examples from FM, MRI, EM and CI. We begin here with a short summary of the effects and benefits, which are detailed in subsequent sections.

The optimal sampling strategy is signal dependent. As the flip test shows, the optimal sampling strategy depends on the structure of the signal. Multilevel sampling takes this into account and allows one to further improve the CS recovery by tailoring the sampling according to e.g. the resolution and expected wavelet structure of the signal. This has additional advantages as one can mitigate application-specific hurdles or target application-specific features (e.g. brain and spine imaging would use different subsampling schemes). This applies to both Type I and Type II problems. The *FM* and *Resolution dependency* sections below provide examples from FM and MRI.

Resolution dependency. An important effect is that regardless of the sampling basis and subsampling scheme, the quality of the reconstruction increases as resolution increases. This is first revealed by fixing the subsampling strategy and fraction across resolutions, and a more striking result is obtained by fixing the number (instead of fraction) of samples, revealing hidden details, previously inaccessible. This is due to signals being typically increasingly (asymptotically) sparse at higher resolutions. The *FM* and *Resolution dependency* sections show this phenomena with examples from FM and MRI.

Infinite dimensional CS. The new theory provides a good fit to some real-world problems that are fundamentally continuous, e.g. EM or MRI. The errors arising from recovering the continuous samples using discrete models are sometimes significant (11). The section *Infinite dimensional problems* discusses this aspect and shows an EM example where such large errors can be overcome by using generalized sampling theory (8) and recovery into boundary wavelets.

Structured sampling vs Structured Recovery. We exploit sparsity structure by using multilevel sampling of asymptotically incoherent matrices and standard ℓ^1 minimization algorithms. Alternatively, sparsity structure can be exploited by using universal sampling matrices (e.g. random Gaussian/Bernoulli) and modified recovery algorithms. The section *Structured sampling vs Structured Recovery* discusses and compares the two approaches, highlighting the advantages of the former, which, in contrast with the latter, allows to choose the sparsity frame, is applicable to both Type I and Type II problems, and yields overall superior results.

Structure vs Universality: Asymptotic vs Uniform incoherence. The universality property of random sensing matrices (e.g. Gaussian, Bernoulli), explained later on, is a reason for their popularity in traditional CS. But is universality desirable when the signal sparsity is structured? Should one use universal matrices when there is freedom to choose the sampling operator, i.e. in Type II problems? Random matrices are largely inapplicable in Type I problems, where the sampling operator is imposed. The *Structure vs Universality* section argues that universal matrices offer little room to exploit extra structure the signal may have, even in Type II problems, and that non-universal matrices, such as Hadamard, coupled with multilevel sampling provide a better solution for both Type I and Type II problems as they can exploit the prevalent asymptotic sparsity of signals in practice.

Storage and speed. Random matrices, popular in traditional CS, besides being inapplicable in many applications, are also slow and require (large) storage. This yields slow recovery and limits the maximum signal size, which severely affects computations and, more importantly, sparsity structure. The *Storage/speed* section discusses this aspect and also shows that simply addressing the speed and storage problems via fast transforms and non-random matrices is not sufficient to achieve improved recovery compared to what multilevel sampling of non-universal matrices can offer.

Frames and TV. Although investigating frames or TV as sparsity systems in CS is not the purpose of this paper, we provide results in the *Frames and TV* section below, which are an experimental verification of the advantages offered by various frames in the CS context. More importantly, they show the added benefit of incorporating signal structure in the sampling procedure, which provides ample freedom to choose the sparsifying system for a CS recovery.

A new CS approach in Fluorescence Microscopy

We start with a FM example which encompasses many of the previously enumerated effects and benefits and shows how they allow to improve performance in a CS setting. The subsequent sections provide focused detailed discussions.

Compressive Fluorescence Microscopy (CFM), a Type II problem, where the sampling operator can be chosen, was first introduced and implemented practically by Studer et al. (12) and we refer the reader to their work for details. In short, a digital micromirror device of $N \times N$ mirrors can form any $N \times N$ pattern of 0's and 1's to project multiple parallel laser beams (the 1's) through a lens onto a specimen, which is excited and emits light (the fluorescence) of a different wavelength, collected and summed by a photodetector, thus taking one CS measurement. Successive measurements are taken, changing the $N \times N$ pattern on the mirrors each time.

Initial approach. Studer et al. used a 2D Hadamard matrix, i.e. each pattern on the mirrors corresponds to one row in the 2D Hadamard matrix, reordered into $N \times N$.

Subsampling pattern. The pattern Ω used in (12) was the "half-half" scheme, i.e. a two-level sampling scheme where the first level samples fully and the second level samples uniformly at random. Hadamard matrices contain 1's and -1's but digital mirrors can only represent 1's and 0's so (12) used the modified sampling operator $\Psi_{\mathrm{m}} = (\Psi + \underline{1})/2$ where $\underline{1}$ is the all-1's matrix and Ψ is the Hadamard operator, and solved $\min_z \|\Phi z\|_1$ s.t. $\|y - \Psi_{\mathrm{m}} x\| < \eta$. This, however, is suboptimal since, unlike Ψ, Ψ_{m} is non-orthogonal and far from an isometry.

Point spread effect. A point emission of light is spread by a lens into an airy disc, a blurring effect. The lens acts as a circular low-pass filter, i.e. its 2D Fourier spectrum is a disc. This is important in the CFM setup where the patterns of 0's and 1's contain discontinuities. To mitigate the lens point-spread effect the above authors *binned* mirrors together into groups of 2×2 or 4×4 to represent a single 0 or 1 value, which narrows the Fourier response of the combined light beam coming from such a group, and is less affected by the lens. The major drawback is that this reduces the resolution of the recovered image by 2 or 4 times. This is a serious limitation: as we shall see, CS recovery improves with resolution so limiting to low resolutions causes a cap in performance. Also, the point-spread function (PSF) of the lens was ignored during the CS minimization recovery.

Photonic noise. A further challenge is the photonic (Poisson) noise at the receptor, which essentially counts the number of photons in a preset time interval. Unlike white Gaussian noise in other systems, the photonic noise mean and variance are signal dependent: the noise power grows with the signal. This is accentuated by the CFM setup since there are always $N^2/2$ light beams at a time, generating a large background luminance (DC offset) and thus impacting the signal-to-noise ratio when measuring higher frequency Hadamard components, e.g. patterns where 0's and 1's are alternating.

New approach. In what follows we present an approach which employs a multilevel subsampling pattern, explaining why it is beneficial, takes into account the lens PSF as well as the photonic noise, and also avoids mirror binning, thus reaching much higher resolutions, of the order of 2048×2048. The new approach employs a few techniques to improve performance and strives to stay loyal to the planned practical setup in collaboration with the Cambridge Advanced Imaging Centre (CAIC), now in the process of building a fluorescence microscope of the scale $N = 2048$.

Given a subsample pattern $\Omega \subseteq \{1, \ldots, N^2\}$ with $|\Omega| = m$, denote with $P_{\mathrm{psf}} \in \{0, 1\}^{N^2 \times N^2}$ the projection matrix corresponding to the Fourier response of the PSF (the disc low-pass filter), with $F \in \mathbb{C}^{N^2 \times N^2}$ the 2D Fourier transform and with $x \in \mathbb{R}^{N^2}$ the original specimen image ordered as a vector. In a noiseless scenario, the measurements $\gamma = \{\gamma_i\} \in \mathbb{N}^m$ would be

$$\gamma = \left\lfloor |P_\Omega F^* P_{\mathrm{psf}} F \Psi_{\mathrm{m}} x| \right\rfloor = \left\lfloor \left| \frac{1}{2} P_\Omega F^* P_{\mathrm{psf}} F (\Psi + \underline{1}) x \right| \right\rfloor. \quad \textbf{[9]}$$

Since the photonic noise is dominant in the CFM setup, the actual measurements y can be modelled as values drawn from a Poisson distribution with mean and variance equal to γ, i.e.

$$y = \{y_i\} \sim \mathrm{Poisson}(\{\gamma_i\}) \in \mathbb{N}^m \quad \textbf{[10]}$$

Knowing that a spatially wide light beam is very little affected by the PSF, the following approximation holds to a high accuracy:

$$F^* P_{\mathrm{psf}} F \underline{1} x \simeq \underline{1} x, \quad \textbf{[11]}$$

since $\underline{1}$ gives the widest combined beam, so we can transform the measurements y into

$$y' = 2y - \underline{y_1}, \quad \textbf{[12]}$$

where $\underline{y_1}$ is the vector with all entries equal to y_1 which represents the measurement taken with the all-1's pattern on the mirrors (the first Hadamard matrix row). This allows us to solve [Eq. 3] using the sampling operator $F^* P_{\mathrm{psf}} F \Psi$ (or even just Ψ as we shall explain), which allows fast transforms and is much closer to an isometry compared to $F^* P_{\mathrm{psf}} F \Psi_{\mathrm{m}}$ (or Ψ_{m}). Thus we solve:

$$\min_{z \in \mathbb{C}^{N^2}} \|z\|_1 \quad \text{s.t.} \quad \|y' - P_\Omega U z\| < \eta, \quad \textbf{[13]}$$

where $U = F^* P_{\mathrm{psf}} F \Psi \Phi^*$ and Φ is an orthobasis like wavelets in which the image is expected to be sparse. This allows us to use fast transforms exclusively for U and its adjoint U^*, needed during the above minimization, since both P_{psf} and Ψ are real and symmetric, hence self-adjoint, so $U^* = \Phi \Psi^* F^* P_{\mathrm{psf}} F$.

Fig. 6 shows zebra fish cells captured by a 2048×2048 raster scan florescence microscope from CAIC. Measurements y were formed
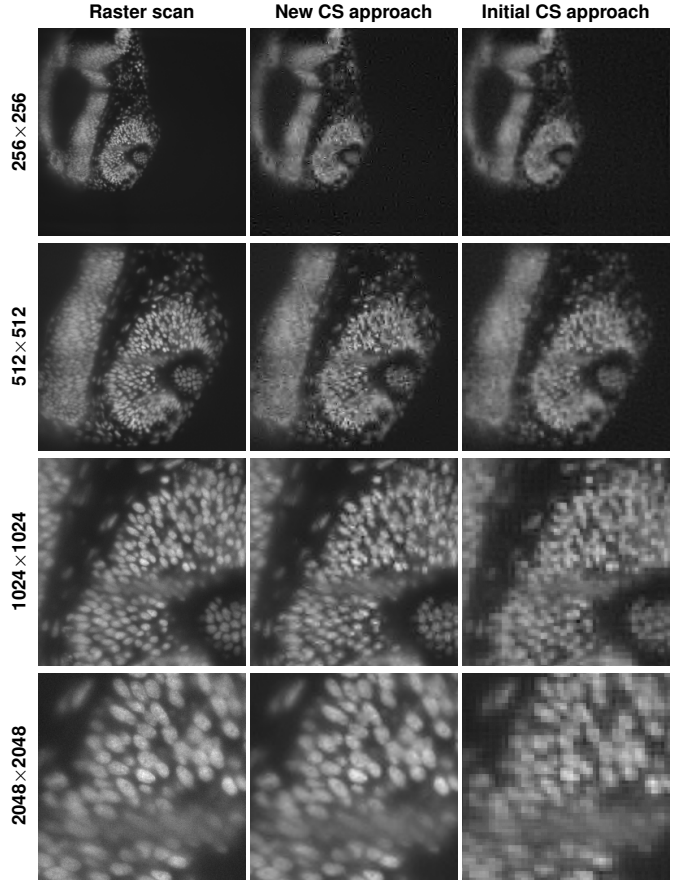

Fig. 6. Fluorescence Microscopy (FM) example. Recovering from 6.25% Hadamard coefficients [Eq. 10] into Daubechies-4. The subsampling pattern for the new CS approach is the the one used in Fig. 2 for the FM case.

using [Eq. 10] (adding large Poisson noise with mean γ [Eq. 9]), where the cutoff value of P_{psf} corresponds to the PSF of the lenses that will be used in the microscope which is now being built, so the real-world measurements are expected to be close to y. As can be seen the new approach CS recovery is much improved compared to the initial approach that uses Ψ_{m} and *half-half* sampling.

Why multilevel subsampling works better. While in Fig. 6 the use of $FP_{\mathrm{psf}}F^*\Psi$ and y' (instead of Ψ_m and y) provide a substantial advantage, a multilevel subsampling pattern Ω intrinsically mitigates the effects of the PSF, thus avoiding any mirror binning. This in turn achieves higher resolutions and also higher quality recovery. The key, as discussed in the section *New CS principles*, is that Ω should follow the asymptotic decrease of the coherence of Hadamard with wavelets. A Hadamard matrix, much like Fourier, captures spatial frequency of increasing orders. The Hadamard rows that give higher coherence with wavelets correspond to lower spatial frequencies, as seen in Fig. 5, i.e. the rows with more adjacent 0's or 1's. The new theory states that favoring those rows when sampling will provide better CS recovery. Importantly, in the CFM setup those rows also inherently emulate mirror binning, owing to the adjaceny of 1's and 0's, so they are bound to be less affected by the lens PSF. For this reason, for an appropriate choice of the multilevel pattern, one could even simply use the faster Ψ (instead of the full $F^*P_{\mathrm{psf}}F\Psi$) as the sampling operator. In contrast, a "half-half" subsampling pattern, besides bound to perform more poorly as it does not closely follow the asymptotic incoherence, also subsamples heavily from the high spatial frequency rows which are more severely affected by the PSF, further decreasing CS recovery quality.

Resolution dependency

A resolution dependency effect could first be noticed in Fig. 6 where the CS recovery gets better as the resolution increases since the image is increasingly (asymptotically) sparser in wavelets, and the coherence between Hadamard and wavelets decreases asymptotically (see Fig. 5). Fig. 7 shows a 2048×2048 MRI image of a pomegranate fruit obtained using a 3T Philips MRI machine, which also contains noise specific to MRI. MRI is an example of Type I problem, where the sampling operator is imposed, but the same new CS principles apply. We subsampled a fixed fraction of 6.25% Fourier samples and solved [Eq. 3] with $U = \Phi_{\mathrm{dft}}\Psi_{\mathrm{dwt}}^*$. The asymptotic sparsity of the wavelet coefficients and the asymptotic incoherence of Fourier and wavelets (see Fig. 5) exploited via multilevel sampling again yield increasingly better reconstruction quality as the resolution increases, in this case to the point where differences are hardly noticeable.

Fixed number of samples. A more striking result of asymptotic sparsity and asymptotic incoherence is obtained by fixing the number of samples taken, instead of the fraction. This was done in Fig. 8, sampling the same number of $512^2 = 262144$ Fourier coefficients in four scenarios, the latter revealing previously hidden details when using multilevel sampling from a broader spectrum.

The explanation? The higher resolution opens up higher-order regions of wavelet coefficients which are mostly sparse, and higher-order regions of coherence between sinusoids and wavelets (see Fig. 5) which is low. As discussed in the section *New CS Principles*, when using a nonlinear recovery, these two asymptotic effects can be fruitfully exploited with a multilevel sampling scheme that spreads the same number of samples across a wider range, aiming for the more coherent regions and reconstructing finer details to a much clearer extent, even in the presence of noise in this example. It is worth noting that other sampling strategies (e.g. half-half) will also benefit from sampling at higher resolutions, provided samples are sufficiently spread out, but a multilevel sampling strategy will provide near optimal guarantees (8).

Sampling strategy is also resolution dependent. The resolution dependency effect also influences the optimum subsampling strategy,

in that the latter will depend on the resolution in addition to signal sparsity and structure. Fig. 9 shows an experiment in which we computed the best subsampling patterns [Eq. 8] for two resolutions of the same image. As is evident, the resulting patterns are different and also yield different results for the same resolution.
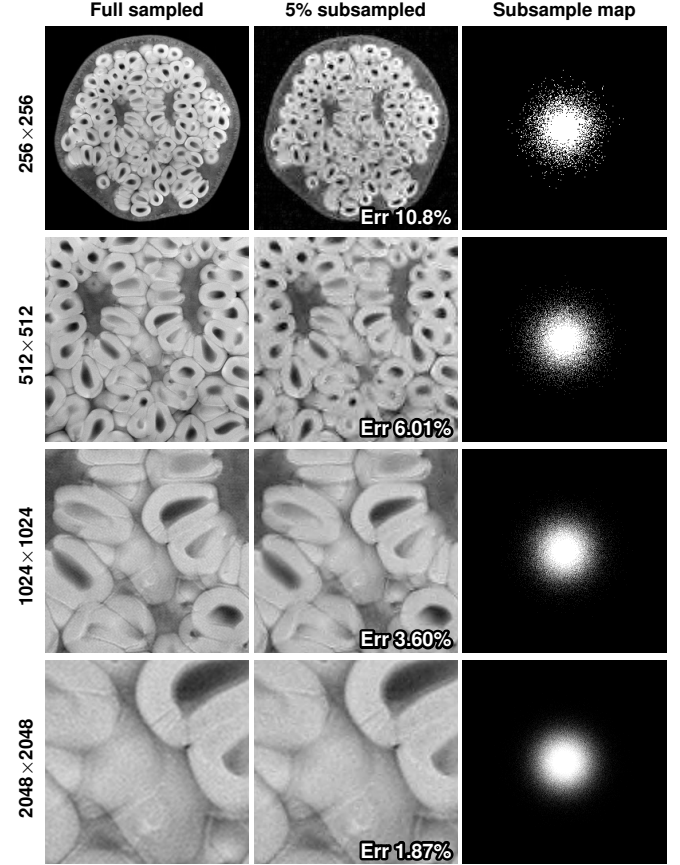

Fig. 7. MRI example. Multilevel subsampling of 5% Fourier coefficients recovered into Daubechies-4.
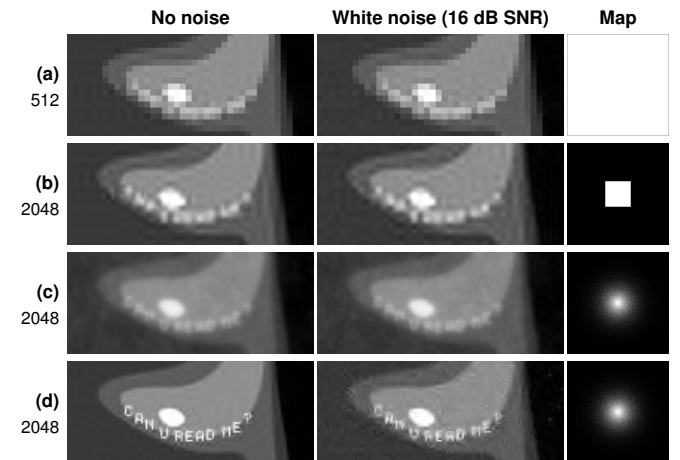

Fig. 8. MRI example. Recovery from a fixed number of $512^2 = 262144$ Fourier coefficients of the phantom from Fig. 1. **(a)** 512×512 linear, from the first 512×512 coefficients. **(b)** 2048×2048 CS into Daubechies-4, from the first 512×512 (6.25%) coefficients. **(c)** 2048×2048 linear, from 512^2 (6.25%) coefficients sampled with a multilevel sampling map. **(d)** 2048×2048 CS into Daubechies-4, from the same 512^2 (6.25%) coefficients from (c).
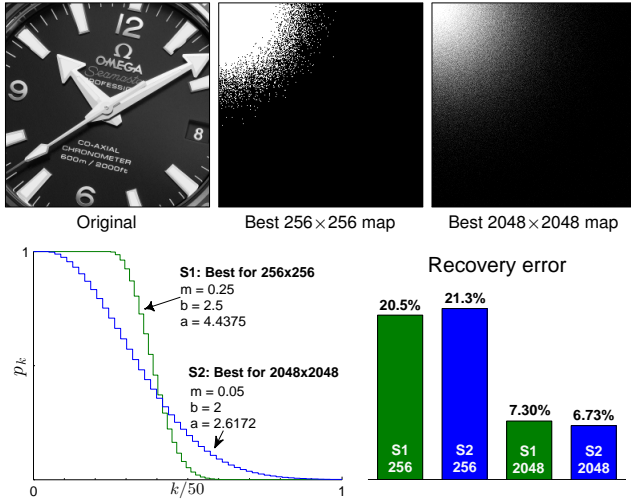
Fig. 9. Compressive Imaging (CI) example. Recovering from 12.5% Hadamard coefficients into Daubechies-4. The best subsampling maps [Eq. 8] were computed heuristically for each resolution of this image.

Remarks. There are a few remarks worth making at this point.
– By simply going higher in resolution (e.g. further in the Fourier spectrum), one can recover a signal much closer to the exact one, yet taking the same number of measurements; or
– By simply going higher in resolution one can obtain the same reconstruction quality, yet taking fewer measurements.
– These experiments showed that it is important in practice to be able to access high resolutions (higher frequencies in the MRI case) in order for CS to provide higher gains. Thus, for MRI, the benefits will be even more visible on future generations of MRI machines.
– Multilevel sampling can better exploit the resolution dependency effect and allows for better tailoring according to sparsity structure, resolution or application specific requirements (e.g. different patterns for different body parts, allow lower overall quality but recover contours better etc) as opposed to uniform random sampling or sampling schemes such as half-half (12) or continuous power laws (13).
– Thirdly, practical CS in MRI has several limitation regarding pointwise sampling. The multilevel patterns used here are the result of our quest for a theoretically optimal sampling pattern, which could then be approximated by realistic MRI patterns or contours, e.g. parametric spirals. The latter is work in progress in collaboration with the Wolfson Brain Imaging Centre.

Infinite dimensional problems

The underlying model in some applications is continuous, such as in MRI, EM, X-ray tomography and its variants. These are Type I problems, where the sensing operator is imposed. In MRI, the measurements y are samples of the continuous (integral) Fourier transform $\mathcal{F}$. The same applies for EM and X-ray and its variants, where the Radon transform is sampled one angle at the time. Via the Fourier slice theorem, the procedure is equivalent to sampling the Fourier transform at radial lines and so the Fourier and Radon transform recovery problems are equivalent to recovering the continuous f from pointwise samples $\hat{g}$, which are evaluations of

$$g = \mathcal{F}f, \quad f \in \mathrm{L}^2(\mathbb{R}^d), \quad \mathrm{supp}(f) \subseteq [0,1]^d. \qquad \textbf{[14]}$$

The issue. Consider the CS recovery into wavelets (see (14) for the general case). Using discrete tools, in this case $U = \Psi_{\mathrm{dft}}\Phi_{\mathrm{dwt}}^*$ in [Eq. 3], to solve a continuous inverse problem can cause large errors, due to *measurement mismatch* (11, 15) and the *wavelet crime* (16). The former assumes a discrete model for f, $\tilde{f} = \sum_{j=1}^N \tilde{\beta}_j\psi_j$, where

ψ_j are step functions, and then (more seriously) replaces the continuous f and $\mathcal{F}$ in [Eq. 14] with $\tilde{f}$ and DFT respectively, leading to the discretization $\tilde{g} = \Psi_{\mathrm{dft}}\tilde{\beta}$, which is a poor approximation of the samples of g. The wavelet crime is as follows. Given scaling and mother wavelet functions φ and ψ, obtaining the wavelet coefficients of f via the DWT should assume $f = \sum_{j=-\infty}^{\infty}\beta_j\varphi(\cdot - j)$ and then compute them from $\{\beta_j\}$ via the DWT. The crime is when one simply replaces β_j with pointwise samples of f.

To illustrate these issues, let $y = P_\Omega\hat{g}$ be the measurements in [Eq. 3], where $\hat{g}$ are the first $2N$ continuous Fourier samples of f. The matrix Ψ_{dft}^* maps $\hat{g}$ to a vector $x \in \mathbb{C}^{2N}$ on an equispaced $2N$ grid of points in $[0,1]$. Specifically, $\Psi_{\mathrm{dft}}x = \hat{g}$ where x is given by the values $f_N(t) = 1/2\sum_{j=1-N}^N \mathcal{F}f(j/2)\mathrm{e}^{2\pi\mathrm{i}\epsilon j}$ on the $2N$ grid. Taking $x_0 = \Phi_{\mathrm{dwt}}x$, the right-hand side of [Eq. 3] becomes $P_\Omega\Psi_{\mathrm{dft}}\Phi_{\mathrm{dwt}}^*x_0$. But [Eq. 3] now requires x_0 to be sparse, which means the truncated Fourier series f_N must be sparse in wavelets, which cannot happen. While f is assumed sparse in wavelets, f_N is not, since it consists of smooth complex exponentials. Large errors thus occur in the recovery, as there is no sparse solution due to the poor approximation of f by f_N. This could be avoided if the measurements arose from the DFT, but that would be the well-known and pervasive *inverse crime* (11): artificially superior performance when data is simulated incorrectly using the DFT, as opposed to the continuous FT, which is the true underlying model.

How to solve. The above crimes stem from discretizing first and then applying finite-dimensional tools. Instead, we shall use the techniques of *infinite-dimensional* CS (10), i.e. first formulate the problem in infinite dimensions and *then* discretize. Let $\{\psi_j\}_{j\in\mathbb{N}}$ and $\{\varphi_j\}_{j\in\mathbb{N}}$ be the sampling and sparsity bases (Fourier and wavelets). If $f = \sum_{j\in\mathbb{N}}\beta_j\varphi_j$, then the unknown vector of coefficients $\beta = \{\beta_j\}_{j\in\mathbb{N}}$ is the solution of $U\beta = \hat{f}$, where

$$U = \begin{pmatrix} \langle\varphi_1,\psi_1\rangle & \langle\varphi_2,\psi_1\rangle & \cdots \\ \langle\varphi_1,\psi_2\rangle & \langle\varphi_2,\psi_2\rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix},$$

and $\hat{f} = \{\hat{f}_j\}_{j\in\mathbb{N}}$ is the infinite vector of samples of f. Suppose we have access to a finite number of samples $\{\hat{f}_j : j \in \Omega\}$, where Ω is the sampling map. To recover β from these samples, we first formulate the infinite-dimensional optimization problem

$$\inf_{z\in\ell^1(\mathbb{N})} \|z\|_1 \text{ subject to } P_\Omega Uz = P_\Omega\hat{f}, \qquad \textbf{[15]}$$
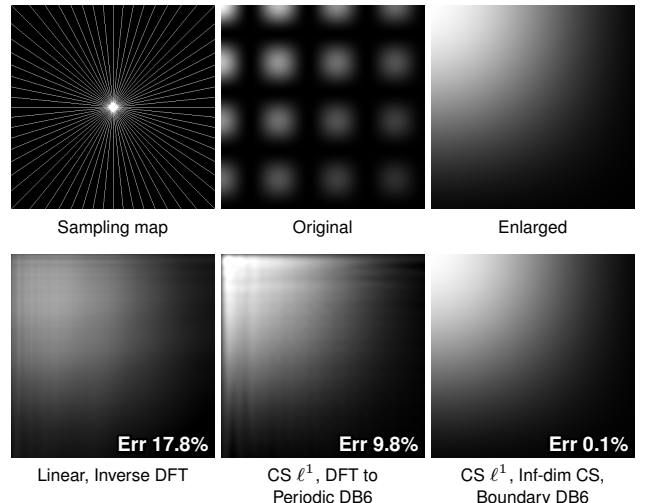

Fig. 10. Electron Microscopy (EM) example. Recovery from 16120 (6.15%) continuous FT samples onto a 512×512 grid.

which commits no crimes. However, [Eq. 15] cannot typically be solved numerically, so we now discretize. We truncate U to $K \in \mathbb{N}$ columns, solving the now finite-dimensional problem

$$\min_{z \in P_K(\ell^2(\mathbb{N}))} \|z\|_1 \text{ subject to } P_\Omega U P_K z = P_\Omega \hat{f}. \qquad \textbf{[16]}$$

We refer to this as infinite-dimensional CS, where K should be sufficiently large to ensure good recovery. Fig. 10 shows an example of such a continuous problem where the continuous function $f(x,y) = \exp(-x - y)\cos^2(x)\cos^2(y)$ is recovered on a 512×512 grid from 16120 (6.15%) continuous Fourier samples taken radially, as in an EM setup. The infinite-dimensional CS recovery is done in boundary DB6 wavelets (17) instead of the periodic DB6 (to preserve the vanishing moments at the boundaries). It is evident that its reconstruction is far superior to both the discrete linear reconstruction (via Ψ_{dft}^*) and discrete CS reconstruction (via $U = \Psi_{\text{dft}}^* \Phi_{\text{dwt}}^*$) which are affected by the crimes.

In conclusion, given sufficiently many vanishing moments, the infinite dimensional CS with boundary wavelets will give substantially better convergence than the slow truncated Fourier convergence f_N whenever f is non-periodic. Knowing that the computational complexity in this case is the same as with FFT, this means that infinite-dimensional CS yields a markedly better approximation of f at little additional cost. The infinite-dimensional CS approach described here is of particular benefit to applications like EM and spectroscopy where smooth functions are encountered. This is work in progress in collaboration with the Department of Materials Science & Metallurgy at University of Cambridge.

Structured sampling vs Structured recovery

The new CS principles in this paper take into account sparsity structure in the sampling procedure via multilevel sampling of non-universal sensing matrices. Sparsity structure can also be taken into account in the recovery algorithm. An example of such an approach is model-based CS (18), which assumes the signal is piecewise smooth and exploits the connected tree structure of its wavelet coefficients to reduce the search space of the matching pursuit algorithm (19). Another approach is the class of message passing and approximate message passing algorithms (20, 21), which exploit the persistence across scales structure (4) of wavelet coefficients by a modification to iterative thresholding algorithms inspired by belief propagation from graph models. This can be coupled with hidden Markov trees to model the wavelet structure, such as in the Bayesian CS (22, 23) and TurboAMP (24) algorithms. Another approach is to assign non-uniform weights to the sparsity coefficients (25), to favor the important coefficients during ℓ^1 recovery by assuming some typical decay rate of the coefficients. A recent approach assumes the actual signal (not its representation in a sparsity basis) is sparse and random, and shows promising theoretical results when using spatially coupled matrices (26, 27, 28), yet it is unclear how a real-world setup can be implemented where signals are sparse in a transform domain.

The main difference is that the former approach, i.e. multilevel sampling of asymptotically incoherent matrices, incorporates sparsity structure in the sampling strategy and uses standard ℓ^1 minimization algorithms, whereas the latter approaches exploit structure by modifying the recovery algorithm and use universal sampling operators which yield uniform incoherence (see section *Structure vs Universality*), e.g. random Gaussian or Bernoulli.

Comparison. *Structured recovery:* Due to the usage of universal operators and assumptions on the sparsity basis, this approach is typically restricted to Type II problems, where the sensing operator can be designed, and is further restricted in the choice of the sparsity frame, whose structure is exploited by the modified algorithm.
Structured sampling: In contrast, this approach practically has no limitation regarding the sparsity frame, thus allowing for further im-

provement of CS recovery (see section *Frames and TV*), and it also works for Type I problems, where the sensing operator is imposed.

To compare performance, we ran a large set of simulations of a CI setup. CI (29, 30), a Type II problem, is an application where universal sensing matrices have been traditionally favored. Here the measurements y are typically taken using a sensing matrix with only two values (usually 1 and -1). Any matrix with only two values fits this setup, e.g. Hadamard, random Bernoulli, Sum-To-One (see section *Storage/speed*), hence we can directly compare the two approaches. Fig. 11 shows a representative example from that set, which points to
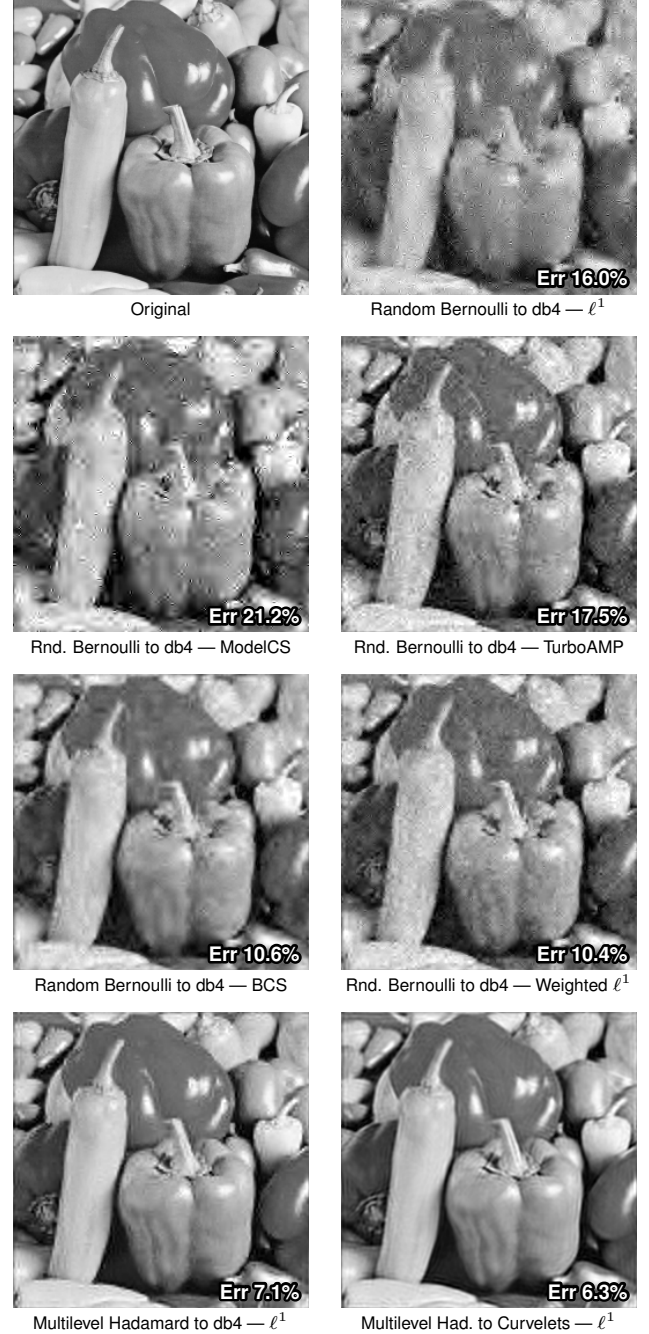

Original	Random Bernoulli to db4 — ℓ^1 — Err 16.0%
Rnd. Bernoulli to db4 — ModelCS — Err 21.2%	Rnd. Bernoulli to db4 — TurboAMP — Err 17.5%
Random Bernoulli to db4 — BCS — Err 10.6%	Rnd. Bernoulli to db4 — Weighted ℓ^1 — Err 10.4%
Multilevel Hadamard to db4 — ℓ^1 — Err 7.1%	Multilevel Had. to Curvelets — ℓ^1 — Err 6.3%

Fig. 11. Compressive Imaging (CI) example. 12.5% subsampling at 256×256. The multilevel subsample map is the one from Fig. 2. The weighted ℓ^1 weights were $w_k = a^k$ for all coefficients in the kth wavelet scale, where $a = 2$ was computed heuristically for this example.

the conclusion that asymptotic incoherence combined with multilevel sampling of highly non-universal sensing matrices (e.g. Hadamard, Fourier) allows structured sparsity to be better exploited than universal sensing matrices, even when structure is accounted for in the recovery algorithm. The figure also shows the added benefit of being able to use a better sparsifying system, in this case curvelets.

Structure vs Universality: Is universality desirable?

Universality is a reason for the popularity in traditional CS of random sensing matrices, e.g. Gaussian or Bernoulli. A random matrix $A \in \mathbb{C}^{m \times N}$ is universal if for any isometry $\Psi \in \mathbb{C}^{N \times N}$, the matrix $A\Psi$ satisfies the RIP with high probability. For images, a common choice is $\Psi = \Psi_{\mathrm{dwt}}^*$, the inverse wavelet transform. Universality helps when the signal is sparse but possesses no further structure.

But is universality desirable in a sensing matrix when the signal is structured? First, random matrices are largely inapplicable in Type I problems, where the sampling operator is imposed. They are applicable mostly in Type II problems, where there is freedom to design the sampling operator. Should then one use universal matrices there? We argue that universal matrices offer little room to exploit extra structure the signal may have, even in Type II problems.

Practical applications typically entail signals that exhibit far more structure than sparsity alone: in particular, asymptotic sparsity structure in some sparsity basis. Thus, an alternative is to use a non-universal sensing matrix, such as Hadamard, Φ_{Had}. As previously discussed and shown in Fig. 5, $U = \Phi_{\mathrm{Had}}\Psi_{\mathrm{dwt}}^*$ is completely coherent with all wavelets yet asymptotically incoherent, and thus perfectly suitable for a multilevel sampling scheme which can exploit the inherent asymptotic sparsity. This is precisely what we see in Fig. 11: multilevel sampling of a Hadamard matrix can markedly outperform solutions employing universal matrices in Type II problems. For Type I problems, an important practical aspect is that many imposed sensing operators happen to be highly non-universal and asymptotically incoherent with popular sparsity bases, and thus easily exploitable using multilevel sampling, as seen in Figs. 1 and 2.

Asymptotic incoherence vs Uniform incoherence. The reasons for the superior results are rooted in the incoherence structure. Universal and close to universal sensing matrices typically provide a relatively low and flat coherence pattern. This allows sparsity to be exploited by sampling uniformly at random but, by definition, these matrices cannot exploit the distinct asymptotic sparsity structure when using a typical (ℓ^1 minimization) CS reconstruction.

In contrast, when the sensing matrix provides a coherence pattern that aligns with the signal sparsity pattern, one can fruitfully exploit such structure. As discussed in the sections *New CS principles* and *Resolution dependency* a multilevel sampling scheme is likely to give superior results by sampling more in the coherent regions, where the signal is also typically less sparse. Even though the optimum sampling strategy is signal dependant (see section *Traditional CS*), real-world signals, particularly images, share a fairly common structure and thus good, all-round multilevel sampling strategies can be designed. An added benefit of multilevel sampling is that it also allows for tailoring of the sampling pattern to target application-specific features rather than an all-round approach, e.g. allowing a slightly lower overall quality but recovering contours better.

Storage/speed: Is non-random/orthogonality enough?

Random matrices have another important practical drawback: they require (large) storage and lack fast transforms. This limits the maximum signal resolution and yields slow recovery. For example, a 1024×1024 experiment with 25% subsampling of a random Gaussian matrix would require 2 Terabytes of free memory and $\mathcal{O}(10^{12})$ time complexity, making it impractical at best.

A low maximum resolution is a big limitation not just for computations. As seen in the section *Resolution dependency*, at low resolutions the asymptotic sparsity has not kicked in and CS yields only marginal benefits. In order to obtain better recovery it is thus of great interest to be able to access high signal resolutions, yet random matrices prevent that.

But is the problem of quality CS recovery solved if we address the storage and speed issues? These were in fact addressed to various extents, e.g. pseudo-random column permutations of the columns of orthogonal matrices such as (block) Hadamard or Fourier (31, 32), Kronecker products of random matrix stencils (33), or even fully orthogonal matrices such as the Sum-To-One (STOne) matrix (34) which has a fast $\mathcal{O}(N \log N)$ transform. However, all these strive to provide universality, i.e. behave like purely random matrices. Fig. 12 shows an extract of a large experiment on various images, resolutions, sparsity bases and subsampling rates, which tests for universality and RIP by performing the flip test (see section *Traditional CS*). It is evident that the orthogonal STOne matrix behaves like random matrices in the CS context which we probed in Fig. 11 (though we note that
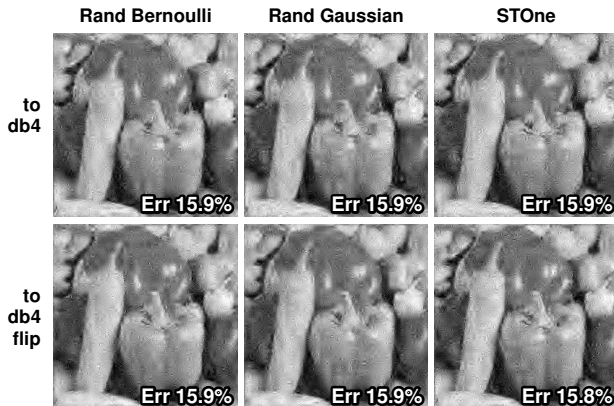

Fig. 12. Subsampling 12.5% at 256×256 with Bernoulli, Gaussian and STOne; part of a larger set of experiments using various images, ratios and sparsity bases. All three matrices yield very similar quality, indicating that they behave the same and that *universality* and *RIP* hold for all three.
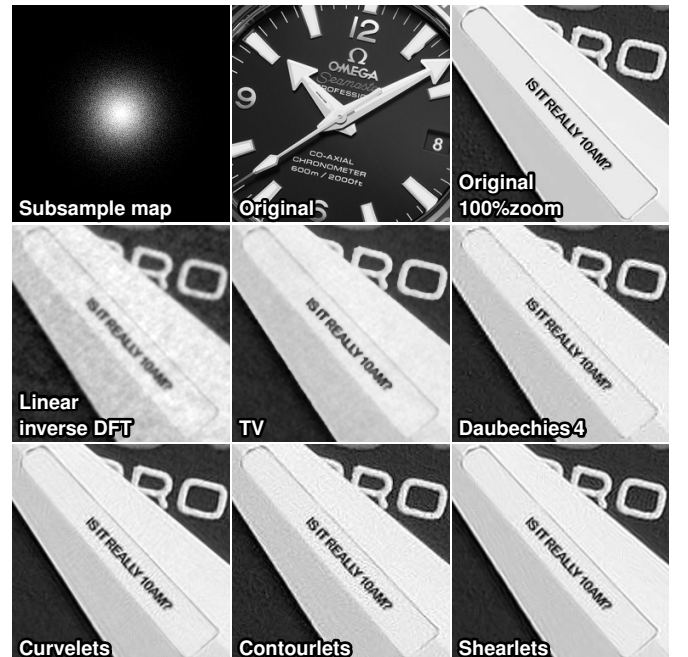

Fig. 13. Recovering from the same 6.25% DFT coefficients at 2048×2048.

the STOne matrix was invented to serve other purposes as well (34)). The poor performance is due to their flat incoherence with sparsity bases, discussed in the section *Structure vs Universality*.

Another solution to the storage and speed problem is to instead use orthogonal and structured matrices like Hadamard, DCT or DFT. These have fast transforms and do not require storage, but also provide asymptotic incoherence with sparsity bases, thus a multilevel subsampling scheme can be used. The important added benefits are that this yields significantly better CS recovery in most Type II problems when compared to universal matrices, as discussed in the section *Structure vs Universality* and probed in Fig. 11, and that, unlike universal matrices, it is also applicable to Type I problems, which impose the sensing operator.

In conclusion, the sensing matrix must contain additional structure besides simply being non-random and/or orthogonalin order to provide asymptotic incoherence. Typically, sensing and sparsifying matrices that are discrete versions of integral transforms, e.g. Fourier, wavelets etc. will provide asymptotic incoherence, but other orthogonal and structured matrices like Hadamard will do so too.

Frames and TV: Freedom of choice

Many images are known to be sparser in TV or frames such as curvelets (5), shearlets (7) or contourlets (6), than in orthobases such as wavelets or DCT. Without going into details regarding frames or TV as sparsifying systems, some further results are shown in Fig. 13, which provides clear experimental verifications of the improvements offered by such sparsifying systems at practical resolution levels.

More important is that, unlike the class of modified recovery algorithms from the section *Structure sampling vs Structure recovery*, and in addition to the benefits discussed in the sections *Structure vs Universality* and *Storage/speed*, incorporating sparsity structure in the sampling procedure also offers complete freedom in the choice of the sparsity system. This holds generally, and is of particular interest in applications where the sampling operator is imposed.

Concluding remarks

The traditional CS pillars: sparsity, incoherence and uniform random subsampling, are often inapplicable in Type I problems, where the sampling operator is imposed (MRI, EM, Tomography, Interferometry etc.), while for Type II problems, where the sampling operator can be designed (FM, CI etc.), they provide little room to exploit extra sparsity structure that real-world signals typically possess. This is due to the coherent nature of Type I problems and of the uniform incoherence of universal sampling operators with sparsity bases.

The new CS principles: asymptotic sparsity, asymptotic incoherence and multilevel subsampling, introduced by the authors (8) to bridge the gap between theoretical and practical CS, provide a better fit for both types of problems. This paper shows how the new principles can be used to better understand the underlying phenomena in practical CS problems, and that an approach based on the new CS principles coupled with non-universal sampling operators can overcome many traditional CS limitations and provide several important benefits and improved CS recovery in real-world applications.

1. Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T Inf Theory* 52(2):489–509.
2. Donoho DL (2006) Compressed sensing. *IEEE T Inf Theory* 52(4):1289–1306.
3. Daubechies I (1992) *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics (Society for Industrial and Applied Mathematics).
4. Mallat SG (2009) *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd ed (Academic Press).
5. Candès E, Donoho DL (2002) Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.* 30(3):784–842.
6. Do MN, Vetterli M (2005) The contourlet transform: An efficient directional multiresolution image representation. *IEEE T Image Proc* 14(12):2091–2106.
7. Dahlke S, et al. (2008) The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.* 6(2):157–181.
8. Adcock B, Hansen AC, Poon C, Roman B (2014) Breaking the coherence barrier: A new theory for compressed sensing. *arXiv:1302.0561* .
9. Candès EJ, Plan Y (2011) A probabilistic and RIPless theory of compressed sensing. *IEEE T Inf Theory* 57(11):7235–7254.
10. Adcock B, Hansen AC (2011) Generalized sampling and infinite-dimensional compressed sensing. Tech Rep NA2011/02, University of Cambridge.
11. Guerquin-Kern M, Lejeune L, Pruessmann KP, Unser M (2012) Realistic analytical phantoms for parallel Magnetic Resonance Imaging. *IEEE T Med Imaging* 31(3):626–636.
12. Studer V, et al. (2011) Compressive fluorescence microscopy for biological and hyperspectral imaging. *Natl Acad Sci USA* 109(26):1679–1687.
13. Krahmer F, Ward R (2012) Compressive imaging: stable and robust recovery from variable density frequency samples. *arXiv:1210.2380* .
14. Adcock B, Hansen AC, Roman B, Teschke G (2014) Generalized sampling: Stable reconstructions, inverse problems and compressed sensing over the continuum. *Advances in Imaging and Electron Physics*, vol 182, ed Hawkes PW (Elsevier), pp 187–279.
15. Chi Y, Scharf LL, Pezeshki A, Calderbank R (2011) Sensitivity to basis mismatch in compressed sensing. *IEEE T Signal Proc* 59(5):2182–2195.
16. Strang G, Nguyen T (1996) *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, MA).
17. Cohen A, Daubechies I, Vial P (1993) Wavelet bases on the interval and fast algorithms. *Appl. Comput. Harmon. Anal.* 1:54–81.
18. Baraniuk RG, Cevher V, Duarte MF, Hedge C (2010) Model-based compressive sensing. *IEEE T Inf Theory* 56(4):1982–2001.
19. Needell D, Tropp J (2009) CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* 26(3):301–321.
20. Baron D, Sarvotham S, Baraniuk R (2010) Bayesian compressive sensing via belief propagation. *IEEE T Signal Proc* 58(1):269–280.
21. Donoho DL, Maleki A, Montanari A (2009) Message-passing algorithms for compressed sensing. *Natl Acad Sci USA* 106(45):18914–18919.
22. He L, Carin L (2009) Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE T Signal Proc* 57(9):3488–3497.
23. He L, Chen H, Carin L (2010) Tree-structured compressive sensing with variational Bayesian analysis. *IEEE Signal Proc Lett* 17(3):233–236.
24. Som S, Schniter P (2012) Compressive imaging using approximate message passing and a markov-tree prior. *IEEE T Signal Proc* 60(7):3439–3448.
25. Khajehnejad M, Xu W, Avestimehr A, Hassibi B (2011) Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models. *IEEE T Signal Proc* 59(5):1985–2001.
26. Wu Y, Verdu S (2010) Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE T Inf Theory* 56(8):3721–3748.
27. Krzakala F, Mézard M, Sausset F, Sun YF, Zdeborová L (2012) Statistical-physics-based reconstruction in compressed sensing. *Phys Rev X* 2:021005.
28. Donoho D, Javanmard A, Montanari A (2013) Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE T Inf Theory* 59(11):7434–7464.
29. Takhar D, et al. (2006) A new compressive imaging camera architecture using optical-domain compression. *Comput Imaging IV at SPIE Electronic Imaging*, pp 43–52.
30. Huang G, Jiang H, Matthews K, Wilford PA (2013) Lensless imaging by compressive sensing. *IEEE Intl Conf Image Proc,*, pp 2101–2105.
31. Candès E, Romberg J (2006) Robust signal recovery from incomplete observations. *IEEE Intl Conf Image Proc*, pp 1281–1284.
32. Gan L, Do TT, Tran TD (2008) Fast compressive imaging using scrambled hadamard ensemble. *Proc Eur Signal Proc Conf*, pp 139–154.
33. Duarte M, Baraniuk R (2012) Kronecker compressive sensing. *IEEE T Image Proc* 21(2):494–504.
34. Goldstein T, Xu L, Kelly KF, Baraniuk RG (2013) The stone transform: Multi-resolution image enhancement and real-time compressive video. *arXiv:1311.3405* .