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

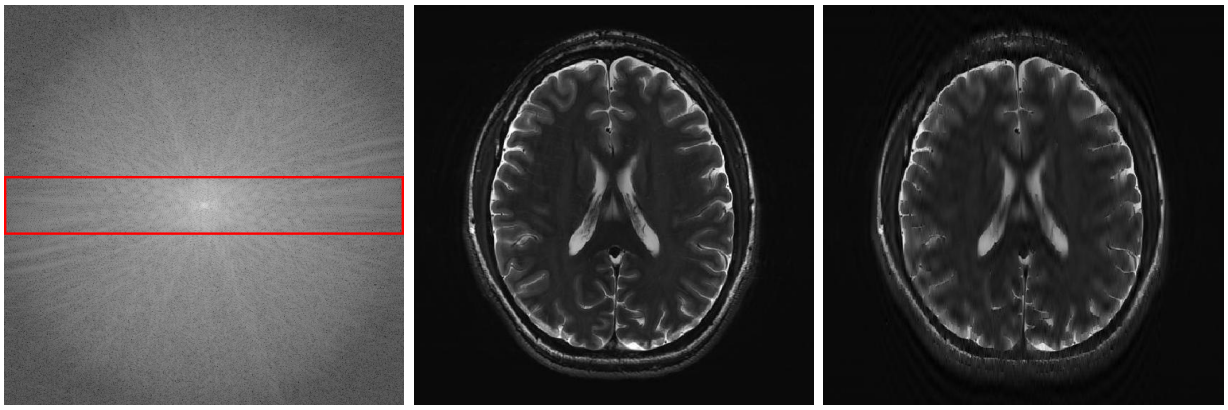
Sameer Pawar and Kannan Ramchandran
Dept. of Electrical Engineering and Computer Sciences
University of California, Berkeley
{spawar, kannanr}@eecs.berkeley.edu

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 $O(n \log n)$. 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 algorithm R-FFAST computes an n-length k-sparse DFT $\vec{X}$ using $O(k \log n)$[1] noise-corrupted time-domain samples, in $O(n \log n)$ computations. Additionally, we also provide a variant of R-FFAST, called an enhanced R-FFAST, that computes an n-length k-sparse DFT $\vec{X}$ using $O(k \log^2 n)$ noise-corrupted time-domain samples, in $O(k \log^3 n)$ computations, i.e., sub-linear time complexity. 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 R-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 using the R-FFAST.

Fig. 1. Application of the $2D$ R-FFAST algorithm to reconstruct the 'Brain' image acquired on an MR scanner with dimension 504×504. The $2D$ R-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. signals relating to audio, image, and video data, biomedical signals, 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 noiseless FFAST framework of [1] to a noise robust R-FFAST algorithm. The robustness against sample noise is achieved by an elegant modification to the FFAST framework. Specifically, the FFAST front-end in [1] has multiple stages of sub-sampling operations, where each stage further has 2 sub-sampling paths called the delay-chains. The delay-chains in each stage of FFAST have an identical sampling period but a circularly time-shifted input signal, i.e., *consecutive shifts*. In contrast, in R-FFAST we use $O(\log n)$ number of delay-chains per stage and the input signal to each delay-chain is circularly shifted by a *random* amount prior to sub-sampling, i.e., *random shifts*, (see Section V for more details). A random choice of the circular shifts endows the effective measurement matrix with a

good *mutual incoherence* and *RIP* [2], thus resulting in stable recovery.

As a motivating example, we demonstrate an application of a $2D$ extension of the R-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 R-FFAST algorithm reconstructs the 'Brain' image acquired on an MR scanner from 60.18% of the Fourier samples as shown in Fig. 1(c). The original MR scanner image is not pre-processed, other than retaining an image of size 504×504. In Section VII-C, we elaborate on the specific details of this experiment. These results are not meant to compete with the state-of-the-art techniques in the MRI literature but rather to demonstrate the feasibility of our proposed approach as a promising direction for future research. More significantly, this experiment demonstrates that while the theory applies to a uniformly-sparse model for the significant DFT coefficients, in practice, our algorithm works even for the non-uniform (or clustered) support setting as is typical of MRI images. 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 R-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. Additionally, we also provide an enhancement to the R-FFAST algorithm that computes an n-length k-sparse DFT $\vec{X}$ using $O(k \log^2 n)$ noise-corrupted time-domain samples, in $O(k \log^3 n)$ computations, i.e., *sub-linear time complexity*. 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.

The rest of the paper is organized as follows: In Section II, we provide the problem formulation and the signal model. Section III provides the main results of this paper. An overview of the related literature is provided in Section IV. In Section V, we explain the R-FFAST architecture and the algorithm using a simple example. An enhancement R-FFAST architecture and the algorithm that has sub-linear time complexity is discussed in Section VI. Section VII provides experimental results that validate the performance of the R-FFAST algorithm.

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:

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

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

$$\vec{y} = \vec{x} + \vec{z}, \tag{2}$$

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

Further, we make the following assumptions on the signal:

- The number of non-zero DFT coefficients $k = O(n^\delta)$, where $0 \leq \delta < 1$, i.e., sub-linear sparsity.
- Let $\mathcal{M} = \{\sqrt{\rho}/2 + i\sqrt{\rho}/M_1\}_{i=0}^{M_1}$ and $\Theta = \{2\pi i/M_2\}_{i=0}^{M_2-1}$, where ρ is a desired signal-to-noise ratio and M_1, M_2 are some constants.
- We assume that all the non-zero DFT coefficients belong to the set $\mathcal{A} = \{\alpha e^{\imath\phi} \mid \alpha \in \mathcal{M}, \phi \in \Theta\}$ (see Remark III.2 for a further discussion on this constraint).
- 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, $\rho = \mathbb{E}_{X[\ell]\neq 0}\{|X[\ell]|^2\}/\mathbb{E}\{||\vec{z}||^2/n\}$.

III. Main results

We provide two variants of the R-FFAST algorithm, with different sample and computational complexity. Both the variants reliably compute a k-sparse n-length DFT $\vec{X} \in \mathcal{A}^n$, of a signal from its noise-corrupted time-domain samples, $\vec{y} = \vec{x} + \vec{z}$. A precise statements of the main results is given by the following theorems.

Theorem III.1. *For any given $0 < \varepsilon < 1$ and a finite signal-to-noise-ratio, there exist (infinitely many) sufficiently large n, such that the R-FFAST algorithm computes a k-sparse DFT $\vec{X} \in \mathcal{A}^n$, where $k = O(n^\delta)$ 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 $1-\varepsilon$.*

  Proof: Please see Appendix C. ■

Remark III.2. *The R-FFAST algorithm reconstructs the k-sparse DFT $\vec{X}$ perfectly, even from the noise-corrupted time-domain samples, with high probability, since in our signal model we assume that $\vec{X} \in \mathcal{A}^n$. The reconstruction algorithm is equally applicable to an arbitrary complex-valued non-zero DFT coefficients, as long as all the non-zero DFT coefficients respect the signal-to-noise-ratio as defined in Section II. The proof technique for arbitrary complex-valued non-zero DFT coefficients becomes much more cumbersome, and we do not pursue it in this paper. However, in Section VII we provide an empirical evidence of the applicability of the R-FFAST algorithm to signals with arbitrary complex-valued DFT coefficients. The reconstruction is deemed successful if the support is recovered perfectly. We also note that the normalized ℓ_1-error is small (see Section VII for more details) when the support is recovered successfully.*

Theorem III.3. *For a finite but sufficiently high ρ and any given $0 < \varepsilon < 1$, there exist (infinitely many) sufficiently large n, such that the enhanced R-FFAST algorithm computes a k-sparse DFT $\vec{X} \in \mathcal{A}^n$, where $k = O(n^\delta)$ 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 $m = O(k \log^2 n)$ samples of $\vec{y}$.*
- **Computational complexity**: *The computational complexity of the algorithm is $O(k \log^3 n)$.*
- **Probability of success**: *The algorithm successfully recovers all the non-zero DFT coefficients of the signal $\vec{x}$, with probability at least $1-\varepsilon$.*

  Proof: Please see Appendix G. ■

IV. Related work

The problem of computing a sparse discrete Fourier transform of a signal is related to the rich literature of frequency estimation [3, 4, 5, 6] in statistical signal processing as well as compressive-sensing [7, 8]. Most works in the frequency estimation literature use well studied subspace decomposition principles e.g., singular-value-decomposition, like MUSIC and ESPRIT [3, 5, 6]. These methods 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 R-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 [8, 9, 10]. A standard tool used for the analysis of the reconstruction algorithms is the *restricted isometry property* (RIP) [2]. 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] have proposed a sub-linear time and sample complexity algorithms for computing a sparse DFT of a high-dimensional signal. Although sub-linear in theory, these algorithms require large number of samples in practice and are robust against limited noise models, e.g., bounded noise. In [20], the authors propose an algorithm for computing a k-sparse $2D$-DFT of an $\sqrt{n} \times \sqrt{n}$ signal, where $k = \Theta(\sqrt{n})$. For this special case, the algorithm in [20] computes $2D$-DFT using $O(k \log n)$ samples and $O(k \log^2 n)$ computations. In contrast, the R-FFAST algorithm is applicable to $1D$ signals and for the entire sub-linear regime of sparsity, i.e., $k = O(n^\delta), 0 < \delta < 1$, when the time-domain samples are corrupted by Gaussian noise.

V. R-FFAST ARCHITECTURE AND ALGORITHM

In this section, we describe the R-FFAST sub-sampling "front-end" architecture as well as the associated "back-end" peeling-decoder. First, we use a simple example to illustrate the key ideas and later we provide a generic description of the architecture as well as the algorithm.

Example V.1. *Consider an $n = 20$ length input signal $\vec{x}$, whose DFT $\vec{X}$, is $k = 5$ sparse. Further,*

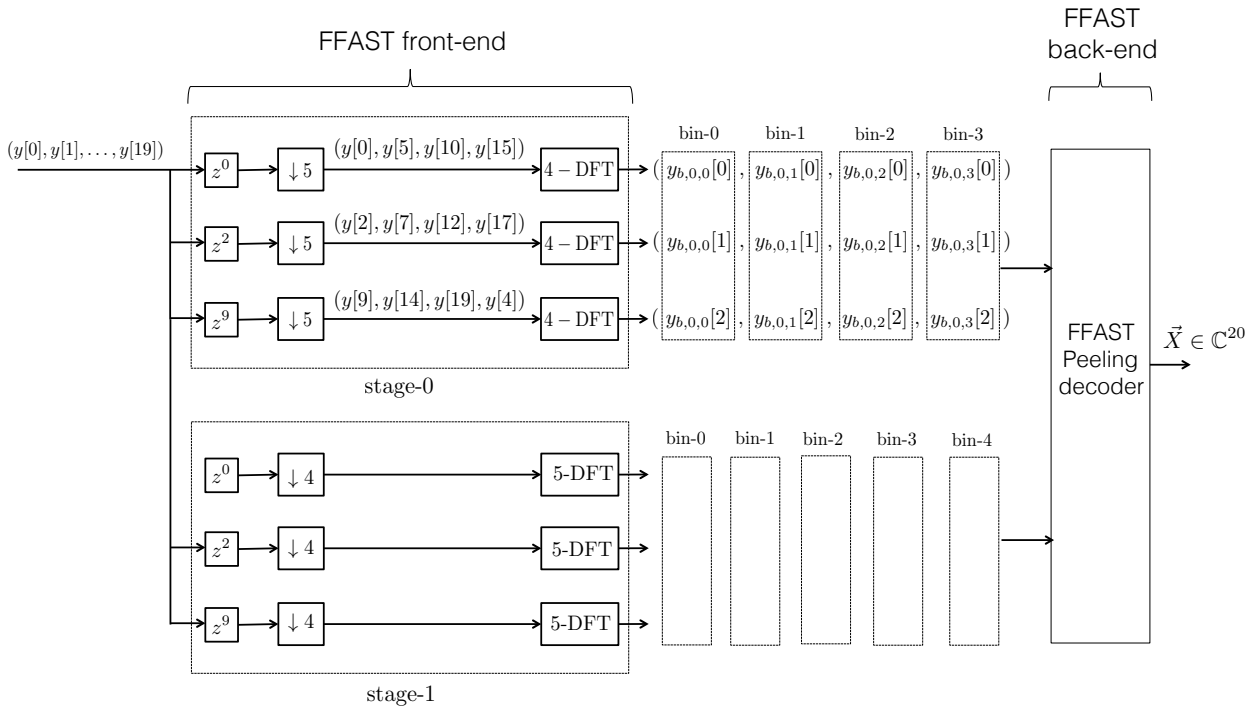


Fig. 2. An example 2-stage R-FFAST architecture. The noise-corrupted time-domain samples $\vec{y} = \vec{x} + \vec{z}$, of a 20-point signal $\vec{x}$ with a 5 sparse DFT $\vec{X}$, are processed using the R-FFAST front-end and the peeling back-end algorithm.

let the 5 *non-zero DFT coefficients of* $\vec{x}$ *be* $X[1]$, $X[3]$, $X[5]$, $X[10]$ *and* $X[15]$. *Let* $\vec{y} = \vec{x} + \vec{z}$ *be the noise-corrupted time-domain signal input to a* 2 *stage R-FFAST architecture, shown in Fig. 2. In its most generic form, the R-FFAST sub-sampling 'front-end' architecture consists of* d (≥ 3) *stages, where* d *is a non-decreasing function of the sparsity-index* δ $(k = O(n^{\delta}))$. *Each stage further has* D *number of subsampling paths or delay-chains with an identical sampling period, e.g.,* 5 *and* 4 *in Fig. 2. The input signal to the different delay-chains is circularly shifted by a random amount. For illustrative purposes, in Fig. 2, we show the processing of* $\vec{y}$ *through a* $d = 2$ *stage R-FFAST architecture, where the sampling period of the two stages are* 5 *and* 4 *respectively. Each stage further has* $D = 3$ *delay-chains. The input signal to each of the* $D = 3$ *delay-chains is circularly shifted by* $0, 2$ *and* 9 *respectively prior to the sub-sampling operation. The output of the R-FFAST front-end is obtained by computing the short DFTs of the sub-sampled data and further grouping them into 'bins', as shown in Fig. 2. The R-FFAST peeling-decoder then synthesizes the big DFT* $\vec{X}$ *from the output of the short DFTs that form the bin observations.*

The relation between the output of the short DFTs, i.e., bin-observation vectors, and the DFT coefficients of the input signal $\vec{x}$, *can be computed using preliminary signal processing identities. Let* $\vec{y}_{b,i,j}$

denote a 3*-dimensional observation vector of bin* j *of stage* i*. Then,*

$$\vec{y}_{b,0,0} = \begin{pmatrix} w_{b,0,0}[0] \\ w_{b,0,0}[1] \\ w_{b,0,0}[2] \end{pmatrix},$$

$$\vec{y}_{b,0,1} = \begin{pmatrix} 1 \\ e^{\imath 2\pi 2/20} \\ e^{\imath 2\pi 9/20} \end{pmatrix} X[1] + \begin{pmatrix} 1 \\ e^{\imath 2\pi 10/20} \\ e^{\imath 2\pi 45/20} \end{pmatrix} X[5] + \begin{pmatrix} w_{b,0,1}[0] \\ w_{b,0,1}[1] \\ w_{b,0,1}[2] \end{pmatrix} \tag{3}$$

$$\vec{y}_{b,0,2} = \begin{pmatrix} 1 \\ e^{\imath 2\pi 20/20} \\ e^{\imath 2\pi 90/20} \end{pmatrix} X[10] + \begin{pmatrix} w_{b,0,2}[0] \\ w_{b,0,2}[1] \\ w_{b,0,2}[2] \end{pmatrix}, \tag{4}$$

where $\vec{w}_{b,i,j} \in \mathcal{CN}(0, I_{3\times 3})$ *consists of the DFT coefficients of the samples of the noise vector* $\vec{z}$*. A bin whose observation vector has no contribution from any of the non-zero DFT coefficients of the signal, e.g., bin* 0 *of stage* 0*, is called a 'zero-ton' bin. Similarly a bin that has contribution from exactly one non-zero DFT coefficient of the signal is called a 'single-ton' bin, and one that has a contribution from more than one non-zero DFT coefficients as a 'multi-ton' bin, e.g., bin* 2 *of stage* 0 *is a singleton while the bin* 1 *of stage* 0 *is a multi-ton bin.*

More generally, the observation vector $\vec{y}_{b,i,j}$*, of bin* j *of stage* i *of a R-FFAST front-end architecture with* d *stages and* D *delay-chains per stage, is given by,*

$$\vec{y}_{b,i,j} = \boldsymbol{A}_{i,j}\vec{X} + \vec{w}_{b,i,j}, \quad 0 \le i < d, \ 0 \le j < f_i, \tag{5}$$

where f_i *is the number of samples per delay-chain in stage* i*,* $\vec{w}_{b,i,j} \sim \mathcal{CN}(0, I_{D\times D})$ *and* $\boldsymbol{A}_{i,j}$ *is the* bin-measurement *matrix. The bin-measurement matrix* $\boldsymbol{A}_{i,j}$ *is a function of the sampling period, number of delay-chains as well as the circular shifts used in stage* i *of the R-FFAST sub-sampling front-end. For example, consider a R-FFAST architecture such that stage* i *has* D *number of delay-chains, sampling period* n/f_i*, i.e.,* f_i *number of samples per delay-chain, and circularly shifts the input signal by* r_s *amount in the* s^{th} *delay-chain before subsampling for* $s = 0, 1, \ldots, D-1$*. Then, the bin-measurement*

matrix $\boldsymbol{A}_{i,j} \in \mathbb{C}^{D\times n}$ *of bin* j *of stage* i *is given by:*

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

where $\vec{\boldsymbol{A}}_{i,j}(\ell)$ *is the* ℓ^{th} *column,* $\ell = 0, \ldots, n-1$*, of* $\boldsymbol{A}_{i,j}$ *and*

$$\vec{a}(\ell)^\dagger = \left(\; e^{-\frac{\imath 2\pi \ell r_0}{n}}, e^{-\frac{\imath 2\pi \ell r_1}{n}}, \ldots, e^{-\frac{\imath 2\pi \ell r_{D-1}}{n}} \; \right). \tag{7}$$

For example in equation (3) *the bin-measurement matrix* $\boldsymbol{A}_{0,1}$ *is,*

$$\boldsymbol{A}_{0,1} = \left[\begin{array}{cccccccccccccccccccc} \vec{0} & \vec{a}(1) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(5) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(9) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(13) & \vec{0} & \vec{0} & \vec{0} & \vec{a}(17) & \vec{0} & \vec{0} \end{array} \right],$$

where $\vec{a}(\ell) = \begin{pmatrix} 1 \\ e^{\imath 2\pi 2\ell/20} \\ e^{\imath 2\pi 9\ell/20} \end{pmatrix}$ *for* $\ell = 0, 1, \ldots, 19$*. In the sequel, we refer to* $\vec{a}(\ell),\ \ell = 0, 1 \ldots, n-1$*, as a* steering-vector *corresponding to the frequency* $2\pi\ell/n$*, sampled at* $(r_0, \ldots, r_{D-1})$*.*

The R-FFAST algorithm effectively computes a 5*-sparse* 20*-point DFT* $\vec{X}$ *of the signal* $\vec{x}$ *as follows,*

- Divide the problem into simpler sub-problems*: The R-FFAST sub-sampling front-end takes a* 20*-length noise-corrupted observation* $\vec{y}$*, and disperses it into* 4 *bin-level problems in stage* 0 *and* 5 *bin-level problems in stage* 1 *respectively, as shown in Fig.* 2*. Each bin-level problem 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 not so trivial, e.g.,* $\vec{y}_{b,0,1}$*.*
- Backend Iterative peeling-decoder*: The R-FFAST backend, in particular "singleton-estimator" (see Algorithm* 2*) identifies which instances of the resulting bin-level sub-problems are* 1*-sparse, i.e., single-ton bins, and reliably computes the support and the value of the non-zero DFT coefficient 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. If one assumes that the singleton-estimator algorithm* 2 *is error free, then, it is easy to verify that the peeling-decoder uncovers the non-zero DFT coefficients of* $\vec{x}$ *in the following order* $X[10], X[1], X[3], X[5]$ *and* $X[15]$*.*

At a high level, the R-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.*

Next, we describe the R-FFAST front-end sub-sampling architecture and the involved design parameters. A pseudo-code of the complete R-FFAST back-end algorithm is provided in Algorithm 1 and Algorithm 2.

Algorithm 1 R-FFAST Algorithm

Input: The noise-corrupted bin observations $\vec{y}_{b,i,j}$, obtained using the R-FFAST sub-sampling 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.

R-FFAST Decoding: Set the initial estimate of the n-point DFT $\vec{X} = 0$. Let ℓ denote the number of iterations performed by the R-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 $q \equiv p \bmod f_s$.
          Set, $X[p] = v_p$.
        else
          bin is a *multi-ton*.
        end if
      end if
    end for
  end for
end for

A. R-FFAST front-end sub-sampling architecture

In its most generic form the R-FFAST front-end sub-sampling architecture consists of d stages and D delay-chains per stage as shown in Fig. 3. The i^{th} delay-chain 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 output

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 and the position p of the non-zero DFT coefficient in case of a single-ton bin.

Singleton-Estimator:
Set the singleton = 'false'.
Set the energy threshold $T = (1 + \gamma)D$ for appropriately chosen γ (see Appendix C).
for each position $q \equiv j \mod f_i$ 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

samples of each delay-chain of stage j. Using identical circular shifts r_i in the i^{th} delay-chain of all the stages is not a fundamental requirement of our result, but a convenience that is sufficient and makes the R-FFAST framework implementation friendly. Next, the (short) DFTs of the output samples of each delay-chain 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 R-FFAST decoder.

1) Number of stages and sampling periods: The number of stages d and the subsampling factor n/f_i are chosen based on the sparsity-index δ, where $k = O(n^\delta), 0 \leq \delta < 1$. Based on the qualitative nature of the design parameters, of the R-FFAST front-end architecture, we identify two regimes of sub-linear sparsity; 1) *very-sparse* regime, where $0 \leq \delta \leq 1/3$, and 2) *less-sparse* regime, where $1/3 < \delta < 1$. The choice of number of stages and the subsampling factors of each stage of R-FFAST front-end, as a function of δ, is identical to that of the FFAST architecture used in the noiseless case. Here, we provide a brief qualitative description about the choice of front-end parameters and we refer the reader to [1] for more details.

- Very-sparse regime: For the very-sparse regime we always use $d = 3$ stages in the R-FFAST front-end. The sampling periods n/f_i, for stages $i = 0, 1, 2$, are chosen such that the integers $f_i = O(k) + O(1)$, i.e., are approximately equal to k, and are relatively co-prime factors of n. For example, $n = 100 * 101 * 103 * 99$, $k = 100$, $f_0 = 100$, $f_1 = 101$ and $f_2 = 103$, achieves an operating point of $\delta = 1/4$.
- Less-sparse regime: For the less-sparse regime the number of stages d is a monotonic non-decreasing

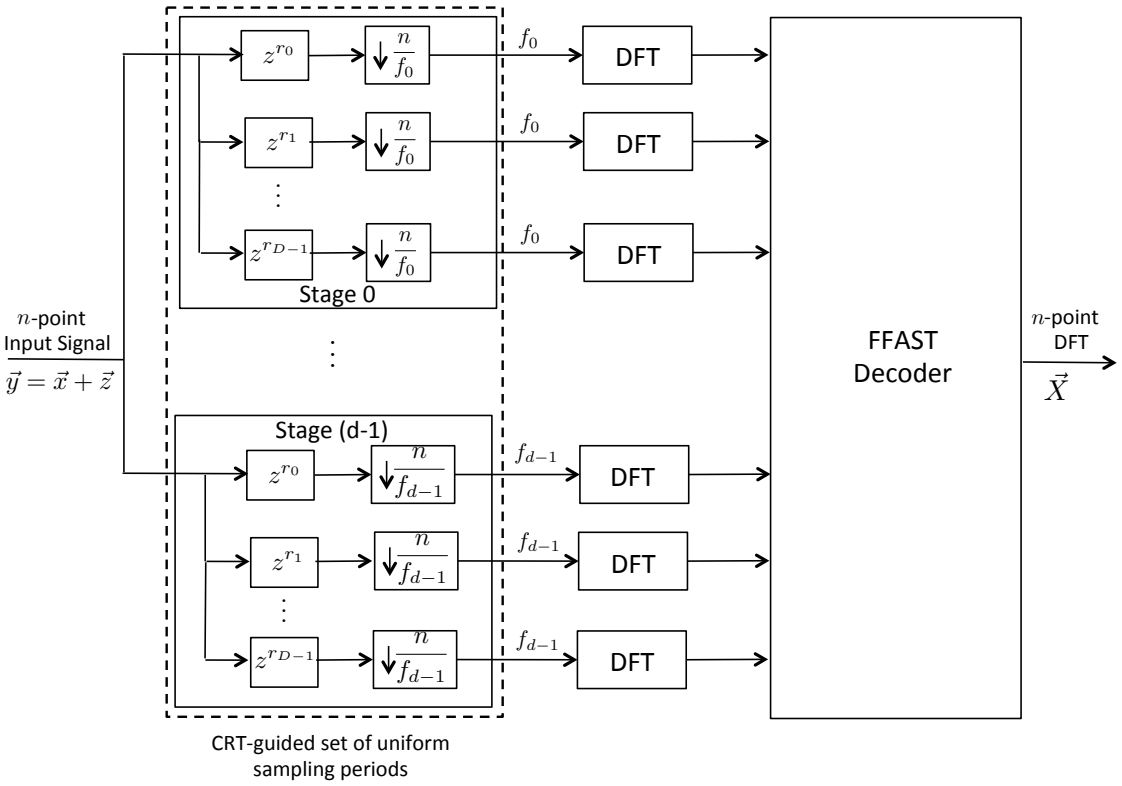


Fig. 3. Schematic block diagram of the R-FFAST architecture for processing noise-corrupted samples $\vec{y} = \vec{x} + \vec{z}$. The n-point input signal $\vec{y}$ is uniformly subsampled by a carefully chosen, guided by the Chinese-Remainder-Theorem, set of d patterns. Each stage has D delay-chains and a delay-chain in stage i has f_i number of samples. The delay shift pattern $(r_0, r_1, \ldots, r_{D-1})$ is carefully chosen so as to induce bin-measurement matrices with "good" mutual incoherence and RIP properties (as explained in Section **??**). Next, the (short) DFTs, of each of the delay-chain samples 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 R-FFAST decoder.

function of the sparsity index δ. In particular, an operating point of fractional sparsity index δ can be achieved as follows. Let the number of stages $d = 1/(1-\delta)$. Consider a set of d relative co-prime integers $\mathcal{P}_i = O(k^{1/(d-1)})$ such that $n = \prod_{i=0}^{d-1} \mathcal{P}_i = O(k^{1/\delta})$. Then, the subsampling periods n/f_i are chosen such that $f_i = \prod_{j=i}^{i+d-2} \mathcal{P}_{(j)_d}$, where $(j)_d = j \mod d$. For example, $n = 10 * 11 * 13$, $k = 100$ and $f_0 = 10 * 11, f_1 = 11 * 13, f_2 = 13 * 10$, achieves $\delta = 2/3$.

2) Delay chains: The number of the delay-chains D as well as the pattern of the circular shifts $\{r_i\}_{i=0}^{D-1}$ used in all the d stages of the R-FFAST front-end architecture influence the noise robustness of the back-end peeling decoder. In (6), we have shown the relation between the choice of the D and $\{r_i\}_{i=0}^{D-1}$ and the bin-measurement matrices, using the signal processing identities. In the compressed sensing literature the *mutual-incoherence* and the *Restricted-isometry-property* (RIP) [2], of the measurement matrix, are widely studied and well understood to play an important role in stable recovery of a high-dimensional

sparse vector from linear measurements in the presence of observation noise. The problem in (5) is a special case of the compressed sensing problem. In particular, in the case of a single-ton bin, the bin-processing algorithm attempts to recover a 1-sparse high-dimensional signal from linear measurements in the presence of the observation noise. As a result, it is crucial to understand the relationship between the parameters D, $\{r_i\}_{i=0}^{D-1}$ and the *mutual-incoherence* and the RIP of the bin-measurement matrices.

Next, we define and establish the mutual-incoherence and the RIP of the bin-measurement matrices for an appropriate choice of D and $\{r_i\}_{i=0}^{D-1}$. Later, we use these properties to prove a stable recovery of the proposed R-FFAST peeling-style iterative recovery algorithm.

Definition V.2. *The mutual incoherence $\mu_{\max}(\boldsymbol{A})$ of a measurement matrix $\boldsymbol{A}$ is defined as*

$$\mu_{\max}(\boldsymbol{A}) \triangleq \max_{\forall p \neq q} \frac{|\vec{\boldsymbol{A}}(p)^\dagger \vec{\boldsymbol{A}}(q)|}{||\vec{\boldsymbol{A}}(p)|| \cdot ||\vec{\boldsymbol{A}}(q)||}, \tag{8}$$

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.3. *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*

$$(1 - \gamma_s(\boldsymbol{A}))||\vec{X}||^2 \leq ||\boldsymbol{A}\vec{X}||^2 \leq (1 + \gamma_s(\boldsymbol{A}))||\vec{X}||^2, \tag{9}$$

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 [21]. 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.

Lemma V.4. *The mutual incoherence $\mu_{\max}(\boldsymbol{A}_{i,j})$ of a bin-measurement matrix $\boldsymbol{A}_{i,j}$, of the R-FFAST front-end with D uniformly random circular shifts, is upper bounded by*

$$\mu_{\max}(\boldsymbol{A}_{i,j}) < 2\sqrt{\log(5n)/D} \ , \ \forall \ i, j \tag{10}$$

with probability at least 0.2.

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 (10) 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 (10) or not. Hence, using offline processing one can choose a pattern $(r_0, \ldots, r_{D-1})$, such that deterministically the bound in (10) is satisfied.

Lemma V.5. *The bin-measurement matrix $\boldsymbol{A}_{i,j}$, of the R-FFAST front-end with D uniformly random circular shifts, satisfies the following RIP condition for all $\vec{X}$ that have $||\vec{X}||_0 \leq s$,*

$$D\left(1-\mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)^{+}||\vec{X}||^2 \leq ||\boldsymbol{A}_{i,j}\vec{X}||^2 \leq D\left(1+\mu_{\max}(\boldsymbol{A}_{i,j})(s-1)\right)||\vec{X}||^2 \ , \ \forall\ i,j \qquad (11)$$

with probability at least 0.2.

Proof: Please see Appendix B. ■

VI. Enhanced R-FFAST architecture and algorithm

In this section, we propose a modification to the R-FFAST front-end sub-sampling pattern to enhance the computational complexity of the R-FFAST algorithm from super-linear (in ambient dimension n) to *sub-linear*. The gains in the computational complexity are achieved at the expense of slightly increased sample complexity. The super-linear complexity of the R-FFAST is due to the exhaustive search performed in the bin-processing algorithm 2. In this section, we propose an enhanced bin-processing algorithm called FSURE, which stands for Fast-Successive-REfinement. The FSURE algorithm is inspired by the frequency estimation literature in signal processing, especially the work of [22]. Next, we briefly review the techniques and results of [22].

A. Frequency estimation of a single complex sinusoid

Consider N samples of a single complex sinusoid buried in white Gaussian noise, obtained at a periodic interval of period T,

$$y(t) = Ae^{j(\omega t+\phi)} + w(t), \quad t = 0, T, 2T, \ldots, (N-1)T, \qquad (12)$$

where the amplitude A, frequency ω, and the phase ϕ are deterministic but unknown constants. The sampling period T determines the maximum frequency $\omega_{\max} \in (0, 2\pi/T)$, that can be estimated using

these samples. Henceforth, for ease of notation WLOG we assume that $T = 1$, i.e., $\omega_{\max} \in (0, 2\pi)$. The sample noise $w(t)$ is assumed to be a zero-mean white complex Gaussian noise with variance σ_w^2.

The optimal maximum likelihood estimator (MLE) is well known to be given by the location of the peak of a periodogram. This estimator attains the Cramer-Rao lower bound on variance for a high-enough signal-to- noise ratio (SNR). In many instances, however, the computation is prohibitive even with an FFT implementation, and so simpler methods are desirable.

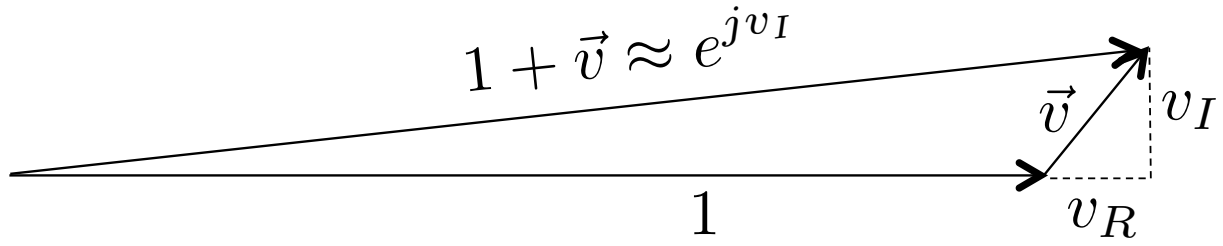


Fig. 4. High SNR approximation

1) High SNR approximation of the phase noise: Let the signal-to-noise-ratio SNR ρ be defined as $\rho = A^2/\sigma_w^2$. Consider the model of (12),

$$\begin{aligned} y(t) &= Ae^{j(\omega t+\phi)} + w(t) \\ &= (1 + v(t))Ae^{j(\omega t+\phi)} \end{aligned} \tag{13}$$

where $v(t) = w(t)e^{-j(\omega t+\phi)}/A$ is a zero-mean complex normal random variable with variance $1/\rho$. For high SNR, i.e., $\rho >> 1$, we can approximate the complex number $(1 + v(t))$ as (also see Fig. 4),

$$\begin{aligned} (1 + v(t)) &= \sqrt{(1 + v_R(t))^2 + v_I^2(t)}e^{j\tan^{-1}(v_I(t)/(v_R(t)+1))} \\ &\approx 1e^{j\tan^{-1}(v_I(t)/(v_R(t)+1))} \\ &\approx 1e^{j\tan^{-1} v_I(t)} \\ &\approx 1e^{jv_I(t)}, \end{aligned} \tag{14}$$

where the phase $v_I(t)$ is a zero mean Guassian random variable with variance $1/2\rho$. The high SNR approximation model was first proposed in [23]. In [22], the author empirically validated that the high SNR approximation holds for SNR's of order $5-7$dB. Thus, under high SNR assumption the observations

in (12) can be written as,

$$y(t) \approx Ae^{j(\omega t+\phi+u(t))}, \tag{15}$$

where $u(t)$ is zero-mean white Gaussian with variance $1/(2\rho)$.

2) Estimator: Using the high SNR approximation model of (15), we have,

$$\begin{aligned} \Delta(t) &= \angle y(t+1) - \angle y(t), \quad t = 0, 1, \ldots, N-2 \\ &= \omega + u(t+1) - u(t) \\ i.e., \ \vec{\Delta} &= \vec{1}\,\omega + \vec{z}, \end{aligned} \tag{16}$$

where $\vec{z} \in \mathbb{R}^{N-1}$ is a zero-mean colored Gaussian noise. The author in [22] then computed the following MMSE rule to estimate the unknown frequency ω,

$$\hat{\omega} = \sum_{t=0}^{N-2} \beta(t) \angle y(t+1) y^{\dagger}(t), \tag{17}$$

where the weights $\beta(t)$ have closed form expression (see [22]) and are computed using the optimal whitening filter coefficients of an MMSE estimator. In [22], the author shows that the frequency estimate of (17) is an unbiased estimator and is relate to the true frequency ω as,

$$\hat{\omega} = \left(\omega + \mathcal{N}\left(0, \frac{6/\rho}{N(N^2-1)}\right)\right)_{2\pi}, \tag{18}$$

where $(\cdot)_{2\pi}$ is a modulo 2π. Equation (18) essentially holds for any value of T, i.e., periodic sampling.

In [22], the author considered a problem where the unknown frequency ω is a real number in an appropriate range, e.g., $\omega \in (0, 2\pi)$. Hence, the result of (18) is expressed in terms of an estimation error as a function of the number of measurements and the signal-to-noise ratio ρ. In contrast, the unknown frequency in the singleton estimation problem of (26) is of the form $2\pi\ell/n$, for some unknown integer $\ell \in [0, n-1]$. Hence, we use the Q-function along with the estimate in (18) to get the following proposition.

Proposition VI.1. *For a sufficiently high SNR and large n, the frequency estimate $\hat{\omega}$ obtained using (17) and $N = O(\log n)$ measurements, satisfies,*

$$Pr(|\hat{\omega} - \omega| > \pi/\log n) < 1/n^3.$$

Proof: Please see Appendix E ■

B. Bin-measurement matrix for FSURE

In Section V, we have shown that the columns of a bin-measurement matrix $\boldsymbol{A}_{i,j} \in \mathbb{C}^{D\times n}$ are determined by circular shifts used in each of the delay-chains in the front-end architecture. In Lemma V.4 and Lemma V.5, we also established the mutual incoherence and RIP properties of a bin-measurement matrix, when the circular shifts are chosen uniformly at random. We build on this structure thus maintaining the incoherence properties. The modified steering vector (or bin-measurement matrix) is shown in (19). It consists of C clusters of measurements, where each cluster further consists of N periodic measurements. Thus, the total number of measurements per bin are $D = NC$. Note, that the first sample of the i^{th} cluster is r_i and the difference between samples for the i^{th} cluster is 2^i, for $i = 0, \ldots, C-1$. First sample locations $r_i's$ in each cluster are chosen uniformly at random between 0 and $n-1$.

$$\vec{a}(\ell) \quad = \quad \begin{pmatrix} e^{\imath 2\pi\ell(r_0+0)/n} \\ e^{\imath 2\pi\ell(r_0+1)/n} \\ \vdots \\ e^{\imath 2\pi\ell(r_0+(N-1))/n} \\ e^{\imath 2\pi\ell(r_1+0)/n} \\ e^{\imath 2\pi\ell(r_1+2)/n} \\ \vdots \\ e^{\imath 2\pi\ell(r_1+2(N-1))/n} \\ \vdots \\ e^{\imath 2\pi\ell(r_{C-1}+2^{C-1}(N-1))/n} \end{pmatrix}. \tag{19}$$

C. Fast-Successive-REfinement (FSURE) bin-processing algorithm

In this section, we propose a fast successive refinement bin-processing algorithm called "FSURE". The FSURE algorithm consists of a two step procedure. First, the frequency estimation algorithm in [22] (i.e., estimator in (17)) is used to process the samples in each cluster to obtain C estimates ω_i for $i = 0, \ldots, C-1$. Then, these estimates are fused together to obtain the final frequency estimate $\hat{\omega}$ corresponding to the location of the non-zero DFT coefficient of a singleton bin.

1) Processing cluster observations: Consider processing the observations of cluster i.

$$\begin{aligned}\Delta_i(t) &= \angle y(t+1) - \angle y(t), \quad t = Ni, Ni+1, \ldots, N(i+1)-2 \\ &= 2^i\omega + u(t+1) - u(t) \\ \vec{\Delta}_i &= \vec{1}\, 2^i\omega + \vec{z}, \end{aligned} \tag{20}$$

where we have used the high SNR approximation model to get $\vec{z} \in \mathbb{R}^{N-1}$ as a zero-mean colored Gaussian noise. Further, applying the MMSE rule of (17) we get an estimate of $2^i\omega$ as follows,

$$\omega_i = \left(2^i\omega + \mathcal{N}\left(0, \frac{6/\rho}{N(N^2-1)}\right)\right)_{2\pi} \tag{21}$$

Thus, each cluster of observations provides an estimate (up to modulo 2π) of a multiple of the true frequency ω. Next, we show how these different estimates can be used to successively refine the search space of the true frequency ω.

2) Successive refinement: Let $\Omega_0 = (\omega_0 - \pi/\log n, \omega_0 + \pi/\log n)$ be a range of frequencies around the estimate ω_0 obtained by processing observations of cluster-0. For a continuous set, with a slight abuse of notation we use $|\cdot|$ to denote the length of the interval, e.g., $|\Omega_0| = 2\pi/\log n$. Using proposition VI.1, we know that $Pr(\omega \notin \Omega_0) < 1/n^3$. Similarly, $Pr(2\omega \notin \Omega_1) < 1/n^3$, where $\Omega_1 = (\omega_1 - \pi/\log n, \omega_1 + \pi/\log n)$.

The operation of multiplying by 2, maps two frequencies θ and $\theta + \pi$ to the same frequency $(2\theta)_{2\pi}$. However, since $|\Omega_0| = 2\pi/\log n$ and n is sufficiently large, $|\Omega_0 \cap \Omega_1/2| \leq 2\pi/(2\log n)$. Similarly, the intersection of estimates from C consecutive clusters is,

$$|\cap_{i=0}^{C-1} \Omega_i/2^i| \leq \frac{2\pi}{2^{C-1}\log n}. \tag{22}$$

Further, for a singleton bin, we know that the frequency corresponding to the location of the non-zero DFT coefficient is of the form $2\pi\ell/n$ for some integer $\ell \in \{0, \ldots, n-1\}$. The FSURE algorithm chooses the number of clusters C and the number of samples per cluster N such that after processing all the clusters the unknown frequency $2\pi\ell/n$ has only constant number of possibilities for ℓ with a high probability. It then exhaustively tries all the candidate values of ℓ to find the final estimate of the location of the non-zero DFT coefficient in a singleton bin. The pseudo code of the FSURE algorithm is provided in 3.

Algorithm 3 FSURE

Inputs:

- A noise-corrupted bin observations $\vec{y}_b \in \mathbb{C}^{NC}$, consisting of C clusters each with N number of measurements.
- Constant c_1 that specifies the number of candidate singletons to be tried exhaustively around the final estimate obtained after processing observations of the C clusters.
- MMSE filter coefficients $\{\beta(0), \ldots, \beta(N-2)\}$.

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

Set $\omega_{prev} = 0$.
for each cluster $i = 0, \ldots, C-1$ **do**
  $\omega_{new} = \frac{1}{2^i} \sum_{t=Ni}^{N(i+1)-2} \beta(t - Ni) \angle y_b(t+1) y_b^\dagger(t)$.
  $\delta_1 = \lceil \frac{\omega_{prev}}{2\pi/2^i} \rceil (2\pi/2^i) + \omega_{new} - \omega_{prev}$.
  $\delta_2 = \lfloor \frac{\omega_{prev}}{2\pi/2^i} \rfloor (2\pi/2^i) + \omega_{new} - \omega_{prev}$.
  if $|\delta_1| < |\delta_2|$ **then**
    $\omega_{prev} = \delta_1 + \omega_{prev}$.
  else
    $\omega_{prev} = \delta_2 + \omega_{prev}$.
  end if
end for
Exhaustive search over a space of constant $2c_1$ number of candidate locations.
Set the singleton = 'false'.
Set initial location estimate $\hat{\ell} = round(\omega_{prev} n/2\pi) - c_1$.
Set the energy threshold $T = (1+\gamma)D$ for an appropriately chosen γ (see Appendix **??**).
for each position $q = \ell, \ldots, \ell + 2c_1$ **do**
  $v_q = \vec{a}(q)^\dagger \vec{y}/D$.
  if $||\vec{y} - v_q \vec{a}(q)||^2 < T$ **then**
    singleton = 'true'.
    $p = q$ and $v_p = v_q$.
  end if
end for

D. Analysis of the FSURE algorithm

Consider a singleton bin with $D = NC$, observations obtained using the measurement matrix consisting of n steering vectors (of the form (19)) as columns. Let $N = c_1 \log n$, for some constant c_1, be the number of measurements per cluster and $C = \log n - \log\log n - \log c_1$ be the number of clusters. Then,

Proposition VI.2. *For a sufficiently high SNR and large n, the FSURE algorithm correctly identifies the location and the value of a singleton bin with probability at least $1 - O(1/n^2)$. The FSURE algorithm uses $D < O(\log^2 n)$ carefully chosen bin-measurements and no more than $O(\log^2 n)$ number of complex*

operations.

Proof: Please see Appendix F ■

VII. Simulations

In Section I, we have shown that the R-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 R-FFAST algorithm on synthetic data and show that the empirical results are in close confirmation with the theoretical claims of Theorem III.1.

The R-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,

$$\begin{aligned} & \#\text{ of samples } m \text{ required by R-FFAST} \\ = \; & \{\#\text{ of bins required for successful decoding} \\ & \text{of the resulting sparse graph}\} \times \\ & \{\#\text{ of delay-chains required for robust bin-processing}\} \\ = \; & \{c_1(\delta)k\} \times \{c_2(\rho)c_3 \log n\} \end{aligned} \tag{23}$$

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 R-FFAST algorithm to reliably compute the DFT $\vec{X}$.

A. Sample complexity m as a function of n

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

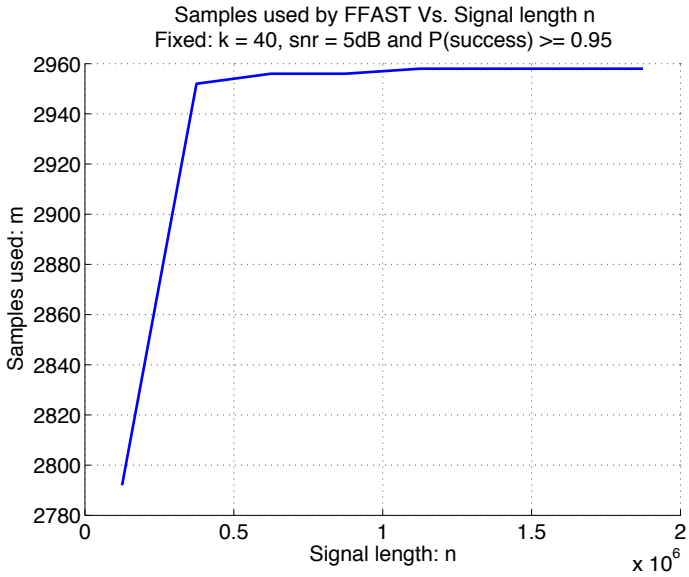


Fig. 5. The plot shows the scaling of the number of samples m required by the R-FFAST algorithm to reliably reconstruct a $k = 40$ sparse DFT $\vec{X}$, from noise-corrupted observations $\vec{y} = \vec{x} + \vec{z}$, for increasing signal length n. For a fixed reliability, signal-to-noise ratio and sparsity, we note that m scales logarithmically with increasing n. This is consistent with the claims of Theorem III.1.

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 $\{\pm\sqrt{\rho}\}$[2] randomly. The input signal $\vec{x}$ is obtained by taking the IDFT of $\vec{X}$. The length 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 $\vec{y} = \vec{x} + \vec{z}$, is generated by adding zero-mean, unit variance complex white Gaussian noise $\vec{z}$ with ρ chosen to have signal-to-noise ratio of 5dB.
- The noise-corrupted signal $\vec{y}$ is input to a $d = 3$ stage R-FFAST architecture with D delay-chains per stage (see Fig. 3). The sampling periods of the 3 stages in the R-FFAST architecture are $50*51$, $51*49$ and $49*50$ respectively. This results in the number of samples per delay-chain, for the

[2]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.

three stages to be $f_0 = 49$, $f_1 = 50$ and $f_2 = 51$ respectively. Thus the total number of bins in the R-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 delay-chains 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 R-FFAST are $m = n_b * D = 3.75 * k * D$.
- Each sample point in the plot is generated by averaging the performance of R-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

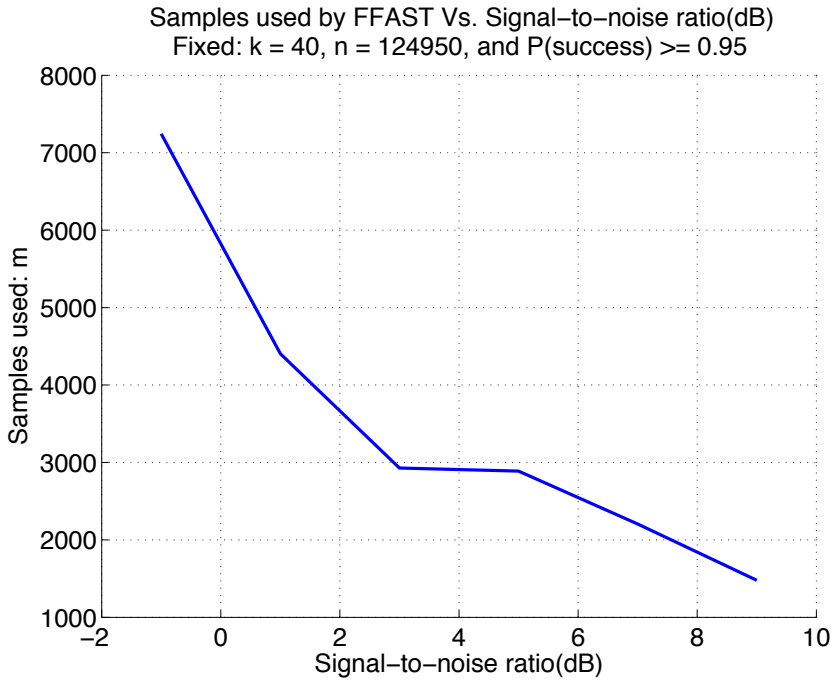


Fig. 6. The plot shows the scaling of the number of samples m required by the R-FFAST algorithm to reliably compute a $n \approx 0.1$ million, $k = 40$ sparse DFT $\vec{X}$, from noise-corrupted observations $\vec{y} = \vec{x} + \vec{z}$, for increasing signal-to-noise ratio ρ. For fixed values of all other parameters, the number of samples decreases roughly in an *inverse relation* with the increasing signal-to-noise ratio on log-scale.

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:

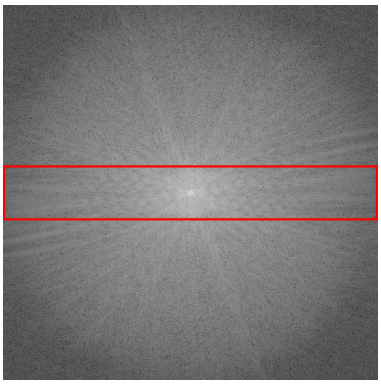
- An $n = 49 * 50 * 51 \approx 0.1$ million length, DFT vector $\vec{X}$, with $k = 40$ non-zero coefficients is generated. The non-zero DFT coefficients are chosen to have uniformly random support from the set $\{0, \ldots, n-1\}$, and the values from the set $\{\pm\sqrt{\rho}\}$. The input signal $\vec{x}$ is obtained by taking IDFT of $\vec{X}$.
- A noise-corrupted signal $\vec{y} = \vec{x} + \vec{z}$, is generated by adding zero-mean, unit variance complex white Gaussian noise $\vec{z}$.
- 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 R-FFAST architecture with D delay-chains per stage (see Fig. 3). The sampling periods of the 3 stages in the R-FFAST architecture are $50 * 51$, $51 * 49$ and $49 * 50$ respectively. This results in the number of samples per delay-chain, 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 R-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 delay-chains 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 R-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.

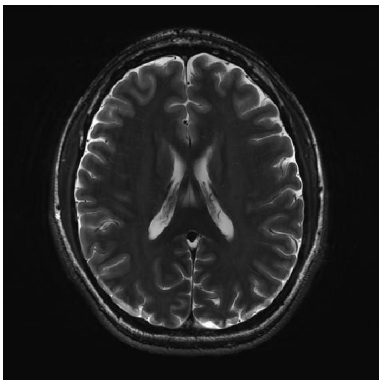
C. Application of the R-FFAST for MR imaging

The $1D$ R-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 $2D$ R-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 from significantly less number of Fourier samples. To reconstruct the full brain image using $2D$ R-FFAST, we perform the following two-step procedure:

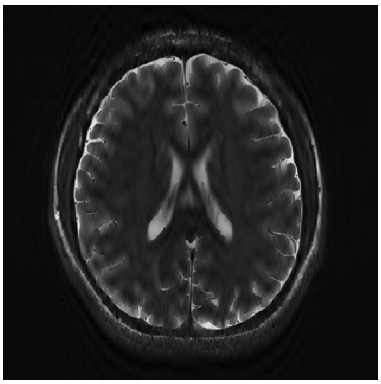
- *Differential space signal acquisition*: We perform a vertical finite difference operation on the image by multiplying the $2D$-DFT signal with $1 - e^{2\pi\imath\omega_0}$. This operation effectively creates an approximately sparse differential image in spatial domain, as shown in Fig. 7(e), and can be reconstructed using R-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



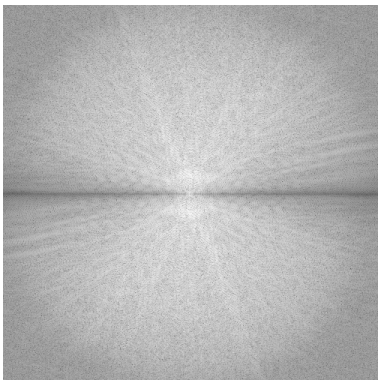
(a) Log intensity plot of the $2D$-DFTof the original 'Brain' image. The red enclosed region is fully sampled and used for the stable inversion.



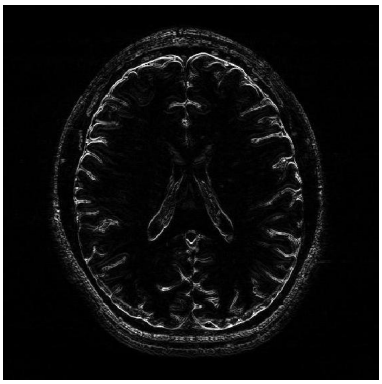
(b) Original 'Brain' image in spatial domain.



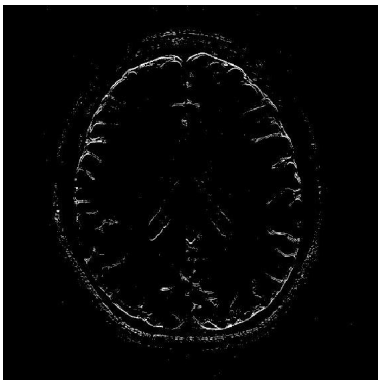
(c) Reconstructed 'Brain' image using the $2D$ R-FFAST architecture along with the fully sampled center frequencies. The total number of Fourier samples used is 60.18%.



(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 $2D$ R-FFAST architecture with 56.71% of Fourier samples.

Fig. 7. Application of the $2D$ R-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. 7(e), using $d = 3$ stage $2D$ R-FFAST architecture with 15 random circular shifts in each of the 3 stages of the R-FFAST architecture. Additionally we acquire all the Fourier samples from the center frequency as shown, by the red enclosure, in Fig. 7(a). Then, we do a stable inversion using the reconstructed differential 'Brain' image of Fig. 7(f) and the fully sampled center frequencies of Fig. 7(a), to get a reconstructed full 'Brain' image as shown in Fig. 7(c). Our proposed two-step acquisition and reconstruction procedure takes overall 60.18% of Fourier samples.

is then sub-sampled and reconstructed using a 3-stage $2D$ R-FFAST architecture. Also, since the brain image is approximately sparse, we take 15 delay-chains in each of the 3 stages of the $2D$ R-FFAST architecture. The R-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. 7(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 $2D$ R-FFAST algorithm uses a total of 60.18% of Fourier samples to reconstruct the brain image shown in Fig. 7(c).

APPENDIX A
MUTUAL INCOHERENCE BOUND

In this section we provide a proof of Lemma V.4. 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 (7). Then,

$$\begin{aligned}\mu_{\max}(\boldsymbol{A}_{i,j}) &\leq \max_{p\neq q} \frac{1}{D}|\vec{a}(p)^\dagger \vec{a}(q)| \\ &= \max_{\ell\neq 0} \frac{1}{D}|\sum_{s=0}^{D-1} \exp(\imath 2\pi \ell r_s/n)| \\ &= \max_{\ell\neq 0} \mu(\ell)\ , \end{aligned} \tag{24}$$

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,

$$Pr(\mu(\ell) > \sqrt{2}t) \leq 4\exp(-t^2 D/2). \tag{25}$$

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 circular shifts $r_0, \ldots, r_{D-1}$, 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 R-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 [24], 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 circular shifts $r_0,\ldots,r_D$, 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 R-FFAST reconstructs the DFT $\vec{X}$, with probability at least $1-\varepsilon$ for any $\varepsilon > 0$, using $m = O(k\log(n))$ samples. In the second part of the proof we show that computational complexity of the R-FFAST decoder is $O(n\log(n))$.

A. Reliability Analysis and sample complexity of the R-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 R-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 R-FFAST algorithm in the presence of the observation noise is

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

2) Reliability analysis: In Lemma V.4, we have shown that a random choice of delay parameters $r_0, \ldots, r_{D-1}$ satisfies the upper bound of (10) with probability at least 0.2. Also, it is easy, i.e., $O(nD)$ complexity, to verify if for a given choice of $r_0, \ldots, r_{D-1}$ (10) 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 (10) 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 R-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 R-FFAST to reconstruct the DFT $\vec{X}$ and over $O(k)$ bins used in the R-FFAST architecture to get,

$$\begin{aligned} Pr(E) &< \text{number of iterations} \times \text{number of bins} \times Pr(E_b) \\ &< O(1/k). \end{aligned} \tag{26}$$

Let E_f denote an event that the R-FFAST algorithm fails to reconstruct the DFT $\vec{X}$. Then putting the pieces together, we get,

$$\begin{aligned} Pr(E_f) &< Pr(E) + Pr(E_f \mid \overline{E}) \\ &\overset{(a)}{<} O(1/k) + O(1/k) \\ &< \varepsilon \end{aligned} \tag{27}$$

where in (a) we used the bound from (26) and the results from [1] along with the fact that if there is no error in processing any bin, the R-FFAST algorithm performs identical to the noiseless algorithm FFAST.

Hence, in order to complete the reliability analysis of the R-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 $\vec{z} \sim \mathcal{CN}(0, I_{D\times D})$, we have,*

$$Pr(||\vec{u}+\vec{z}||^2 < (1+\gamma)D) < 2e^{-D\gamma^2/9} + e^{-(||\vec{u}||-\sqrt{2\gamma D})_+^2}, \tag{28}$$

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 R-FFAST architecture. As shown in the Section **??** 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,

$$\begin{aligned} Pr(E_z) &= Pr(||\vec{w}_{b,i,j}||^2 > (1+\gamma)D) \\ &= P(\chi^2_{2D} > 2(1+\gamma)D) \\ &< 2\exp(-D\gamma^2/9) \quad \forall \quad \gamma \in [0, 1/3]. \end{aligned} \tag{29}$$

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,

$$D > 18\log(k)/\gamma^2 \tag{30}$$

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 $X[\ell] = \sqrt{\rho}e^{\iota\phi}$ and the steering vector $||\vec{a}(\ell)|| = \sqrt{D}$. Hence,

$$Pr(E_{sz}) < 2\exp\{-D\gamma^2/9\} + \exp\{-(\sqrt{\rho D} - \sqrt{2\gamma D})_+^2\}. \tag{31}$$

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,

$$\begin{aligned} Pr(E_{ss}) &= Pr(||\vec{y}_{b,i,j} - X[\ell']\vec{a}(\ell')||^2 < (1+\gamma)D) \\ &= Pr(||X[\ell]\vec{a}(\ell) - X[\ell']\vec{a}(\ell') + \vec{w}_{b,i,j}||^2 < (1+\gamma)D) \\ &\overset{(a)}{=} Pr(||\boldsymbol{A}_{i,j}\vec{v} + \vec{w}_{b,i,j}||^2 < (1+\gamma)D) \\ &< 2\exp\{-D\gamma^2/9\} + \exp\{-(||\boldsymbol{A}_{i,j}\vec{v}|| - \sqrt{2\gamma D})_+^2\} \end{aligned} \tag{32}$$

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.4 and Lemma V.5,

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

Thus, bound in (32) becomes,

$$Pr(E_{ss}) < 2\exp\{-D\gamma^2/9\} + \exp\left\{-\left(\sqrt{2\rho D(1 - 2\sqrt{\log(5n)/D})_+} - \sqrt{2\gamma D}\right)_+^2\right\} \tag{33}$$

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,

$$\begin{aligned} Pr(E_{sm}) &< Pr(E_{sm} \mid \hat{X[\ell]} = X[\ell]) + Pr(\hat{X[\ell]} \neq X[\ell]) \\ &= Pr(E_z) + Pr(\hat{X[\ell]} \neq X[\ell]) \end{aligned} \tag{34}$$

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 (29) in (34), we get,

$$Pr(E_{sm}) < 2\exp\{-D\gamma^2/9\} + \exp\{-D\rho\sin^2(\pi/M)\}. \tag{35}$$

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 (31), (33) and (35), to reliably decode a single-ton bin with probability at least $1-O(1/k^2)$ we need

$$D > \max\left\{\frac{2\log(k)}{(\sqrt{\rho}-\sqrt{2\gamma})_+^2}, 16\log(n), \frac{18\log(k)}{\gamma^2}, \frac{2\log(k)}{\rho\sin^2(\pi/M)}\right\} \tag{36}$$

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:

$$\begin{aligned} Pr(E_m) &< Pr(E_m \mid L < O(1)) + Pr(E_m \mid O(1) < L < \log(k)) + Pr(L \geq \log(k)) \\ &< Pr(E_{m,1}) + Pr(E_{m,2}) + O(1/k^2), \end{aligned} \tag{37}$$

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.4, V.5 and the fact that all the non-zero components of $\vec{u}$ are of the form $\sqrt{\rho}e^{\imath\phi}$, we have,

$$\begin{aligned} ||\boldsymbol{A}_{i,j}\vec{v} - X[\ell]\vec{a}(\ell)||^2 &= ||\boldsymbol{A}_{i,j}\vec{u}||^2 \\ &\geq L\rho D(1-\mu_{\max}(\boldsymbol{A}_{i,j})L)_+ \\ &> L\rho D(1-2L\sqrt{\log(5n)/D})_+ \end{aligned} \tag{38}$$

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 (38).

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 a random choice of circular shifts

$$\begin{aligned} \mathbb{E}[\vec{a}_p^\dagger \vec{a}_q] &= \mathbb{E}\left[\sum_{s=0}^{L-1} e^{\imath(2\pi/n)\ell_s(q-p)}\right] \\ &= 0 \end{aligned} \tag{39}$$

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,

$$\begin{aligned} Pr(||\vec{y}_b||^2 < (1-\gamma)D(\rho L+1)) &= Pr(\chi^2_{2D} < 2D(1-\gamma)) \\ &< 2e^{-D\gamma^2/4} \\ &< O(1/k^2) \end{aligned} \tag{40}$$

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 (30), (36) and (37) we have,

$$Pr(E_b) < O(1/k^2), \tag{41}$$

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

B. Computational complexity of R-FFAST

The computational cost of the R-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 R-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 delay-chains $D = O(\log n)$ is sufficient for reliable reconstruction. Thus, the overall computational complexity of the R-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 $\vec{z} \in \mathcal{CN}(0, I_{D\times D})$. The energy detector fails to identify the presence of the vector $\vec{u}$, if $||\vec{u}+\vec{z}||^2 < (1+\gamma)D$, where $0<\gamma<1$ is a constant. Next, we analyze the probability of this event.

$$\begin{aligned}
& Pr(||\vec{u}+\vec{z}||^2 < (1+\gamma)D) \\
= \; & Pr(||\sqrt{2}\vec{u}+\sqrt{2}\vec{z}||^2 < 2D(1+\gamma)) \\
\overset{(a)}{=} \; & Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 + \chi^2_{2D-1} < 2D(1+\gamma)) \\
< \; & Pr(\chi^2_{2D-1} < 2D(1-\gamma)) + Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 2D(1+\gamma) - 2D(1-\gamma)) \\
= \; & Pr(\chi^2_{2D-1} < 2D(1-\gamma)) + Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 4D\gamma), \qquad (42)
\end{aligned}$$

in (a) we did a change of basis such that $\vec{u}/||\vec{u}||$ is one of the basis vectors. Since $\vec{z}$ 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,

$$Pr(\chi^2_{2D-1} < 2D(1-\gamma)) < 2e^{-D\gamma^2/9}. \qquad (43)$$

Now consider the second term in (42) as,

$$\begin{aligned}
& Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1)^2 < 4D\gamma) \\
< \; & Pr(\mathcal{N}(\sqrt{2}||\vec{u}||,1) < 2\sqrt{D\gamma})) \\
= \; & Pr(\mathcal{N}(0,1) < -(\sqrt{2}||\vec{u}|| - 2\sqrt{D\gamma})) \\
< \; & 2Pr(\mathcal{N}(0,1) < -(\sqrt{2}||\vec{u}|| - 2\sqrt{D\gamma})_+) \\
< \; & \exp\left\{-\left(||\vec{u}|| - \sqrt{2D\gamma}\right)_+^2\right\}. \qquad (44)
\end{aligned}$$

Substituting (43) and (44) in (42), we get,

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

■

APPENDIX E
PROOF OF PROPOSITION VI.1

Consider $N = c_1 \log n$ noise-corrupted time-domain samples of a single complex sinusoid, obtained by sampling the signal at a periodic interval of 1,

$$y(t) = Ae^{j(\omega t+\phi)} + w(t), \quad t = 0, 1, 2, \ldots, (N-1), \tag{45}$$

where the amplitude A, frequency ω, and the phase ϕ are deterministic but unknown constants. The noise samples $w(t) \sim \mathcal{CN}(0,1)$. Then, as shown in (18) the MMSE estimate of the unknown frequency ω is given by,

$$\hat{\omega} = \omega + \mathcal{N}\left(0, \frac{6/\rho}{N(N^2-1)}\right), \tag{46}$$

where $\rho = A^2$. Then,

$$\begin{aligned}
Pr\left(|\omega-\hat{\omega}| > \frac{\pi}{\log n}\right) &= Pr\left(\left|\mathcal{N}\left(0, \frac{6/\rho}{N(N^2-1)}\right)\right| > \frac{\pi}{\log n}\right) \\
&= Pr\left(|\mathcal{N}(0,1)| > \sqrt{\frac{\pi^2}{\log^2 n}\frac{N(N^2-1)}{6/\rho}}\right) \\
&= 2Q\left(\sqrt{\frac{\pi^2}{\log^2 n}\frac{N(N^2-1)}{6/\rho}}\right) \\
&< \exp\left(-\frac{\pi^2}{\log^2 n}\frac{N(N^2-1)}{6/\rho}\right) \\
&< 1/n^3,
\end{aligned} \tag{47}$$

for an appropriate choice of constant c_1, where $N = c_1 \log n$.

APPENDIX F
PROOF OF PROPOSITION VI.2

Let ω_i be an estimate of $2^i\omega$ obtained using the observations of cluster i as per estimation rule (17). Also let $\Omega_i := (\omega_i - \pi/\log n, \omega_i + \pi/\log n)$, i.e., $|\Omega_i| = 2\pi/\log n$. Define E_i an event that $(2^i\omega)_{2\pi} \notin \Omega_i$,

where ω is the true unknown frequency, of the form $2\pi\ell/n$, corresponding to the location ℓ of a singleton bin. Then, from proposition VI.1 we have $Pr(E_i) = Pr(E_0) < 1/n^3$ for some constant c_1 such that $N = c_1 \log n$. Thus, applying a union bound over C clusters, we get,

$$\begin{aligned} Pr(\omega \notin |\cap_{i=0}^{C-1} \Omega_i/2^i|) &\leq \sum_{i=0}^{C-1} Pr(E_i) \\ &< C/n^3 \\ &< 1/n^2, \end{aligned} \tag{48}$$

where in the last inequality we used the value of $C = \log n - \log\log n - \log c_1$.

From (22) we know,

$$\begin{aligned} |\cap_{i=0}^{C-1} \Omega_i/2^i| &\leq \frac{2\pi}{2^{C-1}\log n} \\ &= \frac{2\pi}{n} 2c_1. \end{aligned} \tag{49}$$

Hence using (48), (49) and the fact that there are total of n possible frequencies $2\pi\ell/n, \ell = 0, \ldots, n-1$, we conclude that the FSURE successive refinement algorithm provides an estimate of the location ℓ of a singleton bin within a range of $2c_1$ candidate locations with probability at least $1 - 1/n^2$. Lemma **??** in Section **??** implies that the exhaustive search among these final $2c_1$ candidate locations provides the true location of the non-zero coefficient of a singleton bin with probability at least $1 - O(1/k^2)$.

e) Sample Complexity: The sample complexity of the FSURE algorithm is

$$\begin{aligned} D &= NC \\ &= c_1 \log n(\log n - \log\log n - \log c_1) \\ &< O(\log^2 n). \end{aligned}$$

f) Computational Complexity: The computational complexity of processing each cluster of observations is $N = c_1 \log n$. There are total of C clusters. Hence, FSURE algorithm determines the final range of $2c_1$ candidate locations of a singleton bin in no more than NC arithmetic operations. Further, the complexity of performing an exhaustive search is $O(NC)$. Hence, the overall complexity of the FSURE algorithm is $O(NC) = O(\log^2 n)$.

APPENDIX G
PROOF OF THEOREM III.3

The sample set of the enhanced R-FFAST framework is a super-set of the samples of R-FFAST. Hence all the reliability analysis of all the component failure events analyzed in Appendix C continues to hold. The only new component in the enhanced R-FFAST algorithm is the use of the fast FSURE bin-processing algorithm 3. Also, in proposition VI.2, we have shown that FSURE algorithm correctly identifies the location and the value of a singleton bin with probability at least $1 - O(1/n^2)$. Thus, completing the reliability analysis of the enhanced R-FFAST algorithm.

Next, we analyze the sample and the computational complexity of the enhanced R-FFAST algorithm.

- *Sample complexity*: The enhanced R-FFAST algorithm uses $O(k)$ bins and $D = O(\log^2 n)$ samples per bin. Thus, the total sample complexity of the enhanced R-FFAST algorithm is $m = O(k \log^2 n)$.
- *Computational complexity*: In proposition VI.2, we have shown that the computational complexity of the bin-processing algorithm FSURE is $O(\log^2 n)$. The peeling-process completes in the constant number of iterations. Hence, the total back-end decoding complexity of the enhanced R-FFAST algorithm is $O(k \log^2 n)$. On the other hand, the computational complexity of the enhanced R-FFAST front-end is that of computing $O(k \log^2 n)$, k-point DFTs, i.e., $O(k \log^3 n)$. ■

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

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

[4] 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.

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

[6] 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.

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

[8] 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.

[9] 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.

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

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

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

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

[20] B. Ghazi, H. Hassanieh, P. Indyk, D. Katabi, E. Price, and L. Shi, "Sample-optimal average-case sparse fourier transform in two dimensions," *arXiv preprint arXiv:1303.1209*, 2013.

[21] 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.

[22] S. Kay, "A fast and accurate single frequency estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 12, pp. 1987–1990, 1989.

[23] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression (corresp.)," *Information*

Theory, IEEE Transactions on, vol. 31, no. 6, pp. 832–835, 1985.

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