Robustifying the Sparse Walsh-Hadamard Transform without Increasing the Sample Complexity of $O(K \log N)$

Xiao Li, Joseph Kurata Bradley, Sameer Pawar and Kannan Ramchandran
Dept. of Electrical Engineering and Computer Sciences, U.C. Berkeley.

***Abstract*—The problem of computing a K-sparse N-point Walsh-Hadamard Transforms (WHTs) from noisy time domain samples is considered, where $K = \mathcal{O}(N^\alpha)$ scales sub-linearly in N for some $\alpha \in (0, 1)$. A robust algorithm is proposed to recover the sparse WHT coefficients in a stable manner which is robust to additive Gaussian noise. In particular, it is shown that the K-sparse WHT of the signal can be reconstructed from noisy time domain samples with any error probability ϵ_N which vanishes to zero, using the same sample complexity $\mathcal{O}(K \log N)$ as in the noiseless case.**

I. Introduction

The Walsh-Hadamard Transform (WHT) has been widely deployed in image compression [1], spreading code design in multiuser systems such as CDMA and GPS [2], and compressive sensing [3]. The WHT may be computed using N samples and $N \log N$ operations via a recursive algorithm [4], [5] analogous to the Fast Fourier Transform (FFT). However, these costs can be significantly reduced if the signal is sparse in the WHT domain, as is true in many real world scenarios [6], [7].

Since the WHT is a special case of the multidimensional DFT over a finite field $\mathbb{F}_2^n$, recent advances in computing K-sparse N-point Fourier transforms have provided insights in designing algorithms for computing sparse WHTs. There has been much recent work in computing a sparse Discrete Fourier Transform (DFT) [8]–[13]. Among these works, the Fast Fourier Aliasing-based Sparse Transform (FFAST) algorithm proposed in [13] uses $\mathcal{O}(K)$ samples and $\mathcal{O}(K \log K)$ operations for any sparsity regime $K = \mathcal{O}(N^\alpha)$ with $\alpha \in (0, 1)$ under a uniform sparsity distribution. Following the sparse-graph decoding design in [13] for DFTs, the Sparse Fast Hadamard Transform (SparseFHT) algorithm developed in [14] computes a K-sparse N-point WHT with $K = \mathcal{O}(N^\alpha)$ using $\mathcal{O}(K \log N)$ samples. Since K is sub-linear in N, their results can be interpreted as achieving a sample complexity $\mathcal{O}(K \log N)$. However, the algorithm specifically exploits the noiseless nature of the underlying signals and hence fails to work in the presence of noise.

In this paper, we consider the problem of computing a K-sparse N-point WHT *in the presence of noise*. A key question of theoretical and practical interest is: what price must be paid to be robust to noise? *Surprisingly, there is no cost in sample complexity to being robust to noise, other than a constant factor.* Specifically, we develop a robust algorithm which uses $\mathcal{O}(K \log N)$ samples and has strong performance guarantees. We prove that our algorithm can recover the sparse WHT at constant signal-to-noise ratios (SNRs) with the same $\mathcal{O}(K \log N)$ samples as for the noiseless case in [14]. This result contrasts with the DFT work in [15], for which robustness to noise increases the sample complexity from $\mathcal{O}(K)$ to $\mathcal{O}(K \log N)$.

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 our main results and a brief comparison with related literature. In Section IV, we explain the proposed front-end architecture for acquiring samples and the robust algorithm using a simple example. In Section V, we provide simulation results which empirically validate the performance of our algorithm.

Notations: Throughout this paper, the set of integers $\{0, 1, \cdots, N-1\}$ for some integer N is denoted by $[N]$. Lowercase letters, such as x, are used for the time domain expressions and the capital letters, such as X, are used for the transform domain signal. Any letter with a bar such as $\bar{x}$ or $\bar{X}$ represents a vector containing the corresponding samples. Given a real-valued vector $\bar{v} \in \mathbb{R}^N$ with $N = 2^n$, the i-th entry of $\bar{v}$ is interchangeably represented by $v[i]$ indexed by the decimal representation of i or $v_{i_0,i_1,\cdots,i_{n-1}}$ indexed by the binary representation of i, where $i_0, i_1, \cdots, i_{n-1}$ denotes the binary expansion of i with i_0 and i_{n-1} being the least significant bit (LSB) and the most significant bit (MSB), respectively. The notation $\mathbb{F}_2$ refers to the finite field consisting of $\{0, 1\}$, with defined operations such as summation and multiplication modulo 2. Furthermore, we let $\mathbb{F}_2^n$ be the n-dimensional vector with each element from $\mathbb{F}_2$ and the addition of the vectors done element-wise over this field. The inner product of two binary indices i and j is defined by $\langle i, j \rangle = \sum_{t=0}^{n-1} i_t j_t$ with arithmetic over $\mathbb{F}_2$, and the inner product between two vectors is defined as $\langle \bar{x}, \bar{y} \rangle = \sum_{t=1}^{N} x[t]y[t]$ with arithmetic over $\mathbb{R}$.

II. Signal Model and Problem Formulation

Consider a signal $\bar{x} \in \mathbb{R}^N$ containing $N = 2^n$ samples x_m indexed with elements $m \in \mathbb{F}_2^n$, and the corresponding WHT $\bar{X} \in \mathbb{R}^N$ containing N coefficients X_k with $k \in \mathbb{F}_2^n$. The N-dimensional WHT $\bar{X}$ of the signal $\bar{x}$ is given by

$$X_k = \frac{1}{\sqrt{N}} \sum_{m \in \mathbb{F}_2^n} (-1)^{\langle k, m \rangle} x_m, \tag{1}$$

where $k \in \mathbb{F}_2^n$ denotes the corresponding index in the transform domain. We assume the WHT is a sub-linearly sparse

signal with $K = N^\alpha$ non-zero coefficients X_k in the set $k \in \mathcal{K}$ and $\alpha \in (0, 1)$.

Previous analysis [14] assumes exact measurements of the time-domain signal $\bar{x}$. We generalize this setting by using noise-corrupted measurements:

$$y_m = x_m + w_m, \tag{2}$$

where $w_m \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise added to the clean samples x_m. The SparseFHT algorithm [14] no longer works in the presence of noise. Therefore, the focus of this paper is to develop a robust algorithm which can compute the sparse WHT coefficients $\{X_k\}_{k\in\mathcal{K}}$ reliably from the noisy samples y_m with the same sample complexity as in the noiseless case.

III. Related Work and Our Results

In this section, we first frame our results in the context of previous work on recovering sparse transforms. We then summarize our main results.

A. Related Work

Due to the similarities between the Discrete Fourier Transform (DFT) and the WHT, we give a brief account of previous work on reducing the sample and computational complexity of computing a K-sparse N-point DFT. [8], [9] developed randomized sub-linear time algorithms that achieve near-optimal sample and computational complexities of $\mathcal{O}(K \log N)$ with potentially large big-Oh constants [11]. Then, [10] further improved the algorithm for 2-D Discrete Fourier Transforms (DFTs) with $K = \sqrt{N}$, which reduces the sample complexity to $\mathcal{O}(K)$ and the computational complexity to $\mathcal{O}(K \log K)$, albeit with a constant failure probability that does not vanish as the signal dimension N grows. On this front, the deterministic algorithm in [12] is shown to guarantee zero errors but with complexities of $\mathcal{O}(\text{poly}(K, \log N))$.

A major improvement in terms of both complexities is given by the FFAST algorithm [13], which achieves a vanishing failure probability using only $\mathcal{O}(K)$ samples and $\mathcal{O}(K \log K)$ operations for any sparsity regime $K = \mathcal{O}(N^\alpha)$ and $\alpha \in (0, 1)$. The success of the FFAST algorithm is thanks to peeling-based decoding over sparse graphs, which depends on the *single-ton test* to pinpoint the "parity" Fourier bin containing only one "erasure event" (unknown non-zero DFT coefficient). Given such a single-ton bin, the value and location of the coefficient can be obtained and then removed from other "parity" bins. This procedure iterates until no more single-ton bins are found.

Inspired by [13], the SparseFHT algorithm in [14] computes a K-sparse WHT of $\bar{x}$ using $\mathcal{O}(K \log N)$ samples and $\mathcal{O}(K \log^2 N)$ operations[1]. The tenet of the algorithm is again to intelligently subsample the multidimensional signal to create hashing/aliasing patterns in the transform domain bins. Similar to the single-ton test in [13], the SparseFHT algorithm critically relies on the *collision detection* module to identify parity bins which contain only one unknown WHT coefficient. Since both the single-ton test in [13] and the collision detection in [14] specifically exploit the noiseless nature of signals, they cannot be used in the noisy setting without major algorithmic changes. Our work fills this gap by developing a sparse WHT algorithm which is robust to noise.

[1] In [14], the result suggests a requirement of $\mathcal{O}(K \log(N/K))$ samples and $\mathcal{O}(K \log K \log(N/K))$ operations, which is equivalent to $\mathcal{O}(K \log N)$ and $\mathcal{O}(K \log^2 N)$, respectively, since $K = N^\alpha$ for a fixed constant $\alpha \in (0, 1)$.

B. Our Results

We now summarize our main results on recovering a K-sparse N-point WHT of a signal from noisy time domain samples. For our analysis, we make the following assumptions:

- The support of the non-zero WHT coefficients is uniformly random in the set $[N]$.
- The unknown WHT coefficients take values from $\pm\rho$.
- The signal-to-noise ratio $\text{SNR} = \frac{\rho^2}{N\sigma^2}$ is fixed.

The first assumption is critical to analyzing the peeling decoder. The next two assumptions merely simplify analysis.

Theorem 1. *For any sublinear sparsity regime $K = \mathcal{O}(N^\alpha)$ for $\alpha \in (0, 1)$, our robust algorithm based on the* **randomized hashing** *front-end (Section IV-A) and the associated* **peeling-based decoder** *(Section IV-B) can stably compute the WHT $\bar{X}$ of any signal $\bar{x}$ in the presence of noise $w \sim \mathcal{N}(0, \sigma^2 I_{N\times N})$, with the following properties:*

1) **Sample complexity:** *The algorithm needs $\mathcal{O}(K \log N)$ noisy samples y_m.*
2) **Computational complexity:** *The algorithm requires $\mathcal{O}(N \log^2 N)$ operations.*
3) **Probability of success:** *The algorithm successfully computes the K-sparse WHT $\bar{X}$ with probability at least $1 - \epsilon_N$ for any $\epsilon_N > 0$.*

Proof. See Appendix A. □

Importantly, the proposed robust algorithm can compute the sparse WHT using $\mathcal{O}(K \log N)$ samples, i.e., no more than the SparseFHT algorithm [14] developed for the noiseless case. The overhead in moving from the noiseless to the noisy regime is only in the extra computational complexity.

IV. Stable Fast Walsh-Hadamard Transform via Robust Sparse Graph Decoding

We now describe our *randomized hashing* front-end architecture and the associated *peeling-based decoding algorithm* for computing a K-sparse N-point WHT, which we then connect to the framework of decoding over sparse-graph codes.

A. Randomized Hashing

Our algorithm is based on subsampling to create aliasing patterns, similarly to the SparseFHT algorithm in [14] and the FFAST algorithm in [13]. After subsampling $B = \mathcal{O}(K)$ time domain samples, computing the corresponding B-point WHT creates "bins" of coefficients from the original N-point WHT. Each of these hashed (aliased) WHT coefficients (bins) is composed of zero, one, or several coefficients from the original WHT. Each subsample of B time domain samples is called a *stage*; the hashing front-end consists of C stages, each of which uses a different subsampling matrix $\Psi_c \in \mathbb{F}_2^{n\times b}$ which has rank b.

Algorithm 1 Subsampling and Shifting

`Input:` Noisy time domain samples $y_m \in \mathbb{R}^N$. subsampling matrix $\Psi_c \in \mathbb{F}_2^{n\times b}$.
`Set:` Random shift $d_{c,p}$, where $c \in [C]$, $p \in [P]$.
return Length-B time-domain vector indexed by $\ell \in \mathbb{F}_2^b$:

$$u_{c,p}[\ell] \quad = \quad \sqrt{\frac{N}{B}} y_{\Psi_c \ell + d_{c,p}} \tag{3}$$

The subsample for each stage c is *shifted* by P different binary patterns $d_{c,p} \in \mathbb{F}_2^n$, $p \in [P]$; we call these "substreams." The key change from [13], [14] is that, rather than using deterministic shifts, we use randomized shifts which make our algorithm robust to noise.

We summarize the subsampling and shifting procedure in Algorithm 1. The following proposition describes how original WHT coefficients are hashed to bins.

Proposition 1. *(Randomized Hashing) Suppose that in the c-th stage p-th substream, the noisy time domain samples y_m are subsampled by $\Psi_c \in \mathbb{F}_2^{n\times b}$ and shifted by $d_{c,p}$, as in Algorithm 1. Let $u_{c,p}[\ell]$ be the resulting length-B time-domain vector. Then the B-point WHT of $u_{c,p}[\ell]$ may be written as:*

$$U_{c,p}[j] = \sum_{k\in\mathbb{F}_2^n | h_c(k)=j} X_k (-1)^{\langle d_{c,p},k\rangle} + \xi_{c,p}[j], \tag{4}$$

where $h_c(k) = \Psi_c^T k$ denotes the hash function and where

$$\xi_{c,p}[j] = \frac{\sqrt{N}}{B} \sum_{\ell\in\mathbb{F}_2^b} (-1)^{\langle j,\ell\rangle} w_{\Psi_c\ell + d_{c,p}} \tag{5}$$

is the compound Gaussian noise $\xi_{c,p}[j] \sim \mathcal{N}(0, N\sigma^2/B)$.

Proof. The proof follows from the properties of WHT, similarly to that in [14], and hence is omitted here. □

The hash function $h_c(k)$ maps original WHT coefficients X_k to the hash bin j. The shifts $d_{c,p}$ change the sign of the contribution of each original coefficient X_k to its bin.

1) A Simple Example: A simple example of the randomized hashing front-end is shown in Fig. 1 with $N = 2^n = 64$ and sparsity level $K = B = 2^b = 4$. Suppose the 4 non-zero WHT coefficients of the signal $\bar{X}$ are $X[4], X[8], X[17]$ and $X[62]$. Here the decimal representation $X[k_{10}]$ of X_k is used for convenience: e.g., $X[4] = X_{000100}$. The randomized hashing front-end subsamples the input signal and its shifted version through $C = 2$ stages. The signal is shifted using the random 6-bit patterns $d_{c,p}$.

For illustration, we show the second substream $p = 1$ in stage $c = 0$, where the associated random shift is chosen as $d_{0,1} = [4]_2 = 000100$. The subsampling matrix for stage $c = 0$ is $\Psi_0 = [0_{4\times2}^T, I_{2\times2}^T]^T$, which freezes the 4 MSB. Thus, substream $p = 1$ in this stage is obtained as

$$u_{0,1}[0] = y_{000000+d_{0,1}} = y[4],\ u_{0,1}[1] = y_{000001+d_{0,1}} = y[5],$$
$$u_{0,1}[2] = y_{000010+d_{0,1}} = y[6],\ u_{0,1}[3] = y_{000011+d_{0,1}} = y[7].$$

The second sampling matrix Ψ_1 freezes the 2 MSB and 2 LSB, whose subsampled outputs are shown in Fig. 1. These

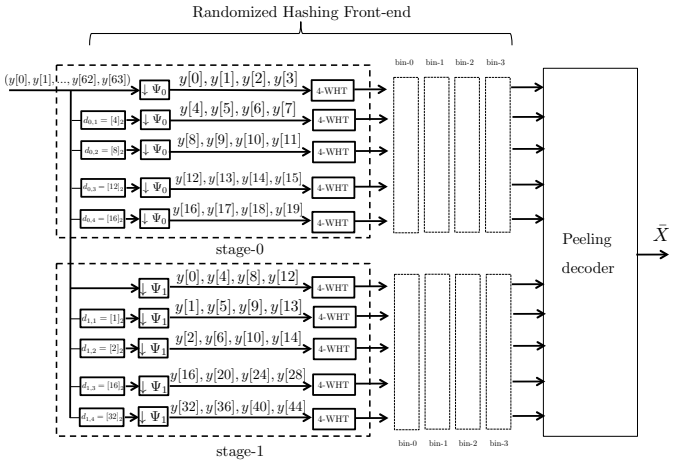

Fig. 1. Consider a $N = 2^n = 64$ length signal $\bar{x}$ that has a $K = 2^b = 4$ sparse WHT, i.e., $n = 6, b = 2$. The signal $\bar{x}$ is processed through a 2 stage FHT architecture. In general there are 3 or more stages but here for purpose of illustration we show an example architecture with 2 stages. The subsampling operation is performed using matrices Ψ_0 in stage-1 and Ψ_1 in stage-2. The first sampling matrix Ψ_0 freezes the 4 MSB bits while the second sampling matrix Ψ_1 freezes the 2 MSB and 2 LSB bits. Various shifts are implemented by shifting the signal using the 6 bit patterns $d_{c,p}$, where $d_{c,p}$ is a random shift. The peeling-decoder then synthesizes the big WHT $\bar{X}$, from the short WHT's of each of the subsampled data streams.

substreams, each containing $B = 2^b = 4$ subsamples, are then passed to a B-point WHT to obtain the hash observations. The output of the short WHT for this particular substream $(c = 0, p = 1)$ is:

$$U_{0,1}[0] = X[0] - X[4] + \cdots - X[60] + \xi_{0,1}[0]$$
$$U_{0,1}[1] = X[1] - X[5] + \cdots - X[61] + \xi_{0,1}[1]$$
$$U_{0,1}[2] = X[2] - X[6] + \cdots - X[62] + \xi_{0,1}[2]$$
$$U_{0,1}[3] = X[3] - X[7] + \cdots - X[63] + \xi_{0,1}[3].$$

2) General Case using Bin-Measurement Matrices: For the general case, the hash outputs of the P substreams in each stage can be stacked in a length-P vector $\bar{U}_c[j] \triangleq [U_{c,1}[j], \cdots, U_{c,P}[j]]^T$ for each bin j and expressed as

$$\bar{U}_c[j] = G_c[j]\bar{X} + \bar{\xi}_c[j], \tag{6}$$

for $c \in [C]$ and $j \in [B]$, where $\bar{\xi}_c[j] \sim \mathcal{N}(0, (N/B)\sigma^2 I_{P\times P})$ and $G_c[j]$ is the *bin measurement matrix* defined by the random shifts as well as the subsampling operator Ψ_c. Let

$$\bar{g}_k = [(-1)^{\langle d_{c,1},k\rangle}, \cdots, (-1)^{\langle d_{c,P},k\rangle}]^T \in \{-1, 1\}^P \tag{7}$$

be the signature associated with a particular WHT coefficient X_k. Then the k-th column of the bin measurement matrix $G_c[j] \in \mathbb{F}_2^{P\times N}$ of bin j of stage c is given by

$$[G_c[j]]_{(:,k)} = \begin{cases} \bar{g}_k, & \text{if } \Psi_c^T k = j \\ 0_{P\times 1}, & \text{otherwise} \end{cases}. \tag{8}$$

Therefore, the outputs $\bar{U}_c[j]$ from stage c in the randomized hashing front-end at a certain bin j becomes the compressed measurement of the unknown sparse WHT vector $\bar{X}$, with the random bin measurement matrix $G_c[j]$ in (6). Thus, each stage divides the computation of the sparse WHT into multiple

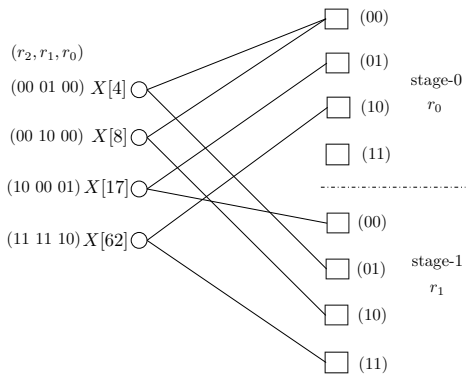

Fig. 2. A 2-left regular degree bi-partite graph representing the relation between the unknown non-zero WHT coefficients and the observations obtained through the architecture shown in Fig. 1, for the 64-point example signal $\bar{x}$. Variable (left) nodes correspond to the non-zero WHT coefficients and the check (right) nodes are the observations.

sparse recovery problems.

To analyze recovery performance, we must place restrictions on $G_c[j]$ to ensure that it is a "good" measurement matrix. Here we adopt the criterion of *mutual coherence* μ between different codewords $\bar{g}_k$ in the non-zero columns of $G_c[j]$, defined as:

$$\mu = \max_{k \neq k'} \frac{1}{P} |\bar{g}_k^T \bar{g}_{k'}|. \tag{9}$$

Intuitively, μ measures similarity between codewords, and a good measurement matrix should have codewords that are as mutually different as possible. We use the following bound on the mutual coherence:

Lemma 1. *The mutual coherence μ is bounded by*

$$\mu \leq \frac{1}{2(L+1)} \tag{10}$$

for some positive integer $L = \mathcal{O}(1)$ with probability at least $1 - \mathcal{O}(N^{-2})$, as long as the number of random shifts is at least $P \geq 24(L+1)^2 \log N$.

Proof. See Appendix B. □

B. Peeling-based Decoder for Stable WHT Recovery

So far, we have described how WHT coefficients are hashed to bins. We now describe how those coefficients are identified and recovered. After explaining the structure of our decoding problem by relating it to sparse graph codes (following [13]), we discuss how to identify coefficients and iteratively "peel" them from the bins.

1) Sparse Graph Codes with Randomized Check Nodes: Each hash observation $\bar{U}_c[j]$ consists of randomized linear combinations of the unknown WHT coefficients $\{X_k\}_{k \in \mathcal{K}}$ in bin j. In terms of sparse graph codes, we need to identify a set of K erasures (coefficients $\{X_k\}$). These erasures/coefficients correspond to K *variable nodes*, and the observations correspond to the *check nodes*, for stages $c \in [C]$. We illustrate this graph decoding problem in Fig. 2, continuing the example from Fig. 1.

Algorithm 2 Robust Bin Identification

`Input:` Noisy observations $\bar{U}_c[j] \in \mathbb{R}^B$ for bin j in stage c.
`Set:` Parameters $\gamma > 0$ and $L \gg 1$.
if $\left\|\bar{U}_c[j]\right\|^2 / P \leq (1+\gamma)\nu^2$ **then**
  return Bin j is a zero-ton
else if $\left\|\bar{U}_c[j]\right\|^2 / P \geq (1+\gamma)(L\rho^2 + \nu^2)$ **then**
  return Bin j is a multi-ton
else
  for $k \in \mathbb{F}_2^b$ **do**
    `Obtain the MLE of the coefficient:`

$$\widehat{X}_k = \frac{1}{P} \bar{g}_k^T \bar{U}_c[j]. \tag{11}$$

  end for
  `Identify the best coefficient:`

$$\widehat{k} = \arg\min_{k \in \mathbb{F}_2^b} \left\| \bar{U}_c[j] - \widehat{X}_k \bar{g}_k \right\|^2 \tag{12}$$

  if $\left\| \bar{U}_c[j] - \widehat{X}_{\widehat{k}} \bar{g}_{\widehat{k}} \right\|^2 / P \leq (1+\gamma)\nu^2$ **then**
    return Coefficient index $\widehat{k}$ and value $\widehat{X}_{\widehat{k}}$
  else
    return Unable to identify bin
  end if
end if

The degree of each check node $\bar{U}_c[j]$ depends on how many non-zero coefficients X_k are hashed into bin j in stage c. Our next goal will be to identify the degree of check nodes, which we categorize as *zero-ton bins* (no non-zero coefficients), *single-ton bins* (one non-zero coefficient), and *multi-ton bins* (multiple non-zero coefficients).

2) Robust Identification of Single-ton Bins: We briefly describe our tests for zero/single/multi-tons, summarized in Algorithm 2. We prove that these tests succeed with high probability in the appendix. For simplicity, we assume that the signal strength ρ and the (hashed) noise variance $\nu^2 = N\sigma^2/B$ are known.

For each type of bin, the observation in bin (c, j) has values:

$$\bar{U}_c[j] = \bar{\xi}_c[j] \qquad \text{(zero-ton)} \tag{13}$$

$$\bar{U}_c[j] = X_k \bar{g}_k + \bar{\xi}_c[j] \qquad \text{(single-ton)} \tag{14}$$

$$\bar{U}_c[j] = \sum_{k \in \text{nonzeros}(c,j)} X_k \bar{g}_k + \bar{\xi}_c[j] \qquad \text{(multi-ton)} \tag{15}$$

For zero-tons and large multi-tons, we can expect the energy $\left\|\bar{U}_c[j]\right\|$ to be small and large, respectively, relative to the energy of a single-ton. Algorithm 2 uses this idea to eliminate zero-tons and large multi-tons. To locate single-tons and distinguish them from small multi-tons, Algorithm 2 uses a Maximum Likelihood Estimate (MLE) test. For each of N/B possible coefficient locations k (for a fixed bin (c, j)), we obtain the MLE for X_k as:

$$\widehat{X}_k = \frac{1}{P} \bar{g}_k^T \bar{U}_c[j]\,. \tag{16}$$

We choose among the locations by finding the location k

which minimizes the residual energy:

$$\widehat{k} = \arg\min_{k} \left\| \bar{U}_c[j] - \widehat{X}_k \bar{g}_k \right\|^2. \tag{17}$$

3) Iterative Decoding: Once we identify a single-ton bin's coefficient (location and value), we can substract its contribution to other bins, possibly creating new single-tons. We detail this iterative method in Algorithm 3. Barring the zero/single/multi-ton testing, peeling may be analyzed the same way as in [13], [14].

Let bin j_c be a single-ton detected in stage c, and let the associated non-zero location be $\widehat{k}$ and coefficient estimate be $\widehat{X}_{\widehat{k}}$ as in (16). For each stage $c' = 1, \cdots, C$, the coefficient $X_{\widehat{k}}$ contributes to bins $j_{c'}$ for which:

$$j_{c'} = \Psi_{c'}^T \widehat{k}, \quad c' = 1, \cdots, C\,. \tag{18}$$

We remove $X_{\widehat{k}}$ from a bin $j_{c'}$ by updating the bin values as:

$$U_{c',p}(j_{c'}) \longleftarrow U_{c',p}(j_{c'}) - \widehat{X}_{\widehat{k}}(-1)^{\langle d_{c',p}, \widehat{k} \rangle}, \quad \forall p. \tag{19}$$

This whole process iterates until no more single-tons are found and K non-zero coefficients have been decoded.

V. Numerical Experiments

We tested our method on samples generated from a sparse WHT signal X of length $N = 2^n$ with $K = 2^b$ randomly positioned non-zero coefficients of magnitude ρ. We added zero-mean Gaussian noise with variance σ^2 to the time-domain signal computed from X. Despite the fact that our analysis is asymptotic, probabilities of failure approach 0 quickly in a range of problem settings.

Sample Complexity: We examine how noise affects the number of samples our algorithm requires. For a fixed problem size (WHT signal of length 2^{14} with 2^6 non-zero coefficients), we varied the SNR and calculated the probability of failure from 100 random problems. In Fig. 3, we can see that the method requires a small constant oversampling factor of about $4\times$ more shifts than the noiseless algorithm from [14] (oversampling factor 1).

Sparsity: We examine how sparsity affects the probability of success in Fig. 4. For a fixed signal length of 2^{14}, we vary the number of non-zero coefficients 2^b. Our method seems to recover dense problems more easily. We posit that this difference is due to the fact that, for a fixed SNR, the expected magnitude of noise in each bin shrinks as $\sqrt{1/B}$; thus, as B increases, it is easier to recognize signals of magnitude ρ amidst the noise. Note that we use $2B$ bins rather than B, which we use in our asymptotic analysis. With B bins, the probability of success does not get quite as close to 0 because of occasional failures in the peeling process.

VI. Conclusions

In this paper, we have proposed a robust algorithm to compute a K-sparse N-point WHT using $\mathcal{O}(K \log N)$ samples generated by the randomized hashing front-end and $\mathcal{O}(N \log^2 N)$ operations. Our approach is based on strategic subsampling of the input noisy signal y_m using a small set of randomly chosen subsampling patterns, which achieves a vanishing failure probability at the same level of complexities as that of the noiseless case.

We are currently developing faster decoding methods to decrease computational complexity. Another valuable but more challenging direction would be to modify the analysis to relax our assumptions, especially the uniformly random sparsity pattern and the assumed knowledge of problem parameters such as SNR.

Algorithm 3 Peeling-based Decoding Algorithm

Require: # of hash blocks C; # of peeling iterations I; Sets of random shifts $d_{c,p}$ for each substream $p \in [P]$ and stage $c \in [C]$.

Ensure: Hash block size $B = \mathcal{O}(K)$ and $P = \beta \log N$ with some sufficiently large constant β.

`Given:` Noisy sequence $\bar{y} = \bar{x} + \bar{\xi} \in \mathbb{R}^N$ with $N = 2^n$, where the WHT of $\bar{x}$ has (unknown) sparsity K.

for $c = 0$ **to** $C - 1$ **do**
  for $p = 0$ **to** $P - 1$ **do**

$$u_{c,p}[\ell] = \sqrt{\frac{N}{B}} y_{\Psi_c \ell + d_{c,p}}, \quad \ell \in \mathbb{F}_2^b \tag{20}$$

$$U_{c,p}[j] = \mathrm{WHT}\left[u_{c,p}[\ell]\right], \quad \ell \in \mathbb{F}_2^b. \tag{21}$$

  end for
end for
for $i = 1$ **to** I **do**
  for $c = 0$ **to** $C - 1$ **do**
    for $j = 0$ **to** $B - 1$ **do**
      `Perform` robust bin identifications in Algorithm 2.
      If not single-ton, continue to next j.
      `Obtain` estimated index $\widehat{k}$ and coefficient $\widehat{X}_{\widehat{k}}$.
      `Infer participating bins`

$$j_{c'} = \Psi_{c'}^T \widehat{k}, \quad c' = 1, \cdots, C.$$

      `Peel Off:`
      for $c' = 0$ **to** $C - 1$ **do**
        Remove the contribution from single-tons

$$U_{c',p}(j_{c'}) \leftarrow U_{c',p}(j_{c'}) - \widehat{X}_{\widehat{k}}(-1)^{\langle d_{c',p}, \widehat{k} \rangle}, \quad \forall p \in [P].$$

      end for
    end for
  end for
end for

Appendix A
Proof of Main Results in Theorem 1

A. Sample Complexity

In [14], it has been shown that for the noiseless case, for any given $0 < \alpha < 1$ and sufficiently large (K, N), their algorithm computes the K-sparse N-length WHT $\bar{X}$, with probability at least $1 - \mathcal{O}(N^{-3/8})$ using a total of $O(K)$ number of bins. Later, we show that $P = \mathcal{O}(\log N)$ number of observations per bin are sufficient to make our algorithm robust against the observation noise. Hence, the total sample

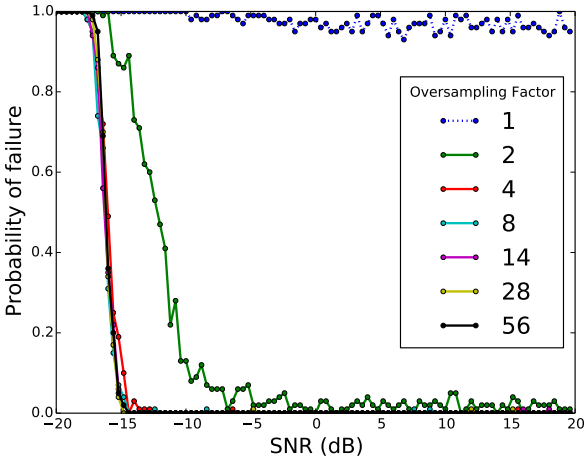

Fig. 3. **Sample complexity:** Our theory for decoding requires that the number P of random delays be on the order of n. These plots of probability of failure (y-axis) vs. SNR (x-axis) show that this value n is fairly tight. The noiseless method from [14] uses $n+b-1$ delays; each plotline uses $(oversampling\,factor)\cdot(n+b-1)$ delays. A small oversampling factor >1 results in high success probability.
Test details: Problems are sparse WHT signals of length $N=2^n$, $n=14$, with $K=2^b$, $b=6$, non-zeros. Algorithm with $C=4$ stages and $2K$ bins. Probability of failure is computed from 100 random problems.

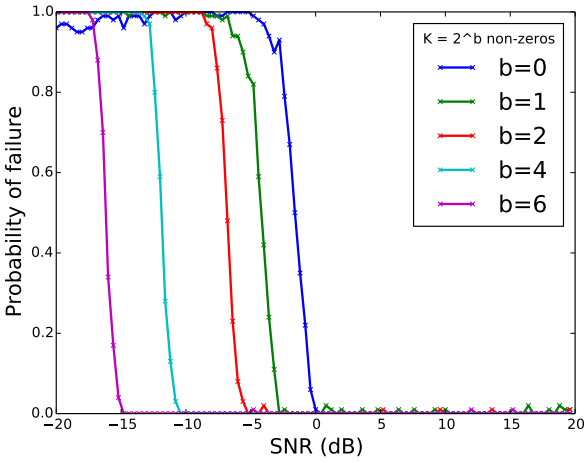

Fig. 4. **Sparsity:** Our algorithm is robust to noise at reasonable SNRs (x-axis), and the probability of failure (y-axis) goes to 0 as SNR increases. Plotlines show different sparsity levels (with 2^b non-zero coefficients); sparser problems seem to require higher SNRs.
Test details: Each test uses twice as many samples as the noiseless algorithm: $P=2(n-b+1)$. Problems are sparse WHT signals of length $N=2^n$, $n=14$, with $K=2^b$ non-zeros. Algorithm with $C=4$ stages and $2K$ bins. Probability of failure is computed from 100 random problems.

complexity of our algorithm in the presence of the observation noise is $\mathcal{O}(K\log N)$.

B. Computational Complexity

The computational cost of our algorithm can be roughly computed as

$$\text{Total \# of arithmetic operations} \tag{22}$$
$$= \text{\# of iterations} \times \text{\# of bins} \times (\text{operations per bin}). \tag{23}$$

Similar to [14], for all values of sparsity index $0<\alpha<1$ our front-end employs no more than $\mathcal{O}(K)$ number of bins and if successful completes decoding in constant number of iterations. Each bin requires first the computation of a K-point WHT, which has complexity $\mathcal{O}(K\log K)$ over $\mathcal{O}(P)$ substreams.

Now, from the pseudocode of bin identification scheme provided in Algorithm 2, it is clear that for each bin, the algorithm performs exhaustive search over $\mathcal{O}(N/B)$ columns of possible signatures $\bar{g}_k$, where each column is of dimension P. Further as shown later, the number of shifts $P=\mathcal{O}(\log N)$ is sufficient for reliable reconstruction. Thus, the overall computational complexity of our algorithm is no more than,

$$\begin{aligned}&\text{Total \# of arithmetic operations}\\ &= \text{\# of iterations} \times \mathcal{O}(K) \times (\mathcal{O}(N/B)\times P)\\ &\quad + \text{\# of iterations} \times \mathcal{O}(K\log K)\times P\\ &= \mathcal{O}(N\log N) + \mathcal{O}(K\log K\log N)\\ &\le \mathcal{O}(N\log^2 N)\end{aligned}$$

where $\log K=\alpha\log N$ is used.

C. Probability of Success

The success rate of the algorithm depends on each bin j to be processed correctly, meaning that each bin is correctly identified as a zero-ton, single-ton or multi-ton. Define $\mathcal{E}_b$ as the the error event where a bin j is decoded wrongly and then, using a union bound over different bins and different iterations, the probability of the algorithm making a mistake in bin identification can be obtained as

$$\mathbb{P}(\mathcal{E}) < \text{number of iterations} \times \text{number of bins} \times \mathbb{P}(\mathcal{E}_b).$$

Furthermore, define $\mathcal{E}_f$ as the error event where the algorithm fails to reconstruct the WHT coefficients $\bar{X}$, then its probability can be obtained as

$$\epsilon_N = \mathbb{P}(\mathcal{E}_f) \tag{24}$$
$$= \mathbb{P}(\mathcal{E}_f|\mathcal{E})\,\mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}_f|\mathcal{E}^c)\,\mathbb{P}(\mathcal{E}^c) \tag{25}$$
$$\le \mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}_f|\mathcal{E}^c). \tag{26}$$

The first term is the error probability of bin processing and the second term is the error probability of the algorithm when it fails to decode all the WHT coefficients given noiseless observations (i.e., bin processing is always correct). According to the result in [14], the error probability of the algorithm with noiseless observations can be upper bounded by $\mathbb{P}(\mathcal{E}_f|\mathcal{E}^c)\le\mathcal{O}(N^{-3/8})$ and therefore, as long as the bin processing error probability can be bounded by

$$\mathbb{P}(\mathcal{E}) < \mathcal{O}(1/N), \tag{27}$$

the overall failure probability remains the same level as the noiseless case [14] such that

$$\lim_{N\to\infty} \epsilon_N = 0. \tag{28}$$

In the following, we focus on showing that the probability in (27) is guaranteed by having a bin-wise error probability

$$\mathbb{P}\left(\mathcal{E}_b\right) < \mathcal{O}(1/N^2), \tag{29}$$

since there are at most $B \approx K$ bins in each iteration and $K < N$. The following section is dedicated to showing that (29) can be guaranteed by having

$$P = \mathcal{O}(\log N) \tag{30}$$

shifts in the randomized hashing front-end.

Probability of Bin Error $\mathbb{P}\left(\mathcal{E}_b\right)$: We define a set $\mathcal{K}_{c,j} \triangleq \{k : h_c(k) = j\} = \{k_1, \cdots, k_{N/B}\}$. Given a specific set of bin measurements $\bar{U}_c[j]$, we have the following possible outcomes from the single-ton identification scheme (i.e., bin processing)

1) It is a zero-ton bin (13).
2) It is a multi-ton bin (15).
3) It is a single-ton bin (14) for some $k \in \mathcal{K}_{c,j}$.

Then the errors made in the identification include the following *miss detection events*

1) Detecting a single-ton as a zero-ton $\mathcal{E}_{z|k_\star}$
2) Detecting a single-ton as another single-ton $\mathcal{E}_{k|k_\star}$
3) Detecting a single-ton as a small L-ton $\mathcal{E}_{m_L|k_\star}$ with $L = \mathcal{O}(1)$
4) Detecting a single-ton as a large L-ton $\mathcal{E}_{M_L|k_\star}$ with $L \neq \mathcal{O}(1)$ and $\lim_{N\to\infty} L = \infty$.

and the following *false detection events*

1) Detecting a zero-ton as a single-ton $\mathcal{E}_{k_\star|z}$
2) Detecting a small L-ton with $L = \mathcal{O}(1)$ as a single-ton $\mathcal{E}_{k_\star|m_L}$
3) Detecting a large L-ton with $L \neq \mathcal{O}(1)$ and $\lim_{N\to\infty} L = \infty$ as a single-ton $\mathcal{E}_{k_\star|M_L}$.

Next we provide the highlights of the analysis to show that the probability of bin error of the peeling-decoder decodes go to zero as $N \to \infty$ at a rate that is at least $\mathcal{O}(1/N^2)$.

- *Vanishing zero-ton error:* The zero-ton error probability $\mathbb{P}(\mathcal{E}_{k_\star|z} \bigcup \mathcal{E}_{z|k_\star})$ is found by evaluating the tail of noise distribution based on the *zero-ton test*.
- *Vanishing large L-ton error:* The large multi-ton error probability $\mathbb{P}(\mathcal{E}_{k_\star|M_L} \bigcup \mathcal{E}_{M_L|k_\star})$ is found by applying the Central Limit Theorem (CLT) to the multi-ton with $L \neq \mathcal{O}(1)$ and $\lim_{N\to\infty} L = \infty$, and evaluating the tail probability with noise based on the *multi-ton test*.
- *Vanishing small L-ton error:* The error exponent of mistaking a small L-ton with the true single-ton $\mathbb{P}(\mathcal{E}_{k_\star|m_L} \bigcup \mathcal{E}_{m_L|k_\star})$ is driven by the minimum distance between any small L-ton and a single-ton, which can be obtained by a proper bound on the mutual coherence of the bin-observation matrix for any two randomly picked columns and applying the worst case mutual coherence bound on $L = \mathcal{O}(1)$ columns.
- *Vanishing single-ton error:* The error exponent of mistaking any single-ton with the true single-ton $\mathbb{P}(\mathcal{E}_{k|k_\star})$ is driven by the minimum distance between every distinct pair of single-tons, which can be obtained by using the mutual coherence on any two randomly picked columns of the bin-observation matrix.

In the following, we analytically study the probability of error in each category.

Proposition 2. *(Vanishing zero-ton error) The probabilities of mistaking a single-ton as a zero-ton (miss detection) as well as mistaking a zero-ton with a single-ton (false detection) can be upper bounded as*

$$\mathbb{P}\left(\mathcal{E}_{z|k_\star} \bigcup \mathcal{E}_{k_\star|z}\right) \leq 2\exp\left(-C_0 P\right)$$

where

$$C_0 = \frac{1}{2}\log\left[\frac{\rho^2}{(1+\gamma)\nu^2}\right] \tag{31}$$

for some γ.

Proof. See Appendix D. □

Proposition 3. *(Large Multi-ton v.s. Single-ton) The probabilities of mistaking a single-ton as a small multi-ton (miss detection) as well as mistaking a small multi-ton with a single-ton (false detection) can be bounded as*

$$\mathbb{P}\left(\mathcal{E}_{M_L|k_\star} \bigcup \mathcal{E}_{k_\star|M_L}\right) < 4\exp\left(-\frac{\gamma^2 P}{8}\right). \tag{32}$$

Proof. See Appendix E. □

Proposition 4. *(Small Multi-ton v.s. Single-ton) The pobabilities of mistaking a single-ton as a small multi-ton (miss detection) as well as mistaking a small multi-ton with a single-ton (false detection) can be bounded respectively as*

$$\mathbb{P}\left(\mathcal{E}_{m_L|k_\star}\right) \leq 2\exp\left(-\frac{\gamma^2 P}{8}\right) + \exp\left(-\frac{P\rho^2}{\nu^2}\right) \tag{33}$$

$$\mathbb{P}\left(\mathcal{E}_{k_\star|m_L}\right) \leq \exp\left(-C_L P\right), \tag{34}$$

where

$$C_L = \frac{1}{2}\log\left(\frac{(L-1)\rho^2}{2(1+\gamma)\nu^2}\right). \tag{35}$$

Proof. See Appendix F. □

Proposition 5. *(Single-ton v.s. Single-ton) The probability of mistaking a single-ton as another single-ton is bounded as*

$$\mathbb{P}(\mathcal{E}_{k_\star|k}) \leq \exp\left(-C_0 P\right). \tag{36}$$

Proof. See Appendix G. □

It can be inferred from Proposition 2 - 5 that the bin error can be bounded by

$$\mathbb{P}\left(\mathcal{E}_b\right) \leq \exp(-C_{\min} P), \tag{37}$$

where

$$C_{\min} \triangleq \min\left\{C_0, \frac{\gamma^2}{8}, C_L\right\}. \tag{38}$$

To drive the error to vanish at the rate of $1/N^2$, it is sufficient to have

$$P \geq \frac{2\log N}{C_{\min}} \tag{39}$$

and thus $P = \mathcal{O}(\log N)$.

Appendix B
Proof of Lemma 1

The mutual coherence μ is obtained as

$$\mu = \max_{n\neq k} \frac{1}{P}\left|\sum_{p=1}^{P}(-1)^{\langle d_{c,p},n\rangle}(-1)^{\langle d_{c,p},k\rangle}\right| \tag{40}$$

$$= \max_{n\neq k} \frac{1}{P}\left|\sum_{p=1}^{P}(-1)^{\langle d_{c,p},n+k\rangle}\right| \tag{41}$$

$$= \max_{n\neq k} \mu_{n,k} \tag{42}$$

where

$$\mu_{n,k} \triangleq \left|\sum_{p=1}^{P}(-1)^{\langle d_{c,p},n+k\rangle}/P\right|. \tag{43}$$

Under the assumption that the shifts $d_{c,p}$'s are selected randomly, each summand is an independent Bernoulli random variable taking values ± 1 equiprobably, which implies that the mean of each term is zero. Furthermore, since each random variable is bounded between $[-1, 1]$, the Hoeffding bound gives us

$$\mathbb{P}\left(\left|\sum_{p=1}^{P}(-1)^{\langle d_{c,p},n+k\rangle}/P\right| \geq \mu_0\right) \leq 2\exp\left(-\frac{P\mu_0^2}{2}\right). \tag{44}$$

Therefore, with probability at least $1 - 2\exp\left(-\frac{P\mu_0^2}{2}\right)$, the variable $\mu_{n,k}$ can be bounded by

$$\mu_{n,k} \leq \mu_0. \tag{45}$$

Therefore, the probability of the mutual coherence μ greather than μ_0 can be bounded by the union bound below (assuming a uniform distribution of n and k)

$$\mathbb{P}(\mu \geq \mu_0) \leq \frac{1}{N}\sum_{n=1}^{N}\sum_{k\neq n}\mathbb{P}(\mu_{n,k} \geq \mu_0) \tag{46}$$

$$\leq 2N\exp\left(-\frac{P\mu_0^2}{2}\right). \tag{47}$$

By choosing $\mu_0 = 1/2(L+1)$ for some positive integer $L > 0$, the mutual coherence μ can be bounded as

$$\mu \leq \frac{1}{2(L+1)} \tag{48}$$

with probability at least $1 - 2N\exp\left(-\frac{P}{8(L+1)^2}\right)$. As long as $P \geq 24(L+1)^2\log N$, the probability can be achieved with

$$1 - 2N\exp\left(-\frac{P}{8(L+1)^2}\right) \geq 1 - \frac{2}{N^2}. \tag{49}$$

Appendix C
Proof of General Tail Probability

We first invoke the following tail bound for later error analysis.

Lemma 2. **[16]** *Given a random Gaussian vector $\bar{\xi} \sim \mathcal{N}(0, \nu^2 I) \in \mathbb{R}^P$ and any length-P vector $\bar{\phi} \in \mathbb{R}^P$ such that*

$$\frac{1}{P}\left\|\bar{\phi}\right\|^2 \geq E_{\min}, \tag{50}$$

then the following tail bounds hold:

$$\mathbb{P}\left(\frac{1}{P}\left\|\bar{\phi}+\bar{\xi}\right\|^2 \leq \tau\right) \leq \exp\left(-C_\tau P\right) \tag{51}$$

where C_τ is some constant given by

$$C_\tau = \frac{1}{2}\log\left(\frac{E_{\min}}{\tau}\right). \tag{52}$$

Appendix D
Proof of Proposition 2

A. Detecting a Single-ton as a Zero-ton $\mathcal{E}_{z|k_\star}$

The event $\mathcal{E}_{z|k_\star}$ occurs under the single-ton model in (14) such that the bin energy obtained as

$$\frac{1}{P}\left\|\bar{U}_c[j]\right\|^2 = \frac{1}{P}\left\|X_{k_\star}\bar{g}_{k_\star} + \bar{\xi}_c[j]\right\|^2 \tag{53}$$

drops below the threshold $\tau = (1+\gamma)\nu^2$. By using $\phi = X_{k_\star}\bar{g}_{k_\star}$ and $\bar{\xi}_c[j] = \bar{\xi}$ in Lemma 2, we have $\left\|\bar{\phi}\right\| = \rho\sqrt{P}$ (i.e., $E_{\min} = \rho^2$) and thus

$$\mathbb{P}\left(\mathcal{E}_{z|k_\star}\right) \leq \exp\left(-C_0 P\right)$$

where C_0 is given in (31).

B. Detecting a Zero-ton as a Single-ton $\mathcal{E}_{k_\star|z}$

The event $\mathcal{E}_{k_\star|z}$ occurs under the zero-ton model in (13) such that the bin energy obtained as

$$\frac{1}{P}\left\|\bar{U}_c[j] - X_{k_\star}\bar{g}_{k_\star}\right\|^2 = \frac{1}{P}\left\|X_{k_\star}\bar{g}_{k_\star} - \bar{\xi}_c[j]\right\|^2 \tag{54}$$

rises above the threshold $\tau = (1+\gamma)\nu^2$.

$$\mathbb{P}\left(\mathcal{E}_{k_\star|z}\right) = \mathbb{P}\left(\frac{1}{P}\left\|X_{k_\star}\bar{g}_{k_\star} - \bar{\xi}_c[j]\right\|^2 \leq \tau\right).$$

Now we let $\bar{\phi} = X_{k_\star}\bar{g}_{k_\star}$ and $\bar{\xi} = -\bar{\xi}_c[j]$. This implies that $E_{\min} = \rho^2$ and therefore by Lemma 2, the probability of error for the event $\mathcal{E}_{k_\star|k}$ can be similarly obtained as

$$\mathbb{P}\left(\mathcal{E}_{k_\star|z}\right) \leq \exp\left(-C_0 P\right) \tag{55}$$

for some γ.

Appendix E
Proof of Proposition 3

Define the set of indices of non-zero coefficients in bin: $\mathcal{L}_{c,j} \subseteq \mathcal{K}_{c,j}$. Since the multi-ton model is a sum of different signatures and noise,

$$\bar{U}_c[j] = \sum_{k\in\mathcal{L}_{c,j}} X_k\bar{g}_k + \xi_c[j], \tag{56}$$

we specifically analyze the following sum in the asymptotic regime $L \to \infty$ as $N \to \infty$

$$\bar{Z} = \frac{1}{L} \sum_{k \in \mathcal{L}_{c,j}} X_k \bar{g}_k. \tag{57}$$

Since X_k and $\bar{g}_k$ are independent random variables, therefore $X_k \bar{g}_k$'s are independent identically distributed. Clearly, by the Central Limit Theorem (CLT), the vector $\bar{Z}$ asymptotically converges in distribution to a multi-variate normal random vector

$$\bar{Z} \sim \mathcal{N}(0_{P\times 1}, \Sigma), \tag{58}$$

where

$$\Sigma = \mathbb{E}\left[\bar{g}_k \bar{g}_k^T |X_k|^2\right] = \rho^2 \mathbb{E}\left[\bar{g}_k \bar{g}_k^T\right]. \tag{59}$$

The (i,j)-th element in Σ can be readily obtained as

$$\Sigma_{ij} = \rho^2 \mathbb{E}\left[(-1)^{\langle d_{c,i} + d_{c,j}, k\rangle}\right] = \begin{cases} \rho^2, & i = j \\ 0, & i \neq j \end{cases}. \tag{60}$$

Therefore, $\bar{Z}$ has P independent unit-variance normal random variables and therefore asymptotically the entries in $\bar{U}_c[j]$ are independent normal random variables each with variance $\sigma_L^2 = L\rho^2 + \nu^2$. Therefore, $\left\|\bar{U}_c[j]\right\|^2$ is a P-dimensional chi-square random variable $\sigma_L^2 \chi_P^2$.

A. Detecting a Single-ton as a Large L-ton $\mathcal{E}_{m_L|k_\star}$

The event $\mathcal{E}_{M_L|k_\star}$ occurs when the underlying bin is a single-ton bin, which is mistaken as a large multi-ton. Such event occurs under the single-ton model (14) whenever

$$\frac{1}{P}\left\|\bar{U}_c(j)\right\|^2 \geq \tau, \quad \tau = (1+\gamma)(L\rho^2 + \nu^2),$$

Substituting the single-ton model into the above expression, we have

$$\frac{1}{P}\left\|X_{k_\star}\bar{g}_{k_\star} + \bar{\xi}_c[j]\right\|^2 \geq (1+\gamma)(L\rho^2+\nu^2).$$

Using the triangular inequality, we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_{M_L|k_\star}\right) &\leq \mathbb{P}\left(\left\|\bar{\xi}_c[j]\right\|^2 \geq P(1+\gamma)(L\rho^2+\nu^2) - \rho^2\right) \\ &< \mathbb{P}\left(\left\|\bar{\xi}_c[j]\right\|^2 \geq P(1+\gamma)\nu^2\right).\end{aligned}$$

Since $\left\|\bar{\xi}_c[j]\right\|^2$ follows a chi-square distribution $\nu^2\chi_P^2$, and χ_P^2 is a sub-exponential random variable with parameters $(4P, 4)$, we obtain a standard tail bound for χ_P^2 for some real number $t \in (0, P)$ as follows

$$\mathbb{P}\left(\left|\chi_P^2 - P\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{8P}\right) \tag{61}$$

and therefore

$$\mathbb{P}\left(\chi_P^2 \geq t + P\right) \leq 2\exp\left(-\frac{t^2}{8P}\right). \tag{62}$$

Now let $P + t = P\tau/\nu^2 = P(1+\gamma)$ such that $t = \gamma P$, we have

$$\mathbb{P}\left(\left\|\bar{\xi}_c[j]\right\|^2 \geq P(1+\gamma)\nu^2\right) < 2\exp\left(-\frac{\gamma^2 P}{8}\right). \tag{63}$$

Therefore,

$$\mathbb{P}\left(\mathcal{E}_{M_L|k_\star}\right) < 2\exp\left(-\frac{\gamma^2 P}{8}\right). \tag{64}$$

B. Detecting a Large L-ton as a Single-ton $\mathcal{E}_{k_\star|M_L}$

The event $\mathcal{E}_{k_\star|M_L}$ corresponds to the error when the underlying bin is a large multi-ton bin of size $L \neq \mathcal{O}(1)$ with $\lim_{N\to\infty} L = \infty$ and is mistaken as a single-ton bin at location $k_\star$. The event $\mathcal{E}_{k_\star|M_L}$ occurs under the multi-ton model (15) whenever the energy

$$\frac{1}{P}\left\|\bar{U}_c[j] - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \leq (1+\gamma)\nu^2.$$

Thus, the error probability can be obtained by

$$\mathbb{P}(\mathcal{E}_{k_\star|M_L}) = \mathbb{P}\left(\frac{1}{P}\left\|\bar{U}_c[j] - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \leq (1+\gamma)\nu^2\right) \tag{65}$$

$$\leq \mathbb{P}\left(\frac{1}{P}\left\|\bar{U}_c[j]\right\|^2 \leq (1+\gamma)\nu^2 + \rho^2\right). \tag{66}$$

Since $(1+\gamma)\nu^2 + \rho^2 \leq (1-\gamma)(L\rho^2+\nu^2)$ as long as $0 < \gamma < 1$ in the asymptotic regime $L \to \infty$, we have

$$\mathbb{P}(\mathcal{E}_{k_\star|M_L}) \leq \mathbb{P}\left(\frac{1}{P}\left\|\bar{U}_c[j]\right\|^2 \leq (1-\gamma)(\nu^2 + L\rho^2)\right) \tag{67}$$

where $\bar{U}_c[j]$ follows the multi-ton model. Therefore, the probability in (67) can be obtained by

$$\mathbb{P}(\mathcal{E}_{k_\star|M_L}) \leq \mathbb{P}\left(\chi_P^2 \leq \frac{P(1-\gamma)(\nu^2+L\rho^2)}{\sigma_L^2}\right) \tag{68}$$

$$= \mathbb{P}\left(\chi_P^2 \leq P(1-\gamma)\right). \tag{69}$$

This can be bounded by the same bound in (61) as

$$\mathbb{P}(\mathcal{E}_{k_\star|M_L}) \leq 2\exp\left(-\frac{\gamma^2 P}{8}\right). \tag{70}$$

Appendix F
Proof of Proposition 4

A. Detecting a Single-ton as a Small L-ton $\mathcal{E}_{m_L|k_\star}$

The event $\mathcal{E}_{m_L|k_\star}$ corresponds to the error when the underlying bin is a single-ton bin and is mistaken as a small multi-ton. Such event occurs under the single-ton model (14) whenever

$$\frac{1}{P}\left\|\bar{U}_c(j) - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \geq (1+\gamma)\nu^2,$$

or

$$\widehat{X}_{k_\star} \neq X_{k_\star},$$

which gives

$$\begin{aligned}&\mathbb{P}\left(\mathcal{E}_{m_L|k_\star}\right) \\ &\leq \mathbb{P}\left(\frac{1}{P}\left\|\bar{U}_c(j) - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \geq (1+\gamma)\nu^2 | \widehat{X}_{k_\star} = X_{k_\star}\right) \\ &+ \mathbb{P}\left(\widehat{X}_{k_\star} \neq X_{k_\star}\right).\end{aligned}$$

Substituting the single-ton model into the above expression, the first term is equivalent to

$$\mathbb{P}\left(\frac{1}{P}\left\|\bar{U}_c(j) - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \geq (1+\gamma)\nu^2 | \widehat{X}_{k_\star} = X_{k_\star}\right) = \mathbb{P}\left(\left\|\bar{\xi}_c[j]\right\|^2 \geq P(1+\gamma)\nu^2\right).$$

The probability of this event can be bounded as that in (63)

$$\mathbb{P}\left(\left\|\bar{\xi}_c[j]\right\|^2 \geq P(1+\gamma)\nu^2\right) < 2\exp\left(-\frac{\gamma^2 P}{8}\right). \quad (71)$$

On the other hand, the value of $\mathbb{P}\left(\widehat{X}_{k_\star} \neq X_{k_\star}\right)$ can be bounded by the error probability of binary detections as

$$\mathbb{P}\left(\widehat{X}_{k_\star} \neq X_{k_\star}\right) \leq \exp\left(-\frac{P\rho^2}{\nu^2}\right). \quad (72)$$

The result in Proposition 4 thus follows by summing up (71) and (72).

B. Detecting a Small L-ton as a Single-ton $\mathcal{E}_{k_\star|m_L}$

The event $\mathcal{E}_{k_\star|m_L}$ is a generalized event that corresponds to the error when the underlying bin is a small multi-ton bin of size $L = \mathcal{O}(1)$ and is mistaken as a single-ton bin at location $k_\star$. The event $\mathcal{E}_{k_\star|m_L}$ occurs under the multi-ton model (15) whenever the energy

$$\frac{1}{P}\left\|\bar{U}_c[j] - X_{k_\star}\bar{g}_{k_\star}\right\|^2 \leq \tau, \quad \tau = (1+\gamma)\nu^2.$$

The probability of this event can be upper bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{k_\star|m_L}) &= \mathbb{P}\left(\frac{1}{P}\left\|\sum_{k\in\mathcal{L}_{c,j}} X_k\bar{g}_k - X_{k_\star}\bar{g}_{k_\star} + \bar{\xi}_c[j]\right\|^2 \leq \tau\right)\\ &= \mathbb{P}\left(\frac{1}{P}\left\|G_c[j]\bar{q} + \bar{\xi}_c[j]\right\|^2 \leq \tau\right),\end{aligned}$$

where $\bar{q} = \sum_{k\in\mathcal{L}} Q_k\bar{g}_k$ is an $(L+1)$-sparse vector with $\mathcal{L} = \mathcal{L}_{c,j}\bigcup\{k_\star\}$ and $G_c[j]$ is the bin-observation matrix in (6).

Now let $\bar{\phi} = G_c[j]\bar{q} = \sum_{k\in\mathcal{L}} Q_k\bar{g}_k$ and $\bar{\xi} = \bar{\xi}_c[j]$. Since the support of $\bar{q}$ is no greater than $(L+1)$, the norm $\|\bar{\phi}\|^2$ can be obtained as follows:

$$\|\bar{\phi}\|^2 = \left\|\sum_{k\in\mathcal{L}} Q_k\bar{g}_k\right\|^2 \quad (73)$$

$$= \sum_{k\in\mathcal{L}} |Q_k|^2\|\bar{g}_k\|^2 + \sum_{n\neq k}\sum_k \bar{g}_n^T\bar{g}_k Q_k Q_n \quad (74)$$

$$\geq P(L-1)\rho^2 - P\rho^2\sum_{n\neq k}\sum_k \mu_{n,k} \quad (75)$$

$$\geq P(L-1)\rho^2 - P\rho^2(L^2+1)\mu, \quad (76)$$

where μ is the mutual coherence. Therefore, the norm $\|\bar{\phi}\|^2$ can be lower bounded by

$$\|\bar{\phi}\|^2 > P(L-1)\rho^2/2,$$

with probability at least $1-\mathcal{O}(N^{-2})$ if $P \geq 24(L+1)^2\log N$ according to Lemma 1. This implies that $E_{\min} = (L-1)\rho^2/2$ and therefore by Lemma 2, the probability of error for the event $\mathcal{E}_{k_\star|m_L}$ can be computed by

$$\mathbb{P}(\mathcal{E}_{k_\star|m_L}) \leq \exp\left(-C_L P\right) \quad (77)$$

where C_L is given in (35).

Appendix G
Proof of Proposition 5

The event $\mathcal{E}_{k|k_\star}$ occurs under the single-ton model (14) whenever the energy

$$\left\|\bar{U}_c[j] - X_k\bar{g}_k\right\|^2 \leq (1+\gamma)\tau, \quad \tau = P\nu^2.$$

The probability of this event can be upper bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{k_\star|k}) &= \mathbb{P}\left(\left\|X_{k_\star}\bar{g}_{k_\star} - X_k\bar{g}_k + \bar{\xi}_c[j]\right\|^2 \leq (1+\gamma)\tau\right)\\ &= \mathbb{P}\left(\left\|G_c[j]\bar{q} + \bar{\xi}_c[j]\right\|^2 \leq (1+\gamma)\tau\right),\end{aligned}$$

where $\bar{q}$ is an 2-sparse vector and $G_c[j]$ is the bin-observation matrix in (6). Now let $\bar{\phi} = G_c[j]\bar{q}$ and $\bar{\xi} = \bar{\xi}_c[j]$. Since the support of $\bar{q}$ is exactly 2, the norm $\|\bar{\phi}\|^2$ can be obtained as follows:

$$\|\bar{\phi}\|^2 \geq P\rho^2, \quad (78)$$

with probability at least $1-\mathcal{O}(N^{-2})$ if $P \geq 24\log N$ according to Lemma 1. This implies that $E_{\min} = \rho^2$ and therefore by Lemma 2, the probability of error for the event $\mathcal{E}_{k_\star|k}$ can be computed by

$$\mathbb{P}(\mathcal{E}_{k_\star|k}) \leq \exp\left(-C_0 P\right). \quad (79)$$

References

[1] W. Pratt, J. Kane, and H. C. Andrews, "Hadamard transform image coding," *Proceedings of the IEEE*, vol. 57, no. 1, pp. 58–68, 1969.

[2] T. R. WGI, "Spreading and modulation (fdd)," 3GPP Tech Rep. TS25. 213, 2000. http://www. 34gpp. org, Tech. Rep.

[3] S. Haghighatshoar and E. Abbe, "Polarization of the rényi information dimension for single and multi terminal analog compression," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 779–783.

[4] M. Lee and M. Kaveh, "Fast hadamard transform based on a simple matrix factorization," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 6, pp. 1666–1667, Dec 1986.

[5] J. Johnson and M. Puschel, "In search of the optimal walsh-hadamard transform," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol. 6, 2000, pp. 3347–3350 vol.6.

[6] K. J. Horadam, *Hadamard matrices and their applications*. Princeton university press, 2007.

[7] A. Hedayat and W. Wallis, "Hadamard matrices and their applications," *The Annals of Statistics*, vol. 6, no. 6, pp. 1184–1238, 1978.

[8] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Simple and practical algorithm for sparse fourier transform," in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2012, pp. 1183–1194.

[9] ——, "Nearly optimal sparse fourier transform," in *Proceedings of the 44th symposium on Theory of Computing*. ACM, 2012, pp. 563–578.

[10] 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.

[11] M. Iwen, A. Gilbert, and M. Strauss, "Empirical evaluation of a sub-linear time sparse dft algorithm," *Communications in Mathematical Sciences*, vol. 5, no. 4, pp. 981–998, 2007.

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

[13] S. Pawar and K. Ramchandran, "Computing a k-sparse n-length discrete fourier transform using at most $4k$ samples and $\mathcal{O}(klogk)$ complexity," *arXiv preprint arXiv:1305.0870*, 2013.

[14] R. Scheibler, S. Haghighatshoar, and M. Vetterli, "A fast hadamard transform for signals with sub-linear sparsity," *arXiv preprint arXiv:1310.1803*, 2013.

[15] S. A. Pawar, "Pulse: Peeling-based ultra-low complexity algorithms for sparse signal estimation," *PhD Dissertation*, 2013.

[16] Y. Jin, Y.-H. Kim, and B. D. Rao, "Limits on support recovery of sparse signals via multiple-access communication techniques," *Information Theory, IEEE Transactions on*, vol. 57, no. 12, pp. 7877–7892, 2011.