A robust FFAST framework for computing a k-sparse n-length DFT in $O(k \log n)$ sample complexity using sparse-graph codes

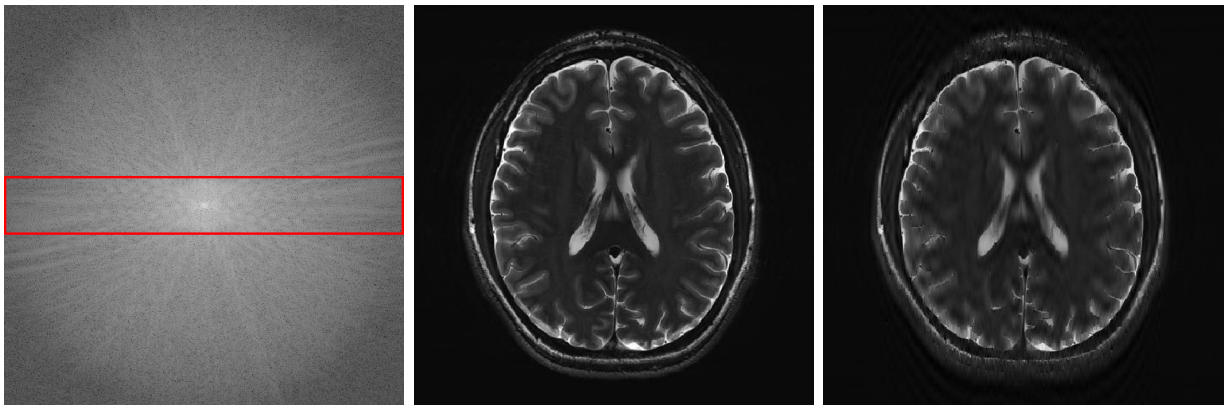
Sameer Pawar and Kannan Ramchandran
Dept. of Electrical Engineering and Computer Sciences
University of California, Berkeley
{spawar, kannanr}@eecs.berkeley.edu

Abstract

The Fast Fourier Transform (FFT) is the most efficiently known way to compute the Discrete Fourier Transform (DFT) of an arbitrary n-length signal, and has a computational complexity of $O(n \log n)$[1]. If the DFT $\vec{X}$ of the signal $\vec{x}$ has only k non-zero coefficients (where $k < n$), can we do better? In [1], we presented a novel FFAST (Fast Fourier Aliasing-based Sparse Transform) algorithm that cleverly induces sparse graph codes in the DFT domain, via a Chinese-Remainder-Theorem (CRT)-guided sub-sampling operation of the time-domain samples. The resulting sparse graph code is then exploited to devise a simple and fast iterative onion-peeling style decoder that computes an n length DFT of a signal using only $O(k)$ time-domain samples and $O(k \log k)$ computations, in the absence of any noise.

In this paper, we extend the FFAST framework of [1] to the case where the time-domain samples are corrupted by white Gaussian noise. In particular, we show that the extended noise robust FFAST algorithm computes an n-length k-sparse DFT $\vec{X}$ using $O(k \log n)$ noise-corrupted time-domain samples, in $O(n \log n)$ computations. While our theoretical results are for signals with a uniformly random support of the non-zero DFT coefficients and additive white Gaussian noise, we provide simulation results which demonstrates that the FFAST algorithm performs well even for signals like MR images, that have an approximately sparse Fourier spectrum with a non-uniform support for the dominant DFT coefficients.

[1]Further, the constants in the big-oh notations are small as is evident from the experimental results provided in Section VII.



(a) Log intensity plot of the $2D$-DFT of the original 'Brain' image.

(b) Original 'Brain' image in spatial domain.

(c) Reconstructed 'Brain' image.

Fig. 1. Application of the $2D$-FFAST algorithm to reconstruct the 'Brain' image acquired on an MR scanner with dimension 504×504. The $2D$-FFAST algorithm reconstructs the 'Brain' image, as shown in Fig. 1(c), using overall 60.18% of the Fourier samples of Fig. 1(a).

I. Introduction

The Fast Fourier Transform is the fastest known way to compute the DFT of an arbitrary n-length signal, and has a computational complexity of $O(n \log n)$. Many applications of interest involve signals that have a sparse Fourier spectrum, e.g. relating to audio, image, and video data, biomedical signals, cognitive radio applications, and satellite imagery, etc. In [1], we have proposed a novel FFAST framework, that under idealized assumptions of no measurement noise and the spectrum being *exactly k-sparse* (i.e., having precisely k non-zero DFT coefficients), computes an n-length DFT $\vec{X}$ using only $O(k)$ time-domain samples in $O(k \log k)$ arithmetic computations. For long signals, i.e., when n is of order of millions or tens of millions, as is becoming more relevant in the Big-data age, the gains over conventional FFT algorithms can be significant. The idealized assumptions in [1] were primarily used to highlight the novel ideas underlying the FFAST architecture with conceptual clarity.

In this paper, we extend the FFAST framework of [1] to the case of computing a sparse DFT *in the presence of additive input sample noise*. We show that the FFAST framework can be efficiently robustified to handle the case where the input samples are noisy by elegant modification of the "delay chain" used in [1]. Specifically, the FFAST front-end in [1] has 2 paths of sub-sampling operations in each stage. The two sub-sampling paths have *identical sampling period* but the input signal is circular-shifted by 1, i.e., *consecutive delays*, before sub-sampling. In contrast, for the noise-robust version, we use $O(\log n)$ number of sub-sampling paths per stage. Further, each sub-sampling path circularly shifts the input signal by a *random* amount before sub-sampling, i.e., *random delays*. A random choice of the circular shifts

endows the effective measurement matrix with a good *mutual incoherence* and *RIP* [10], thus resulting in stable recovery.

As a motivating example, we demonstrate an application of a $2D$-extension of the FFAST algorithm to acquire the Magnetic Resonance Image (MRI) of the 'Brain' as shown in Fig. 1. In MRI, recall that the samples are acquired in the Fourier domain, and the challenge is to speed up the acquisition time by minimizing the number of Fourier samples needed to reconstruct the desired spatial domain image. The FFAST algorithm reconstructs the 'Brain' image acquired on an MR scanner with dimension 504×504, using overall 60.18% of the Fourier samples of Fig. 1(a). The reconstructed 'Brain' image is shown in Fig. 1(c). In Section VII-C, we elaborate on the specific details of this experiment.

This application leads to some interesting observations. Although the theoretical results of this paper are applicable to signals that have exactly k non-zero DFT coefficients with uniformly random support and additional Gaussian noise, in practice our algorithm performs well even for signals that are approximately sparse and have "non-uniform" (or clustered) support for the dominant DFT coefficients, such as the 'Brain' image of Fig 1(b). However, the focus of this paper is to analyze $1D$ signals having an *exactly-sparse DFT plus observation noise*.

In this paper, we show that the FFAST algorithm computes a k-sparse n-length DFT of a signal from $O(k \log n)$ noise-corrupted samples in $O(n \log n)$ computations. This contrasts the best known scaling, i.e., $O(k \log^3 k \log n)$, of the partial Fourier measurements in the compressed-sensing literature for the worst case model for the input signal [11]. We emphasize the following caveats. First, we assume that the non-zero DFT coefficients of the signal $\vec{x}$ have uniformly random support and take values from a finite constellation (as explained in Section II). Secondly, our results are probabilistic and are applicable for asymptotic values of k, n, where k is sub-linear in n. Lastly, we assume an i.i.d Gaussian noise model for observation noise.

A. Main idea

We use a simple example to illustrate the key ideas. Consider an $n = 20$ length input signal $\vec{x}$, whose DFT $\vec{X}$, is $k = 5$ sparse. Further, let the 5 non-zero DFT coefficients of $\vec{x}$ be $X[1]$, $X[3]$, $X[5]$, $X[10]$ and $X[13]$. Let $\vec{y} = \vec{x} + \vec{w}$ be the noise-corrupted observation of the signal. In general, the FFAST sub-sampling 'front-end' consists of d stages, where each stage further has D subsampling paths, see Fig. 3. For illustrative purposes, in Fig. 2 we show the processing of $\vec{y}$ through stage 0 of a 2-stage FFAST architecture. The FFAST front-end subsamples 3 circularly shifted versions of the observation $\vec{y}$, by a sampling period of 5. The output of stage 0, of the FFAST front-end, is then obtained by computing

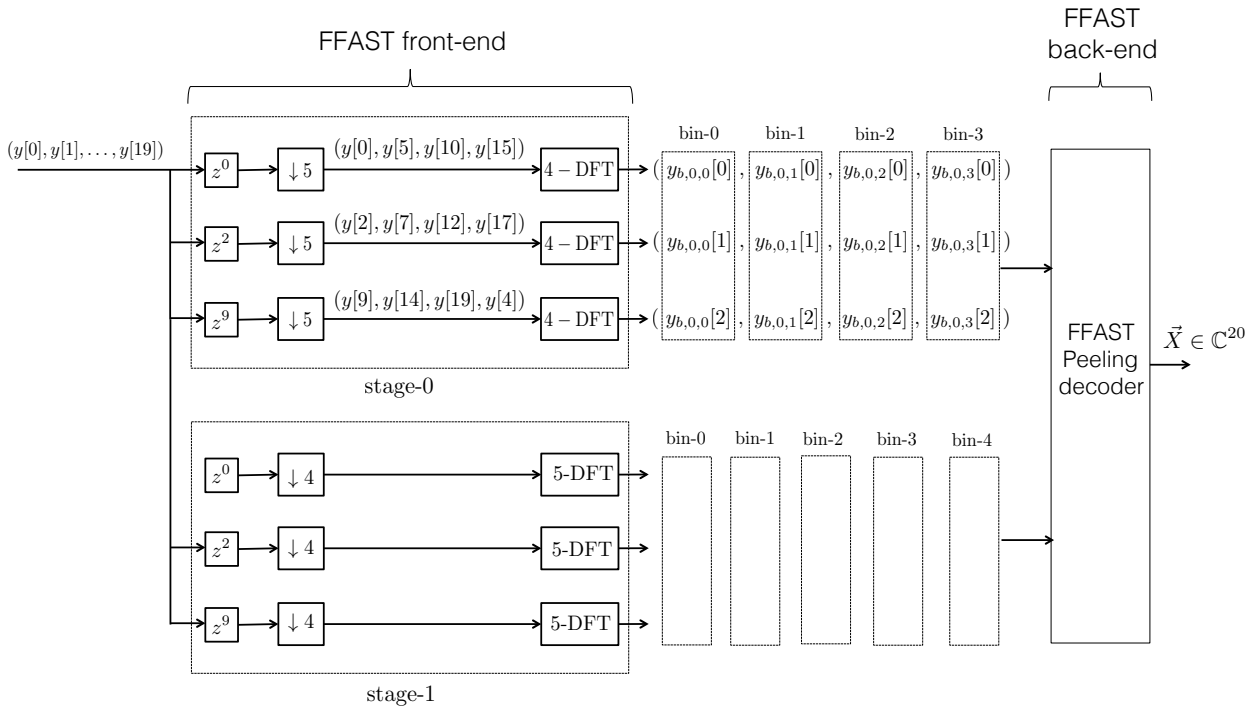


Fig. 2. The noise-corrupted observation $\vec{y} = \vec{x} + \vec{w}$, is processed through a 2-stage FFAST architecture. In general, the sub-sampling front-end of the FFAST architecture consists of 3 or more stages depending on the sparsity index δ, where $k = O(n^{\delta})$. In this example, we consider the FFAST architecture with 2-stages only for the purpose of illustration. Further, here we show only stage 0 of a 2-stage FFAST architecture. The 20-point DFT of the signal $\vec{x}$ is 5 sparse, with the non-zero DFT coefficients being $X[1]$, $X[3]$, $X[5]$, $X[10]$ and $X[13]$. The FFAST front-end subsamples 3 circularly shifted versions of the observation $\vec{y}$, by a sampling period of 5. The number and the pattern of the circular shifts are chosen carefully as explained in Section V-A. The circular shifts z^0, z^2 and z^9 used in this example are only for illustrative purposes. The output of stage 0, of the FFAST front-end, is then obtained by computing the 4-point DFT of each of the sub-sampled stream and further grouping them into 'bins'. A 3-dimensional vector $\vec{y}_{b,i,j}$ denotes the observation of bin j of stage i, e.g., $\vec{y}_{b,0,1}$ is the observation of bin 1 of stage 0.

the 4-point DFT of each of the sub-sampled stream and further grouping them into 'bins', as shown in Fig. 2. Let $\vec{y}_{b,i,j}$ denote the 3-dimensional observation vector of bin j of stage i. Using the basic signal processing identities of sampling-aliasing and circular shifts, the relation between the bin-observation

vectors and the DFT coefficients of the input signal $\vec{x}$, can be written as,

$$\vec{y}_{b,0,0} = \begin{pmatrix} w_{b,0,0}[0] \\ w_{b,0,0}[1] \\ w_{b,0,0}[2] \end{pmatrix}, \quad \vec{y}_{b,0,2} = \begin{pmatrix} 1 \\ e^{\imath 2\pi 20/20} \\ e^{\imath 2\pi 90/20} \end{pmatrix} X[10] + \begin{pmatrix} w_{b,0,2}[0] \\ w_{b,0,2}[1] \\ w_{b,0,2}[2] \end{pmatrix},$$

$$\vec{y}_{b,0,1} = \begin{pmatrix} 1 \\ e^{\imath 2\pi 2/20} \\ e^{\imath 2\pi 9/20} \end{pmatrix} X[1] + \begin{pmatrix} 1 \\ e^{\imath 2\pi 10/20} \\ e^{\imath 2\pi 45/20} \end{pmatrix} X[5] + \begin{pmatrix} 1 \\ e^{\imath 2\pi 26/20} \\ e^{\imath 2\pi 117/20} \end{pmatrix} X[13] + \begin{pmatrix} w_{b,0,1}[0] \\ w_{b,0,1}[1] \\ w_{b,0,1}[2] \end{pmatrix} \quad (1)$$

The FFAST algorithm proceeds as follows:

- *Divides the problem into simpler sub-problems*: The FFAST sub-sampling front-end takes a 20-length noise-corrupted observation $\vec{y}$, and disperses it into 4 bin observations, as shown in Fig. 2. Each bin has 3 output samples (corresponding to the 3 delay chains), and forms an instance of a sub-problem of computing a sparse DFT of a high-dimensional signal. Most sub-problems are trivial, consisting of computing a 0-sparse DFT, e.g., $\vec{y}_{b,0,0}$, or a 1-sparse DFT, e.g., $\vec{y}_{b,0,2}$, while the others are almost trivial, of computing a 3-sparse DFT, e.g., $\vec{y}_{b,0,1}$.
- *Iterative peeling-decoder*: The FFAST decoder identifies an instance of a sub-problem that is 1-sparse and reliably computes the support and the value of the non-zero DFT coefficient, e.g., $X[10]$ in $\vec{y}_{b,0,2}$, participating in this sub-problem. Then, it *peels off* the contribution of the identified non-zero DFT coefficient, from other sub-problems, to create more instances of 1-sparse sub-problems. This peeling-style iterative recovery algorithm eventually uncovers all the non-zero DFT coefficients.
- *Noise robust processing*: The circular shift (or delay) pattern, used in the FFAST front-end architecture, determines the structure of the effective measurement matrix $\boldsymbol{A}_{i,j} \in \mathbb{C}^{3\times 20}$, of the sub-problem corresponding to bin j of stage i. The noise robust decoding of the 1-sparse sub-problems is achieved by judiciously choosing the delay pattern, such that the measurement matrices of all the sub-problems have good mutual incoherence property, i.e., the columns of the measurement matrix are uncorrelated, like Restricted-Isometry-Property (RIP) [10].

At a high level, the FFAST architecture through its multi-stage sub-sampling front-end, divides the original "difficult" (k-sparse) problem into many "simpler" (1-sparse) sub-problems. Then, it solves the 1-sparse sub-problems *reliably*, in the presence of observation noise, using multiple D measurements per sub-problem and iterates. Reliable decoding of 1-sparse sub-problems is achieved by 1) using carefully designed bin-measurement matrices $\boldsymbol{A}_{i,j}$ and 2) using a robust bin-processing/reconstruction algorithm.

The rest of the paper is organized as follows: In Section II, we provide the problem formulation along

with the signal and the noise model. Section III provides the main result of this paper. In Section IV, we provide a brief overview of the related literature and contrast it with the results of this paper. In Section V, we exemplify the connection between the sampling pattern of the FFAST front-end and the resulting bin-measurement matrices. Section VI describes the noise-robust version of the FFAST algorithm. In Section VII, we provide extensive simulations for various settings to empirically validate the performance of the FFAST algorithm in the presence of observation noise. Additionally, we also demonstrate an application of the FFAST framework for an MRI acquisition.

II. Signal model and Problem formulation

Consider an n-length discrete-time signal $\vec{x}$ that is a sum of $k << n$ complex exponentials, i.e., its n-length discrete Fourier transform has k non-zero coefficients:

$$x[p] = \sum_{q=0}^{k-1} X[\ell_q] e^{2\pi \imath \ell_q p/n}, \qquad p = 0, 1, \ldots, n-1, \tag{2}$$

where the discrete frequencies $\ell_q \in \{0, 1, \ldots, n-1\}$ and the amplitudes $X[\ell_q] \in \mathbb{C}$, for $q = 0, 1, \ldots, k-1$. We consider the problem of computing the k sparse n length DFT $\vec{X}$ when the observed time domain samples are corrupted by white Gaussian noise, i.e., we can only observe a subset of the samples of $\vec{y}$, where,

$$\vec{y} = \vec{x} + \vec{w}, \tag{3}$$

and $\vec{w} \in \mathcal{CN}(0, I_{n\times n})$.

Further, we make the following modeling assumptions on the signal:

- All the non-zero DFT coefficients have magnitude $\sqrt{\rho}$, where ρ is a desired signal-to-noise ratio, while the phase is chosen uniformly at random from the set $\{2\pi i/M\}_{i=0}^{M-1}$, for some fixed constant M[2].
- The support of the non-zero DFT coefficients is uniformly random in the set $\{0, 1, \ldots, n-1\}$.
- The signal-to-noise ratio ρ, is defined as,

$$\rho = \min_{X[\ell]\neq 0} \frac{|X[\ell]|^2}{\mathbb{E}\{||\vec{w}||^2\}}.$$

  Note, when $\vec{x}$ is 1-sparse, $\rho = \frac{||\vec{x}||^2}{\mathbb{E}\{||\vec{w}||^2\}}$.

[2]The FFAST framework and the reconstruction algorithm generalize to arbitrary complex-valued non-zero DFT coefficients, as long as all the non-zero DFT coefficients respect the signal-to-noise-ratio. The proof techniques for the arbitrary complex-valued non-zero DFT coefficients become much more cumbersome, and we do not address this issue in this paper.

III. Main results

The proposed FFAST algorithm reliably computes a k-sparse n-length DFT $\vec{X}$, of a signal from its noise-corrupted time-domain samples, $\vec{y} = \vec{x} + \vec{w}$ using $O(k \log n)$ samples and $O(n \log n)$ computations. A precise statement of the main result is given by the following theorem.

Theorem III.1. *For any given $0 < \varepsilon_p < 1$ and a finite signal-to-noise-ratio, there exist (infinitely many) sufficiently large n, such that the FFAST algorithm computes the k-sparse DFT $\vec{X}$, where $k = \Omega(n^\delta)$ and $0 < \delta < 1$, of an n-length signal $\vec{x}$ from its noise-corrupted samples $\vec{y}$, with the following properties:*

- **Sample complexity***: The algorithm needs $m = O(k \log(n))$ samples of $\vec{y}$.*
- **Computational complexity***: The computational complexity of the algorithm is $O(n \log(n))$.*
- **Probability of success***: The algorithm successfully recovers all the non-zero DFT coefficients of the signal $\vec{x}$, with probability at least $1 - \varepsilon_p$.*

Proof: Please see Appendix C. ■

IV. Related work

The problem of computing a sparse discrete Fourier transform of a signal is related to the rich literature of frequency estimation [2, 3, 4, 5] in statistical signal processing as well as compressive-sensing [6, 7]. In frequency estimation, it is assumed that a signal consists of k complex exponentials in the presence of noise and the focus is on 'super-resolution' spectral estimation techniques based on well-studied statistical methods like MUSIC and ESPRIT [2, 3, 4, 5]. The methods used are based on subspace decomposition principles, e.g., singular-value-decomposition, which very quickly become computationally infeasible as the problem dimensions (k, n) increase. In contrast, we take a different approach combining tools from coding theory, number theory, graph theory and statistical signal processing, to divide the original problem into many instances of simpler problems. The divide-and-conquer approach of FFAST alleviates the scaling issues in a much more graceful manner.

In compressive sensing, the bulk of the literature concentrates on random linear measurements, followed by either convex programming or greedy pursuit reconstruction algorithms [7, 8, 9]. A standard tool used for the analysis of the reconstruction algorithms is the *restricted isometry property* (RIP) [10]. The RIP characterizes matrices which are nearly orthonormal or unitary, when operating on sparse vectors. Although random measurement matrices like Gaussian matrices exhibit the RIP with optimal scaling, they have limited use in practice, and are not applicable to our problem of computing a sparse DFT from time-domain samples. So far, to the best of our knowledge, the tightest characterization of the

RIP, of a matrix consisting of random subset of rows of an $n \times n$ DFT matrix, provides a sub-optimal scaling, i.e., $O(k \log^3 k \log n)$, of samples [11]. In contrast, we show that by relaxing the worst case assumption on the input signal one can achieve a much better scaling of $O(k \log n)$ even for partial Fourier measurements. An alternative approach, in the context of sampling a continuous time signal with a finite rate of innovation is explored in [12, 13, 14, 15].

At a higher level though, despite some key differences in our approach to the problem of computing a sparse DFT, our problem is indeed closely related to the spectral estimation and compressive sensing literature, and our approach is naturally inspired by this, and draws from the rich set of tools offered by this literature.

A number of previous works [16, 17, 18, 19, 20] have addressed the problem of computing a 1-D DFT of a discrete-time signal that has a sparse Fourier spectrum, in sub-linear sample and time complexity. Most of these algorithms achieve a sub-linear time performance by first isolating the non-zero DFT coefficients into different bins, using specific filters or windows that have 'good' (concentrated) support in both, time and frequency. The non-zero DFT coefficients are then recovered iteratively, one at a time. Although the time complexity of these algorithms sub-linear, their robustness to observation noise is limited. In [21], the author proposes a sub-linear time algorithm with a sample complexity of $O(k \log^4 n)$ or $O(k^2 \log^4 n)$ and computational complexity of $O(k \log^5 n)$ or $O(k^2 \log^4 n)$ to compute a sparse DFT, with high probability or zero-error respectively. The algorithm in [21] exploits the Chinese-Remainder-Theorem (CRT), along with $O(poly(\log n))$ number of subsampling patterns to identify the locations of the non-zero DFT coefficients. In contrast, the FFAST algorithm exploits the CRT to induce 'good' sparse-graph codes using a small constant number of subsampling patterns and computes the sparse DFT with a vanishing probability of failure.

V. FFAST SAMPLING PATTERN AND THE MEASUREMENT MATRIX

In this section, we describe the connection between the sampling pattern of the FFAST front-end and the measurement matrix of a sparse recovery problem. Consider again the example signal $\vec{x}$ of length $n = 20$, whose 20-point DFT $\vec{X}$ is $k = 5$ sparse. The 5 non-zero DFT coefficients of $\vec{x}$ are $X[1]$, $X[3]$, $X[5]$, $X[10]$ and $X[13]$. The noise-corrupted signal $\vec{y} = \vec{x} + \vec{w}$, is processed through stage 0 of a 2-stage FFAST sub-sampling front-end as shown in Fig. 2. The output samples of the FFAST front-end are grouped into 'bins', as shown in Fig. 2. Using the basic signal processing identities of sampling-aliasing and circular shifts, the relation between the bin-observation vectors and the DFT coefficients of the input signal $\vec{x}$, can be computed, e.g., see equation (1).

More generally, the observation vector $\vec{y}_{b,i,j}$ of bin j of stage i in a FFAST front-end architecture with d stages and D delays is given as,

$$\vec{y}_{b,i,j} \quad = \quad \boldsymbol{A}_{i,j}\vec{X} + \vec{w}_{b,i,j}, \quad 0 \leq i < d, \; 0 \leq j < D, \tag{4}$$

where $\vec{w}_{b,i,j} \sim \mathcal{CN}(0, I_{D\times D})$ and $\boldsymbol{A}_{i,j}$ is the *bin-measurement* matrix. Next, we describe the generic structure of bin-measurement matrices for different bins.

A. Bin-measurement matrix

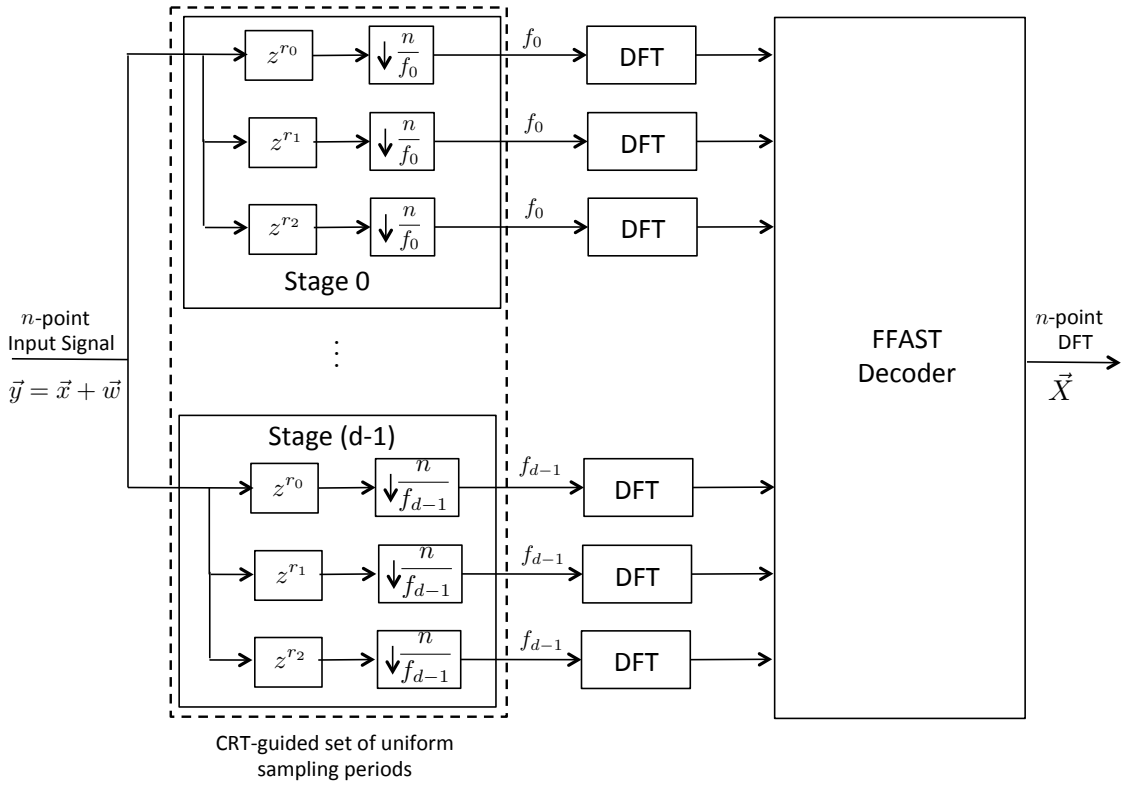


Fig. 3. Schematic block diagram of the FFAST architecture for processing noise-corrupted samples $\vec{y} = \vec{x} + \vec{w}$. The n-point input signal $\vec{y}$ is uniformly subsampled by a carefully chosen, guided by the Chinese-Remainder-Theorem, set of d patterns. Each stage has $D = 3$ delay sub streams and a sub stream in stage i has f_i number of samples. The delay shift pattern (r_0, r_1, r_2) is carefully chosen so as to induce bin-measurement matrices with "good" mutual incoherence and RIP properties (as explained in Section V-A). Next, the (short) DFTs, of each of the sub-streams are computed using an efficient FFT algorithm of choice. The big n-point DFT $\vec{X}$ is then synthesized from the smaller DFTs using the peeling-like FFAST decoder.

Consider the FFAST front-end architecture that processes the noise-corrupted observations $\vec{y} = \vec{x} + \vec{w}$, with d stages and D delay sub-streams per stage as shown in Fig. 3. For the case when the observations

are noiseless, $D = 2$ is sufficient to reconstruct the sparse DFT $\vec{X}$ from the sub-sampled data. However when the observed samples are corrupted by *noise* $D > 2$ is necessary. As shown in Fig. 3, the i^{th} delay sub-stream of the j^{th} stage, circularly shifts the input signal by r_i and then sub-samples by a sampling period of n/f_j, where f_j is the number of samples in each delay sub-stream of stage j. The number of delays D and the shift pattern $(r_0, \ldots, r_{D-1})$ influences bin-measurement matrices as shown below.

Let $\vec{a}(\ell)$, be a discrete D-dimensional complex vector given by:

$$\vec{a}(\ell) = \begin{pmatrix} e^{\imath 2\pi \ell r_0/n} \\ e^{\imath 2\pi \ell r_1/n} \\ \vdots \\ e^{\imath 2\pi \ell r_{D-1}/n} \end{pmatrix}, \tag{5}$$

in the sequel, we refer to $\vec{a}(\ell),\ \ell = 0, 1 \ldots, n-1$, as a *steering-vector* of frequency $2\pi\ell/n$ sampled at $(r_0, \ldots, r_{D-1})$. Then, bin-measurement matrix $\boldsymbol{A}_{i,j} \in \mathbb{C}^{D \times n}$ of bin j of stage i is given by,

$$\vec{\boldsymbol{A}}_{i,j}(\ell) = \begin{cases} \vec{a}(\ell) & \text{if } \ell \equiv j \bmod f_i \\ \vec{0} & \text{other wise,} \end{cases} \tag{6}$$

where $\vec{\boldsymbol{A}}_{i,j}(\ell)$ is the ℓ^{th} column, $\ell = 0, \ldots, n-1$, of $\boldsymbol{A}_{i,j}$. For example in equation (1) bin-measurement matrix $\boldsymbol{A}_{0,1}$ is,

$$\boldsymbol{A}_{0,1} = \begin{bmatrix} \vec{0} & \vec{a}(1) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(5) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(9) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(13) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(17) & \vec{0} & \vec{0} \end{bmatrix},$$

where $\vec{a}(\ell) = \begin{pmatrix} 1 \\ e^{\imath 2\pi \ell/20} \end{pmatrix}$ for $\ell = 0, 1, \ldots, 19$.

Thus, the FFAST front-end effectively divides the problem of recovering a k-sparse DFT $\vec{X}$ of a n-length signal $\vec{x}$ from its noise-corrupted samples $\vec{y}$ into multiple bin-level problems of the form (4). The FFAST peeling-decoder then detects which of these bin-level problems are "single-tons", solves them (identifize the support and the value of the single-ton) using bin observation $\vec{y}_{b,i,j}$ and "peels-off" their contribution from other bin observations to create more singleton bins. In the case when the observations were not corrupted by noise, $D = 2$ measurements per bin were sufficient. However, in the presence of the observation noise we need more measurements per bin, i.e., $D > 2$, to make bin-processing reliable and robust against noise. Moreover, the structure of the delays, i.e., the choice of the circular shifts $(r_0, \ldots, r_{D-1})$, also plays a crucial role in the design of the measurement matrices of the individual bins as shown in (6), which is further used to make the individual bin-processing robust against the

observation noise.

a) Incoherence properties of bin-measurement matrices: In the compressed sensing literature the *mutual-incoherence* and the *Restricted-isometry-property* (RIP) [10], of the measurement matrix, are widely studied and well understood to play an important role in stably recovering a high-dimensional sparse vector from linear measurements in the presence of observation noise. The problem in (4) is a special case of the compressed sensing problem. In particular, if the processed bin is a single-ton, then the bin-processing algorithm attempts to recover a 1-sparse high-dimensional signal from linear measurements in the presence of the observation noise. We establish the *mutual-incoherence* and the RIP of bin-measurement matrices and use them to analyze the noise-robustness of the FFAST algorithm. Next, we define the mutual-incoherence and the RIP for a general measurement matrix $\boldsymbol{A}$.

Definition V.1. *The mutual incoherence $\mu_{\max}(\boldsymbol{A})$ of a measurement matrix $\boldsymbol{A}$ is defined as*

$$\mu_{\max}(\boldsymbol{A}) \triangleq \max_{\forall p \neq q} \frac{|\vec{\boldsymbol{A}}(p)^{\dagger} \vec{\boldsymbol{A}}(q)|}{||\vec{\boldsymbol{A}}(p)|| \cdot ||\vec{\boldsymbol{A}}(q)||}, \tag{7}$$

where $\vec{\boldsymbol{A}}(p)$ is the p^{th} column of the matrix $\boldsymbol{A}$.

The mutual-incoherence property of the measurement matrix indicates the level of correlation between the distinct columns of the measurement matrix. Smaller value of $\mu_{\max}(\boldsymbol{A})$ implies more stable recovery, e.g., $\mu_{\max}(\boldsymbol{A}) = 0$ for an orthogonal measurement matrix $\boldsymbol{A}$.

Definition V.2. *The restricted-isometry constant $\gamma_s(\boldsymbol{A})$ of a measurement matrix $\boldsymbol{A}$, with unit norm columns, is defined to be the smallest positive number that satisfies*

$$(1 - \gamma_s(\boldsymbol{A}))||\vec{X}||^2 \leq ||\boldsymbol{A}\vec{X}||^2 \leq (1 + \gamma_s(\boldsymbol{A}))||\vec{X}||^2, \tag{8}$$

for all $\vec{X}$ such that $||\vec{X}||_0 \leq s$.

The RIP characterizes the norm-preserving capability of the measurement matrix when operating on sparse vectors. If the measurement matrix $\boldsymbol{A}$ has a good RIP constant (small value of $\gamma_{2s}(\boldsymbol{A})$) for all $2s$-sparse vectors, then a stable recovery can be performed for any s-sparse input vector [22]. Since, a small value of $\gamma_{2s}(\boldsymbol{A})$ implies that $||\boldsymbol{A}(\vec{X}_1 - \vec{X}_2)||^2$ is bounded away from zero for any two distinct, $\vec{X}_1 \neq \vec{X}_2$, s-sparse vectors.

Thus stable recovery performance of the FFAST algorithm is a function of the constants $\mu_{\max}$ and γ_s. From equation (6) it is clear that these constants dependent on the number of delays D as well as on the

choice of the delay pattern $(r_0, \ldots, r_{D-1})$, i.e., sub-sampling patterns of the FFAST front-end. There are multiple options for choosing the delay patterns, e.g., uniform periodic delays, aperiodic deterministic delays, random delays, etc. We would like to choose a delay pattern that has good constants $\mu_{\max}$ and γ_s, to guarantee a stable recovery performance. Next, we describe the delay pattern used in the FFAST architecture and establish its mutual incoherence and RIP properties.

B. FFAST sampling patterns

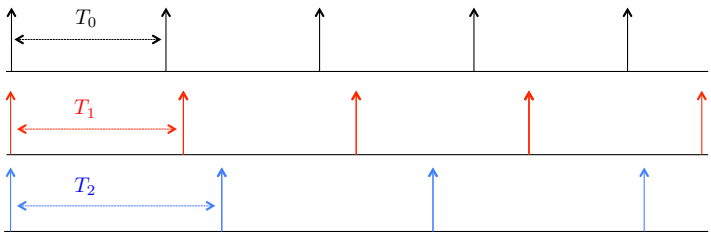


Fig. 4. Principal sampling pattern, corresponding to the $r_0 = 0$ delay, of the FFAST front-end with $d = 3$ stages. The sampling periods of the three stages are T_0, T_1, T_2 respectively. The overall sampling pattern is a union of periodic sampling patterns.

A FFAST front-end architecture with d-stages has $\{f_i\}_{i=0}^{d-1}$ samples per sub-stream in each of the d stages (see Fig. 3). For a given (k, n), in [1], we show how to choose the parameters $\{f_i\}_{i=0}^{d-1}$ using the CRT. As an example consider a $d = 3$ stage FFAST front-end architecture such that the sampling period of stage $i = 0, 1, 2$, is $T_i = n/f_i$. Without loss of generality we assume that $r_0 = 0$. This results in a basic sampling pattern, as shown in Fig. 4, that we refer to as "principal sampling pattern" in the sequel. We use the principal sampling pattern in conjunction with the choice of the remaining $D - 1$ delays $(r_1, \ldots, r_{D-1})$ to construct the overall FFAST sampling pattern. For example, sampling pattern of a $d = 3$ stages and $D = 3$ uniformly random delays, (r_0, r_1, r_2), is shown in Fig. 7.

Next, we characterize the mutual in-coherence and the RIP properties of the *bin-measurement* matrices, of the FFAST architecture, when delays are chosen uniformly at random from the set $\{0, \ldots, n-1\}$. In the sequel we use these properties to prove the stable recovery of the proposed FFAST peeling-style iterative recovery algorithm.

Lemma V.3. *The mutual incoherence $\mu_{\max}(\boldsymbol{A}_{i,j})$ of the bin-measurement matrix $\boldsymbol{A}_{i,j}$, of the FFAST front-end with D delays, is upper bounded by*

$$\mu_{\max}(\boldsymbol{A}_{i,j}) < 2\sqrt{\log(5n)/D}\ , \ \forall\ i, j \tag{9}$$

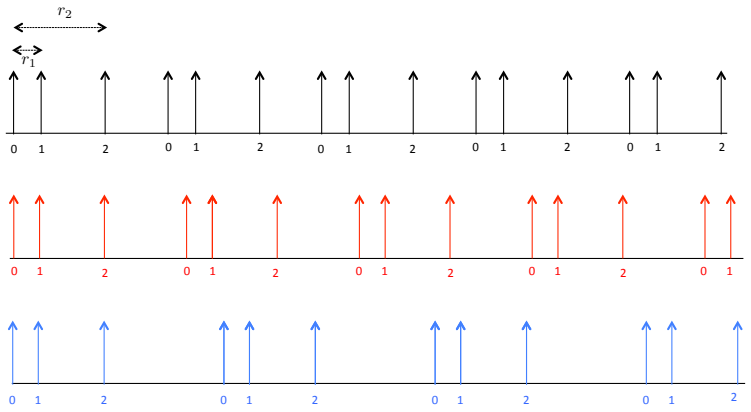


Fig. 5. Sampling pattern of the FFAST front-end with $d = 3$ stages and $D = 3$ delay sub-stream per stage. The sampling periods of the three stages are T_0, T_1, T_2 respectively. Sampling impulse train corresponding to the delay sub stream 0 is indexed by 0 and so on. The delay pattern (r_0, r_1, r_2) is chosen uniformly at random from the set $\{0, 1 \ldots, n-1\}$. For sake of representation WLOG $r_0 = 0$.

with probability at least 0.2*, where the delays* $(r_0, \ldots, r_{D-1})$ *are chosen uniformly at random from the set* $\{0, 1, \ldots, n-1\}$.

Proof: Please see Appendix A. ■

Thus, for a random choice of delay pattern $r_0, \ldots, r_{D-1}$, the coherence parameter $\mu_{\max}(\boldsymbol{A}_{i,j})$ satisfies the bound in (9) with probability at least 0.2. Also, it is easy, i.e., $O(nD)$ complexity, to verify if a given choice of $r_0, \ldots, r_{D-1}$ satisfies (9) or not. Hence, using offline processing one can choose a pattern $(r_0, \ldots, r_{D-1})$, such that deterministically the bound in (9) is satisfied.

Lemma V.4. *The bin-measurement matrix* $\boldsymbol{A}_{i,j}$*, of the FFAST front-end with* D *delays, satisfies the following RIP condition for all* $\vec{X}$ *that have* $||\vec{X}||_0 \leq s$*,*

$$D\left(1 - \mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)^+ ||\vec{X}||^2 \leq ||\boldsymbol{A}_{i,j}\vec{X}||^2 \leq D\left(1 + \mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)||\vec{X}||^2 \ , \ \forall \ i,j \qquad (10)$$

where the delays $(r_0, \ldots, r_{D-1})$ *are chosen uniformly at random from the set* $\{0, 1, \ldots, n-1\}$.

Proof: Please see Appendix B ■

VI. Noise robust FFAST algorithm

A generic structure of the FFAST architecture, that processes noise-corrupted observations $\vec{y} = \vec{x} + \vec{w}$, is shown in Fig. 3. In [1] we described in detail how to carefully design the sub-sampling parameters, i.e., number of stages d and samples per sub-stream $\{f_i\}_{i=0}^{d-1}$ in each of the d stages, of the FFAST

front-end architecture, using the CRT, for a given sparsity-index δ. For the case when there is no noise in the observations, we have shown that $D = 2$ is sufficient to reconstruct the sparse DFT $\vec{X}$ from the sub-sampled data. However when the observed samples are corrupted by *noise* as explained in Section V, we use $D > 2$ number of delay sub-streams per stage. As shown in Fig. 3, the i^{th} delay sub-stream of the j^{th} stage, circularly shifts the input signal by r_i and then sub-samples by a sampling period of n/f_j, where f_j is the number of samples in each delay sub-stream of stage j. The number of delays D and the shift pattern $(r_0, \ldots, r_{D-1})$ influences bin-measurement matrices as shown in Section V-A. We choose the delay pattern $(r_0, r_1, \ldots, r_{D-1})$ such that bin-measurement matrices have good mutual incoherence and RIP properties.

From algorithmic perspective the only difference between the presence and the absence of the observation noise is in the thresholds used by the algorithm and the subroutine 'Singleton-Estimator' that processes individual bin to test if it is a single-ton. The pseudocode of the noise-robust FFAST algorithm is provided in Algorithm 1 and Algorithm 2.

Algorithm 1 FFAST Algorithm

Input: The noise-corrupted bin observations $\vec{y}_{b,i,j}$ obtained through the FFAST front-end of Fig. 3, for each bin j in stage i for all i, j.

Output: An estimate $\vec{X}$ of the k-sparse n-point DFT.

FFAST Decoding: Set the initial estimate of the n-point DFT $\vec{X} = 0$. Let ℓ denote the number of iterations performed by the FFAST decoder.
Set the energy threshold $T = (1 + \gamma)D$ for appropriately chosen γ (see Appendix C).
for each iteration **do**
  for each stage i **do**
    for each bin j **do**
      if $||\vec{y}_{b,i,j}||^2 < T$ **then**
        bin is a *zero-ton*.
      else
        (singleton, v_p, p) = ***Singleton-Estimator*** $(\vec{y}_{b,i,j})$.
        if singleton = 'true' **then**
          Peel-off: $\vec{y}_{b,s,q} = \vec{y}_{b,s,q} - v_p \vec{a}(p)$, for all stages s and bins $q \equiv p \bmod f_q$.
          Set, $X[p] = v_p$.
        else
          bin is a *multi-ton*.
        end if
      end if
    end for
  end for
end for

Algorithm 2 Singleton-Estimator

Inputs: The noise-corrupted bin observation $\vec{y}_{b,i,j}$ and indices (i, j) indicating the stage and the bin number respectively.

Outputs: 1) A boolean flag 'singleton', 2) Estimate of the value v_p of the non-zero DFT coefficient and 3) the position p of the non-zero DFT coefficient.

Singleton-Estimator:
Set the singleton = 'false'.
Set the energy threshold $T = (1 + \gamma)D$ for appropriately chosen γ (see Appendix C).
for each position $q \equiv i \bmod f_j$ for all $q = 0, \ldots, n-1$ **do**
  $v_q = \vec{a}(q)^\dagger \vec{y}_{b,i,j}/D$.
  if $||\vec{y}_{b,i,j} - v_q \vec{a}(q)||^2 < T$ **then**
    singleton = 'true'.
    $p = q$ and $v_p = v_q$.
  end if
end for

VII. Simulations

In Section I, we have shown that the FFAST algorithm performs well even for signals like MR images, that have an approximately sparse Fourier spectrum with a non-uniform support for the dominant DFT coefficients. In this section we evaluate the performance of the FFAST algorithm on synthetic data and show that the empirical results are in close confirmation with the theoretical claims of Theorem III.1.

The FFAST front-end uses CRT-guided sub-sampling pattern to induce LDPC-like sparse-graph in the frequency-domain, which effectively divides the original problem of computing a k-sparse DFT $\vec{X}$, into multiple simpler (mostly 1-sparse) *bin level* problems. Next, we use techniques from *signal processing* and *estimation theory* to robustly solve these individual bin-level problems, in conjunction with iterative peeling. Using this modular approach of solving the problem, the analytical result of Theorem III.1 can be interpreted as follows,

$$\begin{aligned} & \#\text{ of samples } m \text{ required by FFAST} \\ = \; & \{\#\text{ of bins required for successful decoding} \\ & \text{of the resulting sparse graph}\} \times \\ & \{\#\text{ of delays required for robust bin-processing}\} \\ = \; & \{c_1(\delta)k\} \times \{c_2(\rho)c_3 \log n\} \end{aligned} \tag{11}$$

where the constant $c_1(\delta)$ depends on the sparsity index $0 < \delta < 1$ and the constant $c_2(\rho)$ is a function of signal-to-noise-ratio ρ. In [1] we have shown theoretically and validated empirically that $c_1(\delta) < 2,\ \forall\ 0 < \delta < 0.99$. In this section we empirically evaluate the scaling of the number of samples m as a function of the ambient signal dimension n, required by the FFAST algorithm to reliably compute the DFT $\vec{X}$.

A. Sample complexity m as a function of n

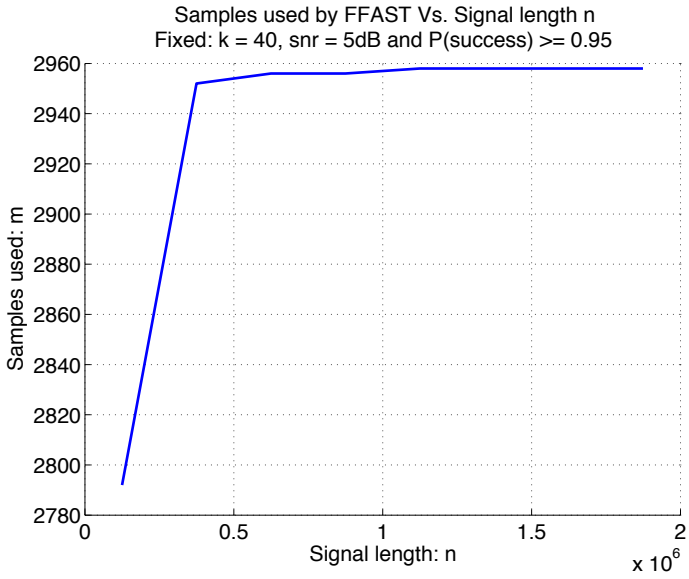


Fig. 6. The plot shows the scaling of the number of samples m required by the FFAST algorithm to reliably reconstruct a $k = 40$ sparse DFT $\vec{X}$, from noise-corrupted observations $\vec{y} = \vec{x} + \vec{w}$, for increasing signal length n. For a fixed reliability, signal-to-noise ratio and sparsity, we note that m scales logarithmically with increasing n. This is consistent with the claims of Theorem III.1.

In this section we empirically validate the scaling of the number of measurements m required by the FFAST algorithm to reliably compute the DFT $\vec{X}$ for various signal lengths n.

Simulation Setup*:*

- An n-length DFT vector $\vec{X}$ with $k = 40$ non-zero coefficients is generated. The non-zero DFT coefficients are chosen to have uniformly random support from the set $\{0, \ldots, n-1\}$, with values from the set $\{\pm\sqrt{\rho}\}$[3] randomly. The input signal $\vec{x}$ is obtained by taking the IDFT of $\vec{X}$. The length

[3]We have also simulated the case where the amplitudes of the non-zero DFT coefficients are arbitrary complex numbers, with fixed magnitude and uniformly random phase, and obtained similar performance plots. In this section, we provide the simulation results for the case when the non-zero DFT coefficients take antipodal values, only to be consistent with the theoretical analysis of the paper.

of the signal n is varied from $n = 49*50*51 \approx 0.1$ million to $n = 49*50*51*15 \approx 1.87$ million. The choice of the ambient dimension n, to be a product of approximately equal sized relatively co-prime numbers, is induced by the Chinese-Remainder-Theorem.

- The noise-corrupted signal $\vec{y} = \vec{x} + \vec{w}$, is generated by adding zero-mean, unit variance complex white Gaussian noise $\vec{w}$ with ρ chosen to have signal-to-noise ratio of 5dB.
- The noise-corrupted signal $\vec{y}$ is input to a $d = 3$ stage FFAST architecture with D sub-streams per stage (see Fig. 3). The sampling periods of the 3 stages in the FFAST architecture are $50*51$, $51*49$ and $49*50$ respectively. This results in the number of samples per sub-stream, for the three stages to be $f_0 = 49$, $f_1 = 50$ and $f_2 = 51$ respectively. Thus the total number of bins in the FFAST front-end architecture is $n_b = \sum_{i=0}^{2} f_i = 150 = 3.75k$, i.e., $c_1(\delta) = 3.75$.
- As the signal length n varies the number of delays D per bin are varied to have reliable decoding, i.e., $Pr(\text{success}) \geq 0.95$. For each value of D, the total number of samples used by the FFAST are $m = n_b * D = 3.75 * k * D$.
- Each sample point in the plot is generated by averaging the performance of FFAST over 1000 runs for each configuration.

We note that the number of samples, in particular D, increase *logarithmically* with increasing n. Thus, empirically confirming the claims of Theorem III.1.

B. Sample complexity m as a function of ρ

In this section we experimentally probe the scaling of the number of measurements m as a function of signal-to-noise-ratio ρ, i.e., $c_2(\rho)$.

Simulation Setup*:*

- An $n = 49 * 50 * 51 \approx 0.1$ million length, DFT vector $\vec{X}$, with $k = 40$ non-zero coefficients is generated. The non-zero DFT coefficients are chosen to have uniformly random support from the set $\{0, \ldots, n-1\}$, and the values from the set $\{\pm\sqrt{\rho}\}$. The input signal $\vec{x}$ is obtained by taking IDFT of $\vec{X}$.
- A noise-corrupted signal $\vec{y} = \vec{x} + \vec{w}$, is generated by adding zero-mean, unit variance complex white Gaussian noise $\vec{w}$.
- The parameter ρ is varied such that the effective signal-to-noise ratio ranges from -1dB to 9dB.
- The noise-corrupted signal $\vec{y}$ is input to a $d = 3$ stage FFAST architecture with D sub-streams per stage (see Fig. 3). The sampling periods of the 3 stages in the FFAST architecture are $50*51$, $51*49$ and $49*50$ respectively. This results in the number of samples per sub-stream, for the three stages

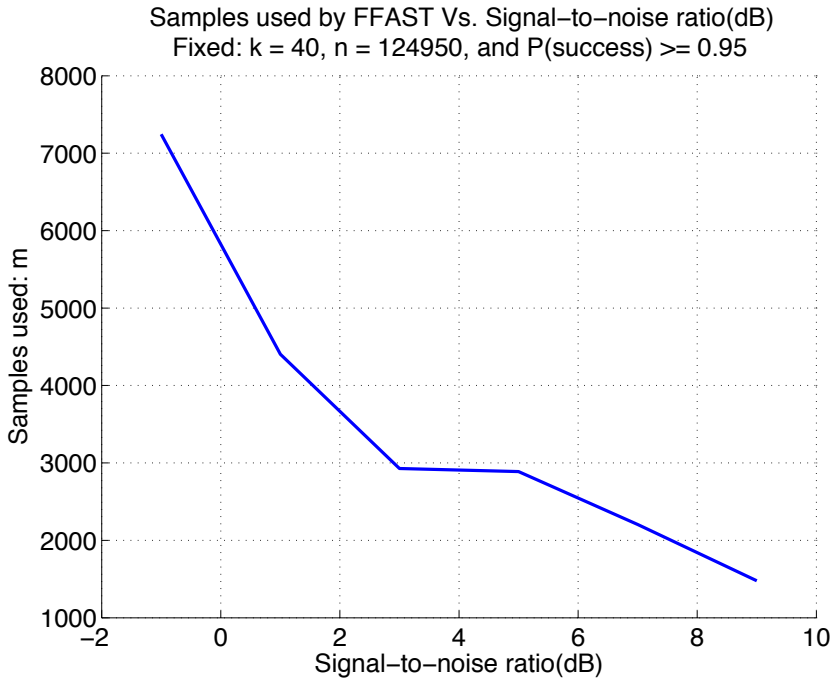


Fig. 7. The plot shows the scaling of the number of samples m required by the FFAST algorithm to reliably compute a $n \approx 0.1$ million, $k = 40$ sparse DFT $\vec{X}$, from noise-corrupted observations $\vec{y} = \vec{x} + \vec{w}$, for increasing signal-to-noise ratio ρ. For fixed values of all other parameters, the number of samples decreases roughly in an *inverse relation* with the increasing signal-to-noise ratio on log-scale.

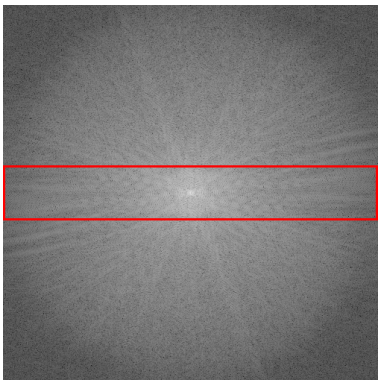
to be $f_0 = 49$, $f_1 = 50$ and $f_2 = 51$ respectively. Thus the total number of bins in the FFAST front-end architecture is $n_b = \sum_{i=0}^{2} f_i = 150 = 3.75k$, i.e., $c_1(\delta) = 3.75$.

- As the signal-to-noise ratio varies the number of delays D per bin are varied to have reliable decoding, i.e., $Pr(\text{success}) \geq 0.95$.
- Each sample point in the plot is generated by averaging the performance of FFAST over 1000 runs for each configuration.

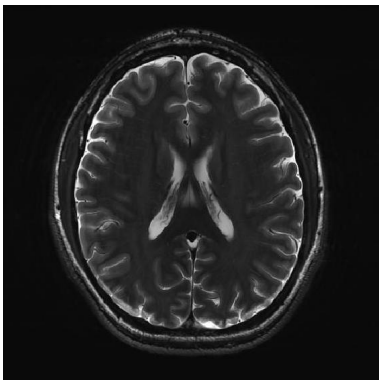
The number of samples decrease roughly in an *inverse relation* with increasing signal-to-noise ratio on log-scale.

C. Application of the FFAST for MR imaging

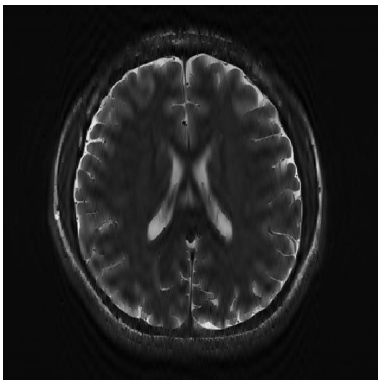
The $1D$-FFAST architecture proposed in this paper can be generalized, in a straightforward manner, to $2D$ signals, with similar performance guarantees. In this section, we apply the $2D$-FFAST algorithm to reconstruct a brain image acquired on an MR scanner with dimension 504×504. In MR imaging, the samples are acquired in the Fourier domain and the task is to reconstruct the spatial image domain signal



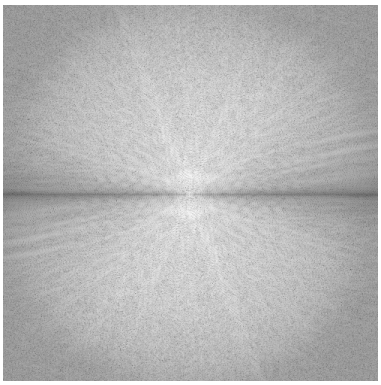
(a) Log intensity plot of the $2D$-DFTof the original 'Brain' image. The red enclosed region is fully sampled and used for the stable inversion.



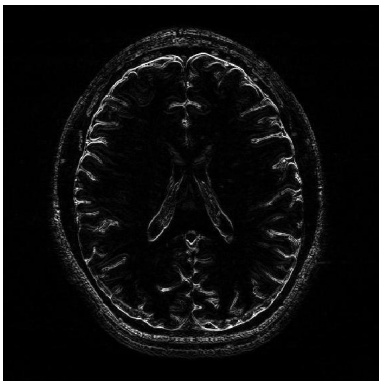
(b) Original 'Brain' image in spatial domain.



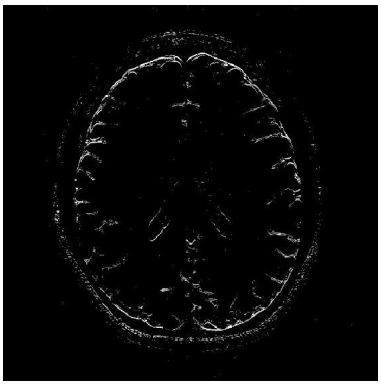
(c) Reconstructed 'Brain' image using the $2D$ FFAST architecture along with the fully sampled center frequencies. The total number of Fourier samples used is 60.18%.



(d) Log intensity plot of $2D$-DFT of the original 'Brain' image, after application of the vertical difference operation.



(e) Differential 'Brain' image obtained using the vertical difference operation on the original 'Brain' image.



(f) Differential 'Brain' image reconstructed using the $2D$ FFAST architecture with 56.71% of Fourier samples.

Fig. 8. Application of the $2D$-FFAST algorithm to reconstruct the 'Brain' image acquired on an MR scanner with dimension 504×504. We first reconstruct the differential 'Brain' image shown in Fig. 8(e), using $d = 3$ stage $2D$-FFAST architecture with 15 random delays in each of the 3 stages of the FFAST architecture. Additionally we acquire all the Fourier samples from the center frequency as shown, by the red enclosure, in Fig. 8(a). Then, we do a stable inversion using the reconstructed differential 'Brain' image of Fig. 8(f) and the fully sampled center frequencies of Fig. 8(a), to get a reconstructed full 'Brain' image as shown in Fig. 8(c). Our proposed two-step acquisition and reconstruction procedure takes overall 60.18% of Fourier samples.

from significantly less number of Fourier samples. To reconstruct the full brain image using $2D$-FFAST, we perform the following two-step procedure:

- *Differential space signal acquisition*: We perform a vertical finite difference operation on the image by multiplying the $2D$-DFT signal with $1-e^{2\pi \imath \omega_0}$. This operation effectively creates an approximately sparse differential image in spatial domain, as shown in Fig. 8(e), and can be reconstructed using FFAST. Note, that the finite difference operation can be performed on the sub-sampled data, and at

no point do we need to access all the input Fourier samples. The differential brain image is then sub-sampled and reconstructed using a 3-stage $2D$-FFAST architecture. Also, since the brain image is approximately sparse, we take 15 delay sub-streams in each of the 3 stages of the $2D$-FFAST architecture, instead of 3 delay sub-streams as in the exactly sparse case. The FFAST algorithm reconstructs the differential brain image using 56.71% of Fourier samples.

- *Inversion using fully sampled center frequencies*: After reconstructing the differential brain image, as shown in Fig. 8(f), we invert the finite difference operation by dividing the $2D$-DFT samples with $1 - e^{2\pi\imath\omega_0}$. Since the inversion is not stable near the center of the Fourier domain, only the non-center frequencies are inverted and the center region is replaced by the fully sampled data.
- Overall, the $2D$-FFAST algorithm uses a total of 60.18% of Fourier samples to reconstruct the brain image shown in Fig. 8(c).

APPENDIX A
MUTUAL INCOHERENCE BOUND

In this section we provide a proof of Lemma V.3. Let $\vec{a}(p)$ be a D dimensional steering vector with frequency $2\pi p/n$, for $p = 0, \ldots, n-1$, as given in equation (5). Then,

$$\begin{aligned} \mu_{\max}(\boldsymbol{A}_{i,j}) &\leq \max_{p \neq q} \frac{1}{D} |\vec{a}(p)^\dagger \vec{a}(q)| \\ &= \max_{\ell \neq 0} \frac{1}{D} |\sum_{s=0}^{D-1} \exp(\imath 2\pi \ell r_s/n)| \\ &= \max_{\ell \neq 0} \mu(\ell) \ , \end{aligned} \tag{12}$$

where $\mu(\ell) \triangleq |\sum_{s=0}^{D-1} \exp(\imath 2\pi \ell r_s/n)|/D$.

Now consider the summation $\sum_{s=0}^{D-1} \cos(\imath 2\pi \ell r_s/n)/D$ for any fixed $\ell \neq 0$. Each term in the summation is a zero-mean random variable i.i.d with bounded support in $[-1/D, 1/D]$. Thus, using Hoeffding's inequality for the sum of independent random variables with bounded support we have, for any $t > 0$,

$$Pr(|\sum_{s=0}^{D-1} \cos(\imath 2\pi \ell r_s/n)/D| > t) \leq 2\exp(-t^2 D/2).$$

Similarly,

$$Pr(|\sum_{s=0}^{D-1} \sin(\imath 2\pi \ell r_s/n)/D| > t) \leq 2\exp(-t^2 D/2).$$

Applying a union bound over the real and the imaginary parts of the summation term in $\mu(\ell)$, we get,

$$Pr(\mu(\ell) > \sqrt{2}t) \leq 4\exp(-t^2 D/2). \tag{13}$$

Further applying a union bound over all $\ell \neq 0$, we have,

$$\begin{aligned} Pr(\mu_{\max}(\boldsymbol{A}_{i,j}) > \sqrt{2}t) &\leq 4n\exp(-t^2 D/2) \\ &= 0.8 \end{aligned}$$

for $t = \sqrt{2\log(5n)/D}$. Thus, over all the random choices of the delays $r_0, \ldots, r_{D-1}$, with probability at least 0.2,

$$\mu_{\max}(\boldsymbol{A}_{i,j}) < 2\sqrt{\log(5n)/D}\ , \quad \forall\ i, j.$$

■

APPENDIX B
RESTRICTED-ISOMETRY-PROPERTY

Consider a s-sparse vector $\vec{X}$ and a measurement matrix $\boldsymbol{A}_{i,j}$ corresponding to bin j of stage i of the FFAST. Using basic linear algebra we get the following inequality,

$$\lambda_{\min}(\boldsymbol{A}_{i,j}^{\dagger}\boldsymbol{A}_{i,j})||\vec{X}||^2 \leq ||\boldsymbol{A}_{i,j}\vec{X}||^2 \leq \lambda_{\max}(\boldsymbol{A}_{i,j}^{\dagger}\boldsymbol{A}_{i,j})||\vec{X}||^2$$

The Gershgorin circle theorem [23], provides a bound on the eigen-values of a square matrix. It states that, every eigen-value of a square matrix lies within at least one of the Gershgoring discs. A Gershgorin disc is defined for each row of the square matrix, with diagonal entry as a center and the sum of the absolute values of the off-diagonal entries as radius. Note that for all (i, j), irrespective of the values of the delays $r_0, \ldots, r_D$, the diagonal entries of the matrix $\boldsymbol{A}_{i,j}^{\dagger}\boldsymbol{A}_{i,j}$ are equal to D. Hence, $D\mu_{\max}(\boldsymbol{A}_{i,j})$ provides an upper bound on the absolute values of the off-diagonal entries of $\boldsymbol{A}_{i,j}^{\dagger}\boldsymbol{A}_{i,j}$.

Then, using the Gershgorin circle theorem we have,

$$D\left(1 - \mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)_{+} ||\vec{X}||^2 \leq ||\boldsymbol{A}_{i,j}\vec{X}||^2 \leq D\left(1 + \mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)||\vec{X}||^2.$$

■

APPENDIX C
PROOF OF THEOREM III.1

In this section we provide a proof of the main result of the paper. The proof consists of two parts. In the first part of the proof, we show that the FFAST reconstructs the DFT $\vec{X}$, with probability at least $1-\varepsilon_p$ for any $\varepsilon_p > 0$, using $m = O(k\log(n))$ samples. In the second part of the proof we show that computational complexity of the FFAST decoder is $O(n\log(n))$.

A. Reliability Analysis and sample complexity of the FFAST

1) Sample complexity: In [1], we have shown that for the noiseless case, for any given $0 < \delta < 1$ and sufficiently large (k, n), the FFAST algorithm computes the k-sparse n-length DFT $\vec{X}$, with probability at least $1-O(1/k)$, using a total of $O(k)$ number of bins. Later in Section C-A2 we show that $D = O(\log n)$ number of samples per bin are sufficient to make Algorithm 2 robust against the observation noise. Hence, the total sample complexity of the FFAST algorithm in the presence of the observation noise is

$$m = O(k\log n).$$

2) Reliability analysis: In Lemma V.3, we have shown that a random choice of delay parameters $r_0, \ldots, r_{D-1}$ satisfies the upper bound of (9) with probability at least 0.2. Also, it is easy, i.e., $O(nD)$ complexity, to verify if for a given choice of $r_0, \ldots, r_{D-1}$ (9) is satisfied or not. Hence WLOG henceforth we assume that the $r_0, \ldots, r_{D-1}$ are chosen such that deterministically the bound in (9) is satisfied.

Let E_b be an event that *a bin* processed by the Algorithm 2 is decoded wrongly. We first show that the Algorithm 2 processes each bin reliably, i.e., $Pr(E_b) < O(1/k^2)$, using $D = O(\log n)$ number of samples per bin. Then, we show that an event E that *some bin* is wrongly decoded by the FFAST algorithm, while reconstructing the DFT $\vec{X}$ has a low probability. In particular, we use a union bound over the constant number of iterations required for the FFAST to reconstruct the DFT $\vec{X}$ and over $O(k)$ bins used in the FFAST architecture to get,

$$\begin{aligned} Pr(E) &< \text{number of iterations} \times \text{number of bins} \times Pr(E_b) \\ &< O(1/k). \end{aligned} \tag{14}$$

Let E_f denote an event that the FFAST algorithm fails to reconstruct the DFT $\vec{X}$. Then putting the

pieces together, we get,

$$\begin{aligned} Pr(E_f) &< Pr(E) + Pr(E_f \mid \overline{E}) \\ &\overset{(a)}{<} O(1/k) + O(1/k) \\ &< \varepsilon_p \end{aligned} \tag{15}$$

where in (a) we used the bound from (14) and the results from [1] along with the fact that if there is no error in any bin-processing, the FFAST algorithm performs as if it has access to the noiseless observations.

Hence, in order to complete the reliability analysis of the FFAST algorithm we need to show that $D = O(\log n)$ samples per bin are sufficient to achieve $Pr(E_b) < O(1/k^2)$. The following lemma, that analyzes the performance of an energy-based threshold rule to detect the presence of a complex vector in the presence of noise, plays a crucial role in the analysis of the event E_b.

Lemma C.1. *For a complex vector $\vec{u} \in \mathbb{C}^D$ and $\vec{w} \sim \mathcal{CN}(0, I_{D\times D})$, we have,*

$$Pr(||\vec{u} + \vec{w}||^2 < (1+\gamma)D) < 2e^{-D\gamma^2/9} + e^{-(||\vec{u}|| - \sqrt{2\gamma D})_+^2}, \tag{16}$$

for any constant $0 < \gamma < 1$ and $D > 1.5/\gamma$.

Proof: Please see Appendix D ■

Without loss of generality consider processing of bin j from stage i of the FFAST architecture. As shown in the Section V bin observation noise is $\mathcal{CN}(0, I_{D\times D})$. Bin j in stage i is either a zero-ton bin, or a single-ton bin or a multi-ton bin. We analyze all the three events below and show that irrespective of the type, the Algorithm 2 decodes bin successfully with probability at least $1 - O(1/k^2)$, as long as $D = O(\log n)$.

a) Analysis of a zero-ton bin: Consider a zero-ton bin with an observation $\vec{y}_{b,i,j} = \vec{w}_{b,i,j}$. Let E_z be an event that a zero-ton bin is not identified as a 'zero-ton'. Then,

$$\begin{aligned} Pr(E_z) &= Pr(||\vec{w}_{b,i,j}||^2 > (1+\gamma)D) \\ &= P(\chi^2_{2D} > 2(1+\gamma)D) \\ &< 2\exp(-D\gamma^2/9) \quad \forall \quad \gamma \in [0, 1/3]. \end{aligned} \tag{17}$$

where the last inequality follows from a standard concentration bound for Lipschitz functions of Gaussian variables, along with the fact that the Euclidean norm is a 1-Lipschitz function. Thus, $Pr(E_z) < O(1/k^2)$ if,

$$D > 18\log(k)/\gamma^2 \tag{18}$$

b) Analysis of a single-ton bin: Let $\vec{y}_{b,i,j} = X[\ell]\vec{a}(\ell) + \vec{w}_{b,i,j}$, be an observation vector of a single-ton bin. The steering vector $\vec{a}(\ell)$ is the ℓ^{th} column of bin-measurement matrix $\boldsymbol{A}_{i,j}$ and $X[\ell]$ is the only non-zero DFT coefficient connected to this bin. Let E_s be an event that a single-ton bin is not decoded correctly. The event E_s consists of the following three events.

Single-ton bin is wrongly classified as a zero-ton bin: [E_{sz}] Let E_{sz} denote an event that the single-ton bin fails the energy test of the Algorithm 2 and is classified as a zero-ton.

$$\begin{aligned} Pr(E_{sz}) &= Pr(||\vec{y}_{b,i,j}||^2 < (1+\gamma)D) \\ &= Pr(||X[\ell]\vec{a}(\ell) + \vec{w}_{b,i,j}||^2 < (1+\gamma)D) \\ &< 2\exp\{-D\gamma^2/9\} + \exp\{-(||X[\ell]\vec{a}(\ell)|| - \sqrt{2\gamma D})_+^2\} \end{aligned}$$

where the last inequality follows from Lemma C.1. The non-zero DFT coefficient $X[\ell] = \sqrt{\rho}e^{\imath\phi}$ and the steering vector $||\vec{a}(\ell)|| = \sqrt{D}$. Hence,

$$Pr(E_{sz}) < 2\exp\{-D\gamma^2/9\} + \exp\{-(\sqrt{\rho D} - \sqrt{2\gamma D})_+^2\}. \tag{19}$$

Single-ton bin is wrongly classified as some other single-ton bin: [E_{ss}] Let E_{ss} denote an event that the Algorithm 2 wrongly concludes that the observation $\vec{y}_{b,i,j}$ corresponds to a single-ton bin with steering vector $\vec{a}(\ell')$ and the non-zero DFT coefficient $X[\ell']$, for some $\ell' \neq \ell$. Then,

$$\begin{aligned} Pr(E_{ss}) &= Pr(||\vec{y}_{b,i,j} - X[\ell']\vec{a}(\ell')||^2 < (1+\gamma)D) \\ &= Pr(||X[\ell]\vec{a}(\ell) - X[\ell']\vec{a}(\ell') + \vec{w}_{b,i,j}||^2 < (1+\gamma)D) \\ &\overset{(a)}{=} Pr(||\boldsymbol{A}_{i,j}\vec{v} + \vec{w}_{b,i,j}||^2 < (1+\gamma)D) \\ &< 2\exp\{-D\gamma^2/9\} + \exp\{-(||\boldsymbol{A}_{i,j}\vec{v}|| - \sqrt{2\gamma D})_+^2\} \end{aligned} \tag{20}$$

where $\vec{v}$ used in (a) is a n-dimensional complex vector with only two non-zero values $v[\ell] = X[\ell]$ and $v[\ell'] = X[\ell']$, i.e., 2-sparse. The last inequality again follows from Lemma C.1.

Using Lemma V.3 and Lemma V.4,

$$\begin{aligned} ||\boldsymbol{A}_{i,j}\vec{v}||^2 &\geq 2||\vec{v}||^2 D(1-\mu_{\max}(\boldsymbol{A}_{i,j}))_+ \\ &= 2\rho D(1-\mu_{\max}(\boldsymbol{A}_{i,j}))_+ \\ &\geq 2\rho D(1-2\sqrt{\log(5n)/D})_+. \end{aligned}$$

Thus, bound in (20) becomes,

$$Pr(E_{ss}) < 2\exp\{-D\gamma^2/9\} + \exp\left\{-\left(\sqrt{2\rho D(1-2\sqrt{\log(5n)/D})_+} - \sqrt{2\gamma D}\right)_+^2\right\} \tag{21}$$

Single-ton bin is wrongly classified as a multi-ton bin: $[E_{sm}]$ Let E_{sm} be an event that bin is processed by the Algorithm 2 but no single-ton is found. Thus,

$$\begin{aligned} Pr(E_{sm}) &< Pr(E_{sm} \mid \hat{X[\ell]} = X[\ell]) + Pr(\hat{X[\ell]} \neq X[\ell]) \\ &= Pr(E_z) + Pr(\hat{X[\ell]} \neq X[\ell]) \end{aligned} \tag{22}$$

From Algorithm 2 we have $\hat{X[\ell]} = \vec{a}(\ell)^\dagger \vec{y}_{b,i,j}/D = X[\ell] + \mathcal{CN}(0, 1/D)$. Then, using the fact that non-zero DFT coefficients take value from a M-PSK constellation with magnitude $\sqrt{\rho}$ we have,

$$\begin{aligned} Pr(\hat{X[\ell]} \neq X[\ell]) &< Pr(|\mathcal{CN}(0,1/D)| > \sqrt{\rho}\sin(\pi/M)) \\ &= \exp\{-D\rho\sin^2(\pi/M)\}. \end{aligned}$$

Substituting the above bound and (17) in (22), we get,

$$Pr(E_{sm}) < 2\exp\{-D\gamma^2/9\} + \exp\{-D\rho\sin^2(\pi/M)\}. \tag{23}$$

Further using a union bound, we get an upper bound on the probability of event E_s as,

$$Pr(E_s) < Pr(E_{sz}) + Pr(E_{ss}) + Pr(E_{sm})$$

Thus, from (19), (21) and (23), to reliably decode a single-ton bin with probability at least $1-O(1/k^2)$ we need

$$D > \max\left\{\frac{2\log(k)}{(\sqrt{\rho}-\sqrt{2\gamma})_+^2}, 16\log(n), \frac{18\log(k)}{\gamma^2}, \frac{2\log(k)}{\rho\sin^2(\pi/M)}\right\} \tag{24}$$

c) Analysis of a multi-ton bin: Consider a multi-ton bin, in particular a L-ton bin where $L \geq 2$. Then, the observation vector of this bin can be written as $\vec{y}_{b,i,j} = \boldsymbol{A}_{i,j}\vec{v} + \vec{w}_{b,i,j}$, where $\vec{v} \in \mathbb{C}^n$ is some

L-sparse vector. Let E_m be an event that a multi-ton bin is decoded as a single-ton bin with a steering vector $\vec{a}(\ell)$ and the non-zero DFT coefficient $X[\ell]$ for some ℓ. We further split the event E_m into two parts: i) $E_{m,1}$, be an event that a multi-ton bin of size $L = O(1)$ is confused as some singleton bin and ii) $E_{m,2}$, be an event that a multi-ton bin of asymptotic size $L \to \infty$ is confused as some singleton bin.

Next, we compute an upper bound on the probability of the failure event E_m as follows:

$$\begin{aligned} Pr(E_m) &< Pr(E_m \mid L < O(1)) + Pr(E_m \mid O(1) < L < \log(k)) + Pr(L \geq \log(k)) \\ &< Pr(E_{m,1}) + Pr(E_{m,2}) + O(1/k^2), \end{aligned} \tag{25}$$

where in last inequality we have used the fact that the number of the non-zero DFT coefficients connected to any bin L is a Binomial $B(1/(\eta k), k)$ random variable (see [1] for more details), for some constant $\eta > 0$. Hence to show that $Pr(E_m) < O(1/k^2)$, we need to show that the probability of each of the two events $Pr(E_{m,1})$ and $Pr(E_{m,2})$ is upper bounded by $O(1/k^2)$.

First we compute an upper bound on event $E_{m,1}$. Let $\vec{u} = \vec{v} - X[\ell]\vec{e}_\ell$, where $\vec{e}_\ell$ is a standard basis vector with 1 at ℓ^{th} location and 0 elsewhere. The vector $\vec{u}$ is either $L+1$ or L sparse. Then, using the Lemmas V.3, V.4 and the fact that all the non-zero components of $\vec{u}$ are of the form $\sqrt{\rho}e^{\imath\phi}$, we have,

$$\begin{aligned} ||\boldsymbol{A}_{i,j}\vec{v} - X[\ell]\vec{a}(\ell)||^2 &= ||\boldsymbol{A}_{i,j}\vec{u}||^2 \\ &\geq L\rho D(1 - \mu_{\max}(\boldsymbol{A}_{i,j})L)_+ \\ &> L\rho D(1 - 2L\sqrt{\log(5n)/D})_+ \end{aligned} \tag{26}$$

Thus,

$$\begin{aligned} Pr(E_{m,1}) &= Pr(||\boldsymbol{A}_{i,j}\vec{v} - X[\ell]\vec{a}(\ell) + \vec{w}_{b,i,j}||^2 < (1+\gamma)D \mid L < O(1)) \\ &\overset{(a)}{<} \max_{2 \leq L < O(1)} \exp\left\{ -\left(\sqrt{L\rho D(1 - 2L\sqrt{\log(5n)/D})_+} - \sqrt{\gamma D} \right)_+^2 \right\}, \end{aligned}$$

where in (a) we used the Lemma C.1 and the lower bound from (26).

Hence, $Pr(E_{m,1}) < O(1/k^2)$, if $D > O(\log n)$.

Next, we analyze the event $E_{m,2}$ that a multi ton with asymptotic number of components L is confused with a singleton. Let Ω be the set of integers representing the support of the non-zero DFT coefficients participating in the multi-ton bin under consideration, i.e., $|\Omega| = L$. Also let $\vec{y}_b$ be a D-dimensional complex bin observation. Then,

$$\vec{y}_b = \boldsymbol{A}_{b,\Omega}\vec{X}_\Omega + \vec{w}_b,$$

where the noise $\vec{w}_b \sim \mathcal{CN}(0, I_{D\times D})$, while $\boldsymbol{A}_{b,\Omega} \in \mathbb{C}^{D\times L}$ and $\vec{X}_\Omega \in \mathbb{C}^L$ denote Ω restricted bin-measurement matrix and DFT vector respectively.

Using Central-Limit-Theorem (CLT), for asymptotic value of L, $\vec{y}_b$ is a zero-mean jointly Gaussian random vector. Let $\vec{a}_p$ and $\vec{a}_q$ be the p^{th} and the q^{th} row of the matrix $\boldsymbol{A}_{b,\Omega}$. Then, over the random choice of delays

$$\begin{aligned}\mathbb{E}[\vec{a}_p^\dagger \vec{a}_q] &= \mathbb{E}\left[\sum_{s=0}^{L-1} e^{\imath(2\pi/n)\ell_s(q-p)}\right] \\ &= 0 \end{aligned} \tag{27}$$

where $\ell_s \in \Omega$ and p, q are uniformly random between 0 and $n-1$. Hence the entries of the observation vector $\vec{y}_b$ are independent zero-mean Gaussian random variables each with variance $L\rho + 1$. Thus, we have,

$$\begin{aligned}Pr(||\vec{y}_b||^2 < (1-\gamma)D(\rho L+1)) &= Pr(\chi^2_{2D} < 2D(1-\gamma)) \\ &< 2e^{-D\gamma^2/4} \\ &< O(1/k^2) \end{aligned} \tag{28}$$

for fixed γ and $D = O(\log n)$. Further, since any singleton component has energy ρD, the residual signal $\vec{y}_b - X[\ell]\vec{a}(\ell)$ has energy at least $\rho D(L(1-\gamma-1))$ which is clearly greater than the threshold $(1+\gamma)D$ used for noise detection. Thus, a multi-ton bin with observation $\vec{y}_b$ that is composed of $L \to \infty$ components cannot be confused with a singleton with probability at least $1 - O(1/k^2)$.

d) Upper bound on the probability of event E_b: We set the threshold $\gamma = \min\{1, \rho/4\}$. Then, using (18), (24) and (25) we have,

$$Pr(E_b) < O(1/k^2), \tag{29}$$

for any fixed $\rho > 0$ and $D = O(\log n)$.

B. Computational complexity of FFAST

The computational cost of the FFAST algorithm can be roughly computed as,

$$\text{Total } \# \text{ of arithmetic operations} = \text{\# of iterations} \times \{\text{\# of bins} \times (\text{operations per bin})\}.$$

As shown in [1] for all values of sparsity index $0 < \delta < 1$ the FFAST front-end employs no more than $O(k)$ number of bins and if successful completes decoding in constant number of iterations. Now, from

the pseudocode of function Singleton-Estimator provided in Algorithm 2, it is clear that per bin FFAST performs exhaustive search over $O(n/k)$ columns of bin-measurement matrix, where each column is of dimension D. Further as shown in Section C-A2, number of delays $D = O(\log n)$ is sufficient for reliable reconstruction. Thus, the overall computational complexity of the FFAST algorithm is no more than,

$$\begin{aligned}\text{Total } \# \text{ of arithmetic operations} &= \text{\# of iterations} \times \{O(k) \times (O(n/k) \times D)\} \\ &= O(n \log(n))\end{aligned}$$

■

APPENDIX D
THRESHOLD BASED ENERGY-DETECTOR

In this section we provide a proof of the Lemma C.1. Let $\vec{u} \in \mathbb{C}^D$ be a complex D dimensional vector embedded in zero mean white Guassian noise $\vec{w} \in \mathcal{CN}(0, I_{D\times D})$. The energy detector fails to identify the presence of the vector $\vec{u}$, if $||\vec{u}+\vec{w}||^2 < (1+\gamma)D$, where $0 < \gamma < 1$ is a constant. Next, we analyze the probability of this event.

$$\begin{aligned}& Pr(||\vec{u}+\vec{w}||^2 < (1+\gamma)D) \\ = \;& Pr(||\sqrt{2}\vec{u}+\sqrt{2}\vec{w}||^2 < 2D(1+\gamma)) \\ \overset{(a)}{=} \;& Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 + \chi^2_{2D-1} < 2D(1+\gamma)) \\ < \;& Pr(\chi^2_{2D-1} < 2D(1-\gamma)) + Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 2D(1+\gamma) - 2D(1-\gamma)) \\ = \;& Pr(\chi^2_{2D-1} < 2D(1-\gamma)) + Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 4D\gamma), \end{aligned} \tag{30}$$

in (a) we did a change of basis such that $\vec{u}/||\vec{u}||$ is one of the basis vectors. Since $\vec{w}$ is circularly symmetric, change of basis does not change its distribution. Using the fact that $\mathcal{N}(0,1)^2 - 1$ is a sub-exponential random variable with parameters $(4,4)$, we obtain a standard tail bound for χ^2_{2D-1} as follows,

$$Pr(\chi^2_{2D-1} < (2D-1)(1-t)) < 2e^{-(2D-1)t^2/8}, \;\forall\; 0 < t < 1$$

Set $t = \frac{2D\gamma-1}{2D-1}$ and using $D > 1.5/\gamma$, we get,

$$Pr(\chi^2_{2D-1} < 2D(1-\gamma)) < 2e^{-D\gamma^2/9}. \tag{31}$$

Now consider the second term in (30) as,

$$\begin{aligned} & Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 4D\gamma) \\ < \; & Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1) < 2\sqrt{D\gamma})) \\ = \; & Pr(\mathcal{N}(0,1) < -(\sqrt{2}||\vec{u}|| - 2\sqrt{D\gamma})) \\ < \; & 2Pr(\mathcal{N}(0,1) < -(\sqrt{2}||\vec{u}|| - 2\sqrt{D\gamma})_+) \\ < \; & \exp\left\{-\left(||\vec{u}|| - \sqrt{2D\gamma}\right)_+^2\right\}. \end{aligned} \tag{32}$$

Substituting (31) and (32) in (30), we get,

$$Pr(||\vec{u}+\vec{w}||^2 < (1+\gamma)D) < 2e^{-D\gamma^2/9} + e^{-\left(||\vec{u}|| - \sqrt{2D\gamma}\right)_+^2}$$

■

REFERENCES

[1] S. Pawar and K. Ramchandran, "Computing a k-sparse n-length discrete fourier transform using at most 4k samples and o (k log k) complexity," *arXiv preprint arXiv:1305.0870*, 2013.

[2] R. Prony, "Essai experimental–,-," *J. de l'Ecole Polytechnique*, 1795.

[3] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical Journal of the Royal Astronomical Society*, vol. 33, no. 3, pp. 347–366, 1973.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, 1986.

[5] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, 1989.

[6] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, 2006.

[7] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[8] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[9] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on IT*, 2007.

[10] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on IT*, 2005.

[11] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, 2012.

[12] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *Trans. on Sig. Proc.*, 2002.

[13] P. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1741–1757, 2007.

[14] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 31–40, 2008.

[15] M. Mishali and Y. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 375–391, 2010.

[16] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse fourier representations via sampling," ser. STOC '02. New York, NY, USA: ACM, 2002.

[17] A. C. Gilbert, S. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal sparse fourier representations," in *in Proc. SPIE Wavelets XI*, 2003.

[18] A. C. Gilbert, M. J. Strauss, and J. A. Tropp, "A tutorial on fast fourier sampling," *Signal Processing Magazine, IEEE*, 2008.

[19] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Nearly optimal sparse fourier transform," in *Proc. of the 44th SOTC*. ACM, 2012.

[20] ——, "Simple and practical algorithm for sparse fourier transform," in *ACM-SIAM Symp. on Discrete Algorithms*, 2012.

[21] M. Iwen, "Combinatorial sublinear-time fourier algorithms," *Foundations of Computational Mathematics*, vol. 10, no. 3, pp. 303–338, 2010.

[22] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[23] S. A. Gershgorin, "Uber die abgrenzung der eigenwerte einer matrix," *Proceedings of the Russian Academy of Sciences. Mathematical series*, pp. 749 – 754, 1931.