Change Detection with Sparse Signals using Quantum Designs

Aditi Jain *, Pradeep Sarvepalli †, Srikrishna Bhaskyam ‡, Arun Pachai Kannu §
Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai - 600036, India
Email: *ee15s081@ee.iitm.ac.in, †pradeep@ee.iitm.ac.in, ‡skrishna@ee.iitm.ac.in, §arunpachai@ee.iitm.ac.in

Abstract—We consider the change detection problem where
the pre-change observation vectors are purely noise and the post-
change observation vectors are noise-corrupted compressive mea-
surements of sparse signals with a common support, measured
using a sensing matrix. In general, post-change distribution of
the observations depends on parameters such as the support
and variances of the sparse signal. When these parameters are
unknown, we propose two approaches. In the first approach, we
approximate the post-change pdf based on the known parameters
such as mutual coherence of the sensing matrix and bounds on
the signal variances. In the second approach, we parameterize
the post-change pdf with an unknown parameter and try to
adaptively estimate this parameter using a stochastic gradient
descent method. In both these approaches, we employ CUSUM
algorithm with various decision statistics such as the energy of
the observations, correlation values with columns of the sensing
matrix and the maximum value of such correlations. We study
the performance of these approaches and offer insights on the
relevance of different decision statistics in different SNR regimes.
We also address the problem of designing sensing matrices with
small coherence by using designs from quantum information
theory. One such design, called SIC POVM, also has an additional
structure which allows exact computation of the post-change pdfs
of some decision statistics even when the support set of the sparse
signal is unknown. We apply our detection algorithms with SIC
POVM based sequences to a massive random access problem and
show their superior performance over conventional Gold codes.

Index Terms—CUSUM algorithm, detection delay, average
run length, sensing matrix design, mutual coherence, quantum
information theory

I. INTRODUCTION

The problem of change detection using statistical tests
has been studied over several decades [1]–[3]. The simplest
model for change detection problems is described below. The
observation at time \( t \) is denoted as \( y[t] \). Let \( \nu \in \mathbb{N}^+ \) denote
the change point such that the observations before and after
change follow different statistics. Specifically, the observations
\( \{y[t], \ t \geq 0\} \) are independent and follow the statistics,

\[
y[t] \sim \begin{cases} \begin{align*}
f_0 & \text{ if } 0 \leq t < \nu, \\ f_1 & \text{ if } t \geq \nu,
\end{align*} \end{cases}
\]

where \( f_0 \) and \( f_1 \) are the pre-change and post-change probability
density functions (pdf) respectively. When the change point
\( \nu \) is unknown and non-random, the quantities of interest are
the average run length \( T_\nu \) and the worst-case detection delay
\( D_w \). These quantities are mathematically defined below. We
use \( E_{\nu} \) to denote the expectation with respect to the probability
measure on the observations when the change point is \( \nu \). We
set \( \nu = \infty \) when there is no change. With \( T \) being the time at
which a given algorithm declares change (which is random),
the average run length and the worst case detection delay of
the algorithm is given as

\[
T_\nu = E_{\infty}\{T\}, \tag{2}
\]

\[
D_w = \sup_{\nu \geq 1} E_{\nu}\{(T - \nu) | T \geq \nu\}. \tag{3}
\]

The CUSUM algorithm [4] for change detection uses the log
likelihood ratio (LLR) for each observation, which is given as
\( L(y[t]) = \log \frac{f_1(y[t])}{f_0(y[t])} \). The CUSUM metric \( W[t] \) is initialized
to \( W[-1] = 0 \) and is recursively computed as

\[
W[t] = (W[t-1] + L(y[t]))^+, \tag{4}
\]

where \((\cdot)^+\) denotes \( \max\{\cdot, 0\} \). The CUSUM decision rule \( \mathcal{R} \)
using the metric \( W[t] \), with threshold \( \tau \in (0, \infty) \), is given as

\[
\mathcal{R} = \begin{cases} \text{Declare change at time } t & \text{if } W[t] > \tau, \\ \text{Continue otherwise.} \end{cases}
\]

The threshold parameter \( \tau \) in the above rule controls the
average run length and the detection delay. It is shown in [1]
that CUSUM algorithm asymptotically minimizes the worst
case detection delay, subject to a constraint on the average
run length \( T_\nu \geq \gamma \), as the threshold \( \tau \to \infty \) (or equivalently
as \( \gamma \to \infty \)).

Several variations of the model in [1] have been addressed
in the literature, considering cases where the pre-change [5]
or post-change distributions [6] have unknown parameters.
Adaptive algorithms to estimate the unknown parameters in
the post-change distributions have been developed in [7], [8].

In natural and practical scenarios, most signals have sparse
representations in an appropriately chosen basis. Compressive
sensing deals with the problem of reconstructing sparse signals
from under-determined linear measurements [9]. Orthogonal
matching pursuit (OMP) is a popular sparse signal reconstruc-
tion technique [10] which works based on the correlation of
the observation with columns of the sensing matrix. Detecting
sparse signals in the presence of noise has been addressed in
several papers such as [11]–[15], where detection is performed
based on various statistics such as energy, correlation values
and partially recovered support. In [16], [17], the authors have
developed a sequential approach based on LLR to detect sparse
signals in the presence of noise.
In this paper, we consider the change detection problem wherein the pre-change observation vectors are purely noise and the post-change observation vectors are noise-corrupted compressive measurements of sparse signals with a common support, measured using a sensing matrix. When the support and the variances of the non-zero entries of a sparse signal are unknown, the post-change distributions of the observations (and other decision statistics) are not known perfectly. Change detection with sparse signals have been previously addressed in [18], [19]. While [18] addresses the problem where the sparsifying dictionary of the signal is unknown, [19] addresses the case where the observation has the same dimension as that of the sparse signal. In our work, the sparsifying dictionary is assumed to be known. However, we allow the dimension of the observation to be much smaller than the dimension of the sparse signal. We develop change detection algorithms using various decision statistics and show their relevance in regimes with different signal to noise ratio (SNR). We also design sensing matrices using constructions from quantum information theory and show that they perform better than random constructions. More details on our system model and contributions are discussed in the following section.

II. SYSTEM MODEL

A. Sparse Signal Model

For the change detection problem with sparse signals, we consider the vector observation model,

\[ y[t] = \begin{cases} n[t] & 0 \leq t < \nu, \\ Ax[t] + n[t] & t \geq \nu. \end{cases} \quad (5) \]

Here, \( n[t] \in \mathbb{C}^{M \times 1} \) denotes the complex additive white Gaussian noise (AWGN) with pdf \( \text{CN}(0, \sigma_n^2 I) \). \( A \in \mathbb{C}^{M \times N} \) denotes the sensing matrix (with \( M \leq N \)) and \( x[t] \) denotes the sparse signal with the number of non-zero entries \( \|x[t]\|_0 = K \ll N \). We refer to \( K \) as the sparsity level of the signal \( x[t] \). We consider the case where the support (i.e. locations of non-zero entries) of \( x[t] \) remains the same for all \( t \geq \nu \). Let \( S \) denote the ordered support set, containing the locations of non-zero entries of \( x[t] \). Note that \( x_S[t] \) is of size \( K \) and contains the non-zero entries of \( x[t] \). After the change, \( y[t] \) can be restated as

\[ y[t] = A_S x_S[t] + n[t] = \sum_{i \in S} a_i x_i[t] + n[t], \quad \forall t \geq \nu. \quad (6) \]

1 Notation: Scalars are denoted by lowercase letters. Matrices (vectors) are denoted by uppercase (lowercase) boldface letters. The \( i \)-th column (entry) of \( A \) (\( x(t) \)) is denoted by \( a_i(x_i[t]) \). The entry in \( i \)-th row and \( f \)-th column of \( A \) is denoted by \( a_{if} \). We denote transpose by \( (\cdot)^T \), conjugate transpose by \( (\cdot)^* \), inverse by \( (\cdot)^{-1} \), trace by \( tr(\cdot) \), \( \ell_p \) norm by \( \| \cdot \|_p \). Calligraphic letters denote sets like \( S \). We use \( A_S \) (\( x_S \)) to denote the sub-vector (sub-matrix) of \( A \) (\( x \)) consisting of columns (entries) whose index belongs to \( S \). \( \| \cdot \|_\infty \) denotes the absolute value of a scalar, as well as the cardinality of a set, which will be apparent from the context, \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) denote the floor and ceiling values of their arguments. We use \( \hat{\theta} \) to denote an approximation of \( \theta \). We denote a zero vector of dimension \( N \) by \( 0_N \), identity matrix by \( I \) and \( \sqrt{\cdot} \) by \( J \). We use \( \text{diag}(\cdot) \) to denote a diagonal matrix with elements of vector \( d \) as its diagonal entries. For observations \( d[t] \), we denote pre-change by \( f_0^P \) and post-change pdf by \( f_0^P \). \( f_0^P \) denotes the approximation of \( f_P^0 \) and \( f_P^1 \) denotes the post-change pdf of \( d[t] \) parametrized by \( \theta \). \( N(\mu, \sigma^2) \) denotes Gaussian and \( \text{CN}(\mu, \sigma^2) \) denotes complex Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \). \( \exp(\lambda) \) denotes exponential distribution with rate parameter \( \lambda \). \( \chi^2_n \) denotes central and \( \chi^2_n(\mu) \) denotes non central chi squared distribution with \( k \) degrees of freedom and non centraality parameter \( \mu \).

We are interested in detecting the change and finding the support of the sparse signal \( x[t] \), once the change is declared. To proceed further, we assume that the pdf of \( x_S[t] \) is \( \text{CN}(0, C_x) \) and the signal covariance matrix \( C_x = \text{diag}(\sigma_1^2, \ldots, \sigma_K^2) \) is diagonal. We also assume that the non-zero entries of \( x[t] \) are independent across time \( t \). For the model in [5], we assume that the noise variance \( \sigma_n^2 \) is known. Mutual coherence of the sensing matrix, defined as

\[ \alpha = \max_{1 \leq k, p \leq N} \frac{\langle a_k, a_p \rangle}{\|a_k\| \|a_p\|} \quad (7) \]

plays an important role in the performance of sparse signal recovery algorithms [10]. In general, smaller the value of \( \alpha \), better will be the sparse signal recovery performance.

1) Change Detection Algorithms: For the special case when the sensing matrix \( A = I \) in [5], the change detection problem is addressed in [19]. On the other hand, we consider the general case, which allows compressive measurements \( M \ll N \) on the sparse signal. The change detection algorithms and their performance vary greatly depending on whether the three parameters, namely, support set \( S \), sparsity level \( K \) and signal covariance \( C_x \), are known or unknown. We consider all the combinations regarding the knowledge of these three parameters and develop corresponding change detection algorithms. When the signal variance is unknown, we assume that the lower \( \sigma_{min}^2 \) and the upper \( \sigma_{max}^2 \) bounds are available such that \( \sigma_{min}^2 \leq \sigma_i^2 \leq \sigma_{max}^2 \), for \( 1 \leq i \leq K \). For change detection, we use CUSUM algorithm with different decision statistics such as energy \( \|y[t]\|_2^2 \), correlation values \( g[t] = A^*y[t] \) and the maximum of these correlations \( \|g[t]\|_\infty \). When any or all of the three parameters \( S, K \) and \( C_x \) are unknown, we have the following two approaches:

- In the pdf-approximation based approach, we approximate the post-change pdf based on the known parameters such as \( \alpha, \sigma_{min}^2 \) and \( \sigma_{max}^2 \) and use this approximate pdf for LLR computations. Here, we use the philosophy that the change detection will be most difficult when the post-change pdf is closest (in terms of Kullback-Leibler (KL) distance) to the pre-change pdf and hence try to obtain the worst-case post-change pdf. We also ensure that some of our post-change pdf approximations are exact when the sensing matrix has some additional structure (such as when it is unitary).

- In the parameter-estimation based approach, we parameterize the post-change pdf with an unknown parameter and try to adaptively estimate this parameter using a stochastic gradient descent method [8], [20]. LLR computations are done using the estimated parameter value in the parameterized post-change pdf.

The various change detection algorithms are presented in Section III.

2) Sensing Matrix Design: From the theory of compressive sensing [9], sensing matrices with small coherence using designs from quantum information theory [7]. Specifically, we use symmetric informationally complete positive operator valued measure (SIC POVM) from the theory of equi-angular
B. Applications of the Model

In this section, we discuss some of the applications of our change detection model in [5].

1) Random Access in Direct Sequence-Code Division Multiple Access (DS-CDMA): Consider the synchronous DS-CDMA system with a codebook \( \{b_i : 1 \leq i \leq N\} \), where the codes \( b_i \) are \( M \times 1 \) vectors. Suppose \( Q \) users indexed by a known set \( Q \) are currently active and at time \( t = \nu \), \( K \) new users indexed by an unknown set \( S \) become active. The corresponding observation model is,

\[
r[t] = \begin{cases} 
B_Qx_Q[t] + n[t] & 0 \leq t < \nu, \\
B_Qx_Q[t] + B_S x_S[t] + n[t] & t \geq \nu.
\end{cases}
\]

Each entry in \( x_Q \) (and \( x_S \)) is the product of the (flat) fading channel gain and the constellation symbol sent by the corresponding user in the index sets \( Q \) and \( S \). The goal is to detect the change and identify the new users entering into the system. The authors in [26] consider the above model in [3] for the special case of \( K = 1 \). We allow \( K > 1 \), that is, more than one user can enter the system at a given time \( \nu \).

If the information related to already active users \( x_Q[t] \) is known (from detection/estimation), then it can be simply subtracted out from the received signal as \( y[t] = r[t] - B_Qx_Q[t] \), resulting in the model [5]. On the other hand, if \( x_Q[t] \) is not perfectly known, we can project the observation \( r[t] \) onto the orthogonal complement of \( B_Q \) as \( y[t] = P_Q^\perp r[t] \) with \( P_Q^\perp = I - B_Q(B_Q^* B_Q)^{-1} B_Q^* \). With this projection, the effective sensing matrix becomes, \( A_S = P_Q^\perp B_S \), resulting in the model specified in [5].

2) Localized Change Detection in Sensor Networks: Consider a wireless sensor network which has \( N \) sensor nodes and a fusion center. To convey their identity to the fusion center and enable transmission at the same time, each of the \( N \) sensors is assigned a unique \( M \)-length identification code, \( \{a_i : i = 1, \ldots, N\} \). All the sensors are initially in the OFF state, so that the observation at the fusion center is purely noise. When a change \( / \) event occurs at time \( t = \nu \), a subset of sensors \( S \) get affected by that event and enter into the ON state. The sensors in the ON state send their information symbol multiplexed with their corresponding code. The observation at the fusion center after the change is \( y[t] = \sum_{i \in S} a_i x_i[t] + n[t] \), with \( x_i[t] \) being the product of the channel between the \( i \)-th sensor to the fusion center and the information symbol sent by that sensor at time \( t \). This resembles the model in [3]. Here, recovering the support set \( S \) reveals the identities of the affected sensors, which in turn can reveal information on the location of the event in the network.

3) User Activity Detection in Massive Random Access: Massive random access systems [27] with applications in Internet of things (IoT), consist of a single receiving station and \( N \) number of users, with \( N \) being quite large. Each user is assigned an \( M \)-length identification code \( \{a_i : 1 \leq i \leq N\} \), which is known to the receiver. Initially, there is no active transmission. At some point in time, a small group of users indexed by the set \( S \) become active and send their transmission using the codes assigned to them, resulting in an observation model specified in [6].

C. Main Contributions

Some of the main contributions of our work are:

1) We develop change detection algorithms using the compressive measurements on the sparse signal and compare their performance in terms of worst-case detection delay versus average run length.

2) We develop an aggregate CUSUM algorithm using the entire correlation vector \( A^* y[t] \), which performs better than energy \( \|y[t]\|_2^2 \) based detection and maximum correlation \( \|A^* y[t]\|_\infty^2 \) based detection, in most of the scenarios.

3) We show that energy \( \|y[t]\|_2^2 \) based detection works better than maximum correlation \( \|A^* y[t]\|_\infty^2 \) based detection in the low SNR regime. On the other hand, when SNR is high, we show that correlation based detection performs better than energy based detection.

4) We show that quantum information theory based (deterministic) sensing matrices perform better than randomly generated matrices with i.i.d. Gaussian or Bernoulli distributed entries. Among the deterministic matrices, we show that SIC POVM yields the best detection performance when compared to MUB and approximate MUB based constructions.

5) We consider an application of our algorithms in massive random access and show that SIC POVM based codes have better detection performance when compared to Gold codes.

III. CHANGE DETECTION ALGORITHMS

In this section, we develop change detection algorithms for the model considered in [5]. The statistics of the post-change observations \( \{y[t], t \geq \nu\} \) depend on the parameters such as support set \( S \), sparsity level \( K \) and the signal covariance matrix \( C_x \). The change detection mechanism depends on whether these parameters are known or unknown. We address all the possible cases in this section.

A. Both Support and Signal Variance Known

For this case, both support set \( S \) and \( C_x \) are assumed to be known. We always have the pre-change pdf of \( y[t] \) as \( f_0^Y = CN(0, \sigma_n^2 I) \). When the support is known, the pdf of \( y[t] \) after the change is also perfectly known,

\[
f_1^Y = CN(0, A_S C_x A_S^* + \sigma_n^2 I). \tag{9}
\]

This resembles the standard change detection problem, for which the CUSUM algorithm is asymptotically optimal. It
is described below for completeness. Computing the LLR as 
\[ L^Y(y[t]) = \log \frac{f^G_y(y[t])}{f_0^G(y[t])}, \]
the CUSUM metric at each time \( t \) is

\[ W^Y[t] = (W^Y[t-1] + L^Y(y[t]))^+, \]

with the initialization \( W^Y[-1] = 0 \). We use the metric \( W^Y[t] \) to
detect change based on the rule specified in (4). We refer
to this method as Ideal-CUSUM since the support is known
perfectly in advance.

In some situations, we find it useful to implement the change
detection algorithms using the vector of inner products \( g[t] \), which is defined as

\[ g[t] \triangleq A^*y[t]. \]

We note that, when \( A \) is full rank, \( g[t] \) serves as a sufficient
statistic for detection since \( y[t] \) can be obtained from \( g[t] \) as
\( y[t] = (A A^*)^{-1}Ag[t]. \) The pre-change and post-change pdf
of \( g[t] \) is given as

\[ f^G_0 = CN(0,\sigma_n^2 A^*A), \]
\[ f^G_t = CN(0, A^*A_S C_xA_S A + \sigma_n^2 A^*A). \]

We compute the LLR as \( L^G(g[t]) = \log \frac{f^G_t(g[t])}{f^G_0(g[t])} \) and
the CUSUM metric as \( W^G[t] = (W^G[t-1] + L^G(g[t]))^+. \) We can
implement the CUSUM rule as given in (4).

Consider the special case when sensing matrix \( A \) is unitary.
In this case,

\[ g[t] = \begin{cases} \tilde{n}[t] & 0 \leq t < \nu, \\ x[t] + \tilde{n}[t] & t \geq \nu, \end{cases} \]

where \( \tilde{n}[t] = A^*n[t] \) and \( n[t] \sim CN(0,\sigma_n^2 I). \) Notice that
whenever the index \( i \notin S \), the \( i^{th} \) entry \( g_i[t] \) in \( g[t] \), is purely
noise before and after the change point. Thus, we have \( f^G_0 = CN(0,\sigma_n^2) \) and \( f^G_t = CN(0,\sigma_n^2 + \sigma_x^2) \), whenever \( i \in S \).
Here, \( \sigma_x^2 \) denotes the variance of the \( i^{th} \) entry in \( x_S[t] \). Hence, the
LLR \( L^G[g[t]] \) simplifies to,

\[ L^G(g[t]) = \sum_{i \in S} \log \frac{f^G_t(g_i[t])}{f^G_0(g_i[t])}. \]

Hence, for a unitary sensing matrix, the LLR computation using
\( g[t] \) boils down to summing the individual LLRs of only
those entries in \( g[t] \) which belong to the support set \( S \). This
observation will be useful in designing algorithms when the
support set is unknown.

B. Both Signal Variance and Sparsity Level Known, Support
Unknown

For the model in (5), we consider the case where the support
set of \( x[t] \) is unknown. However, we assume that the sparsity
level \( K \) and the signal covariance matrix \( C_x \) are known.
In addition, we assume that the covariance matrix is of the
form \( C_x = \sigma_n^2 I \), i.e., all the non-zero entries are i.i.d. Under
these assumptions, we develop the (asymptotically) optimal
CUSUM and other sub-optimal detection techniques in this
section.

1) Optimal CUSUM: When the support \( S \) of \( x[t] \) is unknown,
the covariance matrix of the post-change observations in (9)
is unknown. Thus, we can treat the support \( S \) as an
unknown parameter in the post-change pdf. Since there are
only finite number of possibilities for the support, which is \( N^K \), we run the CUSUM algorithm simultaneously for each
possible support. The exact details of the algorithm are given
below.

Let \( \mathcal{S} \) denote the set of all the subsets of \( \{1,2,\cdots,N\} \)
which have cardinality equal to \( K \). Thus, the true support \( S \)
is also one of the entries in \( \mathcal{S} \). Consider a candidate entry \( \hat{S} \) in \( \mathcal{S} \). For the candidate support set \( \hat{S} \), the associated post-change
pdf of the observation is \( f^Y_{\hat{S}} = CN(0,\sigma_n^2 A_{\hat{S}}A_{\hat{S}}^* + \sigma_x^2 I). \) Thus, we compute the CUSUM for each of the candidate support
set as

\[ L^Y_{\hat{S}}(y[t]) = \log \frac{f^Y_{\hat{S}}(y[t])}{f^G_0(y[t])}. \]

It can be easily verified that, for \( t \geq \nu \), we have
\( E_{\nu}(L^Y_{\hat{S}}(y[t])) > E_{\nu}(L^Y_S(y[t])) \), for any \( \hat{S} \in \mathcal{S} \setminus S \). Hence, after the change, the expected value of LLR will be
the highest when the candidate support set \( \hat{S} \) is identical to
the true support \( S \). Since, we do not know the true support,
we run CUSUM for each of the candidate support set and
make a decision based on the CUSUM metric which has the
largest magnitude. Specifically, for each \( \hat{S} \in \mathcal{S} \), we compute the CUSUM metric as

\[ W^Y_{\hat{S}}[t] = \left(W^Y_{\hat{S}}[t-1] + L^Y_{\hat{S}}(y[t])\right)^+, \]

with the initialization \( W^Y_{\hat{S}}[-1] = 0 \). The CUSUM change
detection rule \( R \) is given as

\[ R = \begin{cases} \text{Declare change at time } t \text{ if } \max_{\hat{S} \in \mathcal{S}} W^Y_{\hat{S}}[t] > \tau, \\ \text{Continue otherwise.} \end{cases} \]

In (13), Lorden also considered the case when the post-change
pdf could be any one from a finite set of pdfs. He established
that running CUSUM separately for each possible post-change
pdf and making the decision based on the maximum of these
CUSUM metrics, as done in (13), is asymptotically optimal,
as the average run length constraint approaches infinity. Though
optimal, for large values of \( N \), running \( \binom{N}{K} \) separate CUSUMs
can be prohibitively complex.

We now develop some sub-optimal detection techniques,
which are described below.

2) Aggregate CUSUM: We develop an algorithm, which we refer
to as Aggregate CUSUM, which uses the vector of inner-
products \( g[t] \), defined in (10). To get some insight, we start by
considering the case when \( A \) is unitary. In that case, \( g[t] \) is
given by (11) and the LLR computation (12) in Ideal-CUSUM
is equivalent to summing the LLRs of individual entries of
\( g[t] \) corresponding to the non-zero locations. Since we do not
know the support, we compute LLR for each entry \( g_i[t] \) in \( g[t] \),
assuming that \( i \) belongs to the support \( S \). Specifically, with
\( f^G_0 = CN(0,\sigma_n^2) \) and \( f^G_t = CN(0,\sigma_n^2 + \sigma_x^2) \), we compute the
LLR for \( i^{th} \) entry as

\[ L^G_i(g_i[t]) = \log \frac{f^G_t(g_i[t])}{f^G_0(g_i[t])}. \]
We compute CUSUM metric for each entry $g_i[t]$ parallely as
\[ W^{G_i}[t] = (W^{G_i}[t-1] + L^{G_i}(g_i[t]))^+, \]
with $W^{G_i}[0] = 0$. Again, we can easily verify that, after the change $(t \geq \nu)$, for any $i \in S$ and any $\ell \notin S$, we have $E((L^{G_i}[t]) > 0$ and $E((L^{G_i}[t]) < 0$. Hence, after the change, LLR in (14) tends to be larger when the entry belongs to the true support. This implies that the CUSUM metrics corresponding to the non-zero locations tend to be higher after the change. This indirectly provides a way of identifying the unknown support $S$ of $x[t]$. In order to detect the change, we sum the $K$-largest CUSUM metrics at each time $t$ and compare it with a threshold. We declare change at time $t$ based on the following decision rule,
\[
\mathcal{R} = \begin{cases} 
\text{Declare change} & \text{if } \sum_{i=0}^{K-1} W^{G_i}(N-1[t] > \tau, \quad (15) 
\text{Continue otherwise,} & \end{cases}
\]
where $W^{G_i}[m][t]$ denotes the $m^{th}$ ordered statistic of the set \{ $W^{G_i}[t] : 1 \leq i \leq N$ \}, with $W^{G_i}[N][t]$ being the largest. Using Aggregate CUSUM, we can also estimate the support by picking those $K$ locations that correspond to the $K$-largest values of $W^{G_i}[T]$ where $T$ is the time at which change is declared. When $A = I$, the Aggregate CUSUM algorithm described above has been studied in [19] and it was shown to have asymptotic optimality properties under some specific conditions.

We extend the Aggregate CUSUM algorithm for a non-unitary sensing matrix $A$ as follows. In this case, the the post-change pdf for each entry $j^{G_i}_1$ is not known perfectly and hence we use approximations for the post-change pdf. We need these approximations to be near-exact in order to ensure good detection performance. Towards getting the approximate pdf, we use the mutual coherence $\alpha$ of the sensing matrix $A$ defined in (7). Based on the derivation in Appendix [B], we set the approximate post-change pdf as
\[
\tilde{j}^{G_i}_1 = \begin{cases} 
\text{CN}(0, \sigma_n^2 + K_1 \sigma^2_0 \sigma_x^2), & i \notin S, \text{CN}(0, \sigma_n^2 + K_1 \sigma^2_0 \sigma_x^2 + (1 - \alpha^2)\sigma_x^2), & i \in S. 
\end{cases} \quad (16)
\]
Once we get the approximating pdf $\tilde{j}^{G_i}_1$, we proceed in the same way as before, by replacing $f_1^{G_i}$ with $\tilde{j}^{G_i}_1$ for $i \in S$ in (14).

3) Energy CUSUM: In this section, we describe another suboptimal technique, which uses energy of the received signal as the decision statistic. Energy detector has been used previously to detect sparse signals in [14], [15]. Let us define the energy of the observation vector to be $e[t] = \|y[t]\|^2$. Before the change, $e[t]$ is the sum of squares of $2M$ i.i.d zero mean Gaussian random variables of variance $\sigma_n^2$ and follows $\chi^2$ distribution with $2M$ degrees of freedom. This implies $f_0^E = \chi_{2M}^2$. However, for sufficiently large values of $M$, the $\chi^2$ distribution of $e[t]$ can be approximated as Gaussian with appropriate mean and variance, using the central limit theorem (CLT). The approximate pre-change distribution is
\[
\tilde{f}_0^E = N(M\sigma_n^2, M\sigma_n^4), \quad \text{Based on the derivation in Appendix [A], we also obtain the approximate post-change pdf as}
\[
\tilde{f}_1^E = N(\mu_E, \sigma_E^2) \quad \text{where, } \mu_E = K\sigma_n^2 + M\sigma_n^2 \quad (17)
\]
\[
\sigma_E^2 = K\phi_{min}^2 + 2\sigma_n^2 K\phi_{min} + M(\sigma_n^2)^2.
\]
Here, $\phi_{min} = \max \{0, \sigma_n^2(1 - (K - 1)\alpha)\}$. Using these pdf approximations, we run the CUSUM algorithm for energy function $e[t]$. Note that, this method does not use the fact that support of all the signals $x[t]$ remains same after the change.

4) Correlator CUSUM: We now describe the matched-filter/correlator based metric as the decision statistic. Specifically, considering the vector of inner products $g[t] = A^\top y[t]$, we use the maximum inner product (correlation value) $c[t] = \|g[t]\|^2$ as the decision statistic. The correlator based statistic has been previously used for detection of sparse signals in [11]. The pre-change pdf is the distribution of maximum of $N$ i.i.d. exponential random variables, $f_0^C = N(1 - e^{-\lambda_0 c(t)}N^{-1} - \lambda_n e^{-\lambda_n c(t)})$, where $\lambda_n = \frac{1}{\sigma_n^2}$. Based on the derivations in Appendix [B], we get the approximate post-change pdf of $c[t]$ as
\[
\tilde{f}_1^C = K(1 - e^{-\lambda_0 c(t)}N^{-1} - \lambda_n e^{-\lambda_n c(t)})(1 - e^{-\lambda_0 c(t)}N^{-1} - \lambda_n e^{-\lambda_n c(t)}N^{-1} - \lambda_0 e^{-\lambda_0 c(t)})(1 - e^{-\lambda_n c(t)}N^{-1} - \lambda_0 e^{-\lambda_0 c(t)}N^{-1} - \lambda_n e^{-\lambda_n c(t)}), \quad (18)
\]
where $\lambda_0 = \frac{1}{\sigma_n^2 + \lambda_n \sigma_n^2}$ and $\lambda_n = \frac{1}{\sigma_n^2 + \lambda_n \sigma_n^2 + (1 - \alpha^2)\sigma_n^2}$. Correlator CUSUM also does not use the fact that the unknown sparse signal $x[t]$ has the same support for all $t \geq \nu$.

Correlator and Energy CUSUM do not provide a direct mechanism to identify the support set $S$. Hence, at the time (say $T$) when the change is declared by Energy (or Correlator) CUSUM, we run a sparse signal recovery algorithm such as orthogonal matching pursuit (OMP) [23] on the observation $y[T]$ and identify the support.

5) Partial Support Estimation (PSE) CUSUM: A technique to detect sparse signals using a partial estimate of the support is presented in [13]. We combine this detection technique with CUSUM algorithm and employ the same for our change detection problem. Sparse signal recovery algorithms, like OMP, can be employed to obtain a partial estimate $\hat{S}_p$ of support having cardinality $|\hat{S}_p| = K_p$, where $1 \leq K_p \leq K$. The sensing matrix in [13] is chosen to satisfy the condition $AA^\top = I_M$, i.e., its rows are orthonormal. The decision statistic considered here is the total power of the received signal $y[t]$ projected on to the subspace spanned by the partial support estimate $\hat{S}_p$ and normalized by the noise variance $\sigma_n^2$. Specifically, the decision statistic is $p[t] = \|P_{\hat{S}_p} y[t]\|^2$ where the projection matrix is given by $P_{\hat{S}_p} = (A_{\hat{S}_p} A_{\hat{S}_p}^\top)^{-1} A_{\hat{S}_p}^\top$. The pre-change distribution of $p[t]$ is $\chi_{K_p}^2$, which can be approximated as $\tilde{f}_0^E = N(K_p, 2K_p)$, using CLT. From [13], the post-change distribution of $p[t]$ is $\chi_{K_p}^2$ with approximate non centrality parameter as $\tilde{\mu}_{\hat{S}_p} = E[\|P_{\hat{S}_p} x[t]\|^2] = \frac{MK_p}{N} \left(1 + \frac{K - K_p}{M} \right) E[\|x[t]\|^2]$. Here, $E[\|x[t]\|^2]$ can be replaced by $K_0^2$ when $C_\infty = \sigma_n^2 I$. Thus, using a Gaussian approximation due to CLT, the post-change pdf is $\tilde{f}_1^P = N(K_p + \tilde{\mu}_{\hat{S}_p}, 2(K_p + 2\tilde{\mu}_{\hat{S}_p})$. The performance of
C. Sparsity Level Known, Both Support and Signal Variance Unknown

In this section, we consider the case when both $S$ and $C_\mathbf{X}$ are unknown and the signal covariance matrix can take the form $\mathbf{C}_\mathbf{X} = \text{diag}([\sigma^2_1, \cdots, \sigma^2_K])$. We assume that the sparsity level $K$ is known. Also, we assume the knowledge of the upper bound $\sigma_{\text{max}}^2$ and the lower bound $\sigma_{\text{min}}^2$ on the signal variances, such that $\sigma_{\text{min}}^2 \leq \sigma_i^2 \leq \sigma_{\text{max}}^2 \forall i \in S$.

1) Based on pdf-approximation: We use this approach for Aggregate, Energy and Correlator CUSUM algorithms when support and signal variance is unknown. The only difference from the previous case is that the approximations for the post change pdfs of the decision statistics are obtained in terms of $\sigma_{\text{min}}^2$ and $\sigma_{\text{max}}^2$. We obtain these approximations based on the post-change pdf which gives the lowest KL distance from the pre-change pdf, in order to account for the worst case detection delay.

Using the derivations in Appendix A, the post-change distribution for Aggregate CUSUM in (16) is replaced with the approximation,

$$f_{\hat{G}}(t) = \begin{cases} \text{CN}(0, \sigma^2_n + K\alpha^2\sigma_{\text{min}}^2), & i \notin S, \\ \text{CN}(0, \sigma^2_n + K\alpha^2\sigma_{\text{min}}^2 + (1 - \alpha^2)\sigma_{\text{max}}^2), & i \in S. \end{cases}$$

(19)

Using the derivations in Appendix A for the Energy CUSUM metric $e[t] = \|\mathbf{y}[t]\|^2_2$, the post-change distribution is $f_{\hat{E}}(t) = \text{N}(\hat{\mu}_E, \hat{\sigma}_E^2)$ and mean and variance of $e[t]$ are approximated as

$$\hat{\mu}_E = K\phi_{\text{min}} + M\sigma_n^2,$$

$$\hat{\sigma}_E^2 = K\phi_{\text{min}}^2 + 2\sigma_n^2 K\phi_{\text{min}} + M(\sigma_n^2)^2,$$

(20)

(21)

where $\phi_{\text{min}} = \max\{0, \sigma_{\text{min}}^2(1 - \alpha(1 - K - 1))\}$.

From the derivations in Appendix B for the Correlator CUSUM metric $c[t] = \|\mathbf{g}[t]\|^2_2$, the post-change pdf $f_{\hat{C}}(t)$ is same as that in (15), but with the parameters $\lambda_0$ and $\lambda_S$, replaced with their approximations,

$$\hat{\lambda}_0 = \frac{1}{\sigma_n^2 + K\alpha^2\sigma_{\text{min}}^2},$$

$$\hat{\lambda}_S = \frac{1}{\sigma_n^2 + K\alpha^2\sigma_{\text{min}}^2 + (1 - \alpha^2)\sigma_{\text{max}}^2}.$$

(22)

(23)

2) Based on parameter-estimation: So far, we have followed the approach of approximating the post-change pdfs using bounds on the signal variances. In an alternative approach, we can adaptively estimate the unknown parameters [7], [8] in the post-change pdfs and compute the LLRs using these estimated parameters. One such approach, which we refer to as stochastic gradient descent (SGD) CUSUM, is described below.

Let $d[t]$ denote the decision statistic used for change detection and $d[t] = g_i[t], e[t]$ or $c[t]$ for Aggregate, Energy or Correlator CUSUM respectively. Let $\theta$ be the unknown parameter in the post-change pdf. Let the actual value of $\theta$ be equal to $\hat{\theta}$ and let $\hat{\theta}$ be its estimate. Hence, $f_{1,\hat{D},\hat{\theta}}$ is the true post-change pdf. We define the LLR parameterized by $\theta$ as $L_{D,\theta}(d[t]) = \log \frac{L_{D,\theta}(d[t])}{L_{D,\hat{\theta}}(d[t])}$ and the corresponding CUSUM metric as $W_{D,\theta}(t)$. Let the regression function denoting the expected value of LLR at time instant $t$ be $V_t(\theta) = \mathbb{E}_d[L_{D,\theta}(d[t])]$. Since expectation is taken w.r.t the true pdf of $d[t]$, it can be shown from [8] that, $V_t(\theta) < 0$ for $t < \nu$ and $V_t(\theta) = D_{KL}(f_{1,\theta} \parallel f_{0,\theta}) - D_{KL}(f_{1,\hat{\theta}} \parallel f_{0,\theta})$, for $t \geq \nu$. Here, $D_{KL}(f \parallel g)$ is used to denote the KL distance from $f$ to $g$. The post-change $(t \geq \nu)$ regression function $V_t(\theta)$ is maximized when $\hat{\theta} = \theta$, i.e., when the argument of the regression function is equal to the true value of the parameter. This motivates a gradient descent based approach to estimate the unknown parameter, which is described below.

At time $t$, the gradient of the regression function at the present estimate $\hat{\theta}[t]$ is $\frac{V_t(\hat{\theta}[t] - \hat{\theta})}{c}$, as the limit $c \to 0$. For SGD, using stochastic approximation principle [3], [20], we replace the expectation (ensemble average) in $V_t(\theta)$ with an instantaneous approximation using the LLR from the actual values of $d[t]$ and $\hat{\theta}[t]$. Specifically, at time $t = 0$, the estimate $\hat{\theta}[0]$ is initialized to zero. With small positive constants $a$ and $c$, for $t \geq 0$, the estimate is updated as

$$\hat{\theta}[t + 1] = \hat{\theta}[t] + \frac{L_{D,\hat{\theta}}[t] + \epsilon}{c} \sum\limits_{i \in S} \frac{V_t(\hat{\theta}[t] - \hat{\theta})}{c}$$

(24)

(25)

where $L_{D,\hat{\theta}}[t] = \log \frac{L_{D,\hat{\theta}}(d[t])}{L_{D,\hat{\theta}}(d[t])}$ and the algorithm terminates according to rule $R$ in (4).

We use the Aggregate, Energy and Correlator decision statistics for SGD-CUSUM and the implementation with each statistic is described below. From [5] in Appendix B we consider $\theta = \alpha^2 \sigma_{\text{sum}}^2 + (1 - \alpha)\sigma_n^2$ to be the unknown parameter for Aggregate-SGD-CUSUM. The estimate $\hat{\theta}[0]$ is initialized to zero. The post-change pdf of $g_i[t]$ for $i \in S$, parameterized by $\theta$ is given by $f_{1,\hat{G},\theta} = \text{CN}(0, \sigma_n^2 + \theta)$.

For Energy-SGD-CUSUM, $\theta = \phi_{\text{min}}$ in (20) and (21), is treated as the unknown parameter which must be initialized to zero at the start of the algorithm. The approximate parameterized post-change pdf is $f_{1,\hat{E},\theta} = \text{CN}(\mu_{\hat{E}}, \sigma_{\hat{E}}^2)$ with $\mu_{\hat{E}} = K\theta + M\sigma_n^2$ and $\sigma_{\hat{E}}^2 = K\theta^2 + 2\sigma_n^2 K\theta + M\sigma_n^4$.

For Correlator-SGD-CUSUM, the post-change pdf is given by (22) in Appendix B. We consider $\theta = \sigma_n^2$ to be the unknown parameter in $\lambda_S$, so that

$$\hat{\lambda}_S = \frac{1}{\sigma_n^2 + K\alpha^2\sigma_{\text{min}}^2 + (1 - \alpha^2)\theta^2}$$

and initialize $\hat{\theta}[0] = 0$. The parameter $\lambda_0$ in (22) does not depend on $\theta$.

The LLR and CUSUM metric in SGD CUSUM for all the above decision statistics is updated using (24) and (25) respectively.

D. Support, Signal Variance and Sparsity level are Unknown

One must observe that the post-change distributions enlisted in Section III-B and Section III-C depend on the knowledge
of the exact value of sparsity order \( K \) of \( x[t] \). In this section, we address the case when support set \( S \), signal covariance \( C_x \) and sparsity level \( K \) are unknown. However, we assume that an upper bound on the sparsity level \( K_{\text{max}} \) is known, such that \( K \leq K_{\text{max}} \).

1) Based on pdf-approximation: We consider the decision statistics \( d[t] \) equal to \( g_x[t], e[t] \) and \( c[t] \) for change detection using Aggregate, Energy and CorrelatorCUSUM respectively. With a decision statistic \( d[t] \) and unknown sparsity \( K \), we run the CUSUM algorithm \emph{parallelly} for all values of \( \{k : 1 \leq k \leq K_{\text{max}}\} \) and declare change based on the largest CUSUM metric. We use the approximate post-change pdfs of various running CUSUM for each value of \( d \) and sparsity level \( K \) as approximate moments of the parameterized post-change pdf. The CUSUM metric is updated as \( L^{D,k}(d[t]) = \frac{K_{\text{phi}}(d[t])}{K_n(d[t])} \). The CUSUM metric is updated as \( W^{D,k}(d[t]) = (W^{D,k}[t-1] + L^{D,k}(d[t]))^{-1} \), with \( W^{D,k}[t] = 0 \).

The parallel CUSUM change detection rule \( R \) is given by,

\[
R = \begin{cases} 
\text{Declare change} & \text{if } \max_{k \in (1, \ldots, K_{\text{max}})} W^{D,k}[t] > \tau, \\
\text{Continue otherwise.} & 
\end{cases}
\]

2) Based on parameter-estimation: With unknown sparsity level, the implementation of Aggregate-SGD-CUSUM remains same as that described in the previous section. For Energy-SGD-CUSUM algorithm, we treat \( \theta = K_{\phi_{\text{min}}} \) as the unknown parameter in \([20],[21],[22],[23]\) which must be initialized to zero at the start of the algorithm. The approximate moments of the parameterized post-change pdf \( f_{\theta,\sigma_{\theta}} = \text{CN}(\mu_{\theta},(\sigma_{\theta})^2) \), are given by \( \mu_{\theta} = \theta + M\sigma_n^2 \) and \( (\sigma_{\theta})^2 = \sigma_{\phi_{\text{min}}}^2 + 2\sigma_n^2 \theta + M\sigma_n^2 \), where the term \( K_{\phi_{\text{min}}}^2 \) in \([21]\) is approximated as \( K_{\phi_{\text{min}}} = \theta_{\phi_{\text{min}}} \approx \theta_{\phi_{\text{min}}} \).

Since the post-change pdf of the correlator statistic in \([18]\) depends implicitly on sparsity \( K \) and cannot be isolated in the form of a separate parameter, Correlator-SGD-CUSUM cannot be implemented for this case.

IV. SENSING MATRIX DESIGN

In this section, we present deterministic constructions of sensing matrices based on designs from quantum information theory [?]. In addition to low mutual coherence, one of these quantum theoretic constructions has an additional structure in the sensing matrix which allows exact computation of the post-change pdfs of some decision statistics.

A. Unitary Matrix

For a unitary sensing matrix, the approximations for the post change pdfs of decision statistics are obtained by setting mutual coherence \( \alpha \) to be zero. However, these approximations are exact for some scenarios, as given below.

Lemma 1. With unitary sensing matrix, when the signal covariance \( C_x \) and sparsity level \( K \) are known but the support \( S \) is unknown, the exact post-change distributions of the decision statistics \( g_x[t], e[t] \) and \( c[t] \) are obtained by substituting \( \alpha = 0 \) in \([16],[17],[18]\) respectively.

Proof. Follows from Appendix A and Appendix B.

B. Symmetric Informationally Complete Positive Operator Valued Measure (SIC POVM)

In a \( d \)-dimensional Hilbert space, SIC POVM is described by a set of \( d^2 \) rank-1 projectors, \( \mathcal{P}_d = \{P_i = \frac{1}{\sqrt{d}} a_i a_i^* : 1 \leq i \leq d^2\} \), with the property,

\[
\text{tr}(\Pi_i \Pi_j) = \frac{1}{d^2} |\langle a_i, a_j \rangle|^2 = \frac{1 + \delta_{i\ell}d}{d^2(1+d)},
\]

where \( \delta_{i\ell} = 1 \) if \( i = \ell \) and zero, otherwise. From these SIC-POVM projectors, we obtain \( d^2 \) equi-angular vectors \( \{a_i\} \) of unit length such that \( |\langle a_i, a_j \rangle| = \frac{1}{\sqrt{d+1}}, i \neq \ell \). Setting \( M = d \) and \( N \leq d^2 \), we construct a sensing matrix \( A_{M \times N} \) using a subset of the SIC POVM vectors as its columns. With this construction, the magnitude of the inner product between any two distinct columns of \( A \) will be equal to \( \alpha \).

Lemma 2. For an \( M \times N \) sensing matrix constructed using SIC POVM of dimension \( M \), when the signal covariance \( C_x \) and sparsity level \( K \) are known but the support \( S \) is unknown, the exact post-change pdf of the decision statistic \( g_x[t] \) in Aggregate CUSUM algorithm is obtained by substituting \( \alpha = \frac{1}{\sqrt{d+1}} \) in \([19]\).

Proof. Follows from derivations in Appendix B.

Though it is conjectured that SIC POVMs exist for every dimension \( d \), the actual constructions for SIC POVMs are available only for some specific values of \( d \) \([22],[23]\). One of the popular techniques to obtain the SIC POVM vectors \( \{a_i\} \) is to apply Weyl-Heisenberg (WH) displacement group operators to a fiducial vector \([22],[23]\). The Weyl-Heisenberg displacement group operators in dimension \( d \) are generated by the cyclic shift operation \( X \) and its Fourier-transformed version \( \hat{X} \) on a fixed orthonormal basis, say, \( \{e_0, e_1, \ldots, e_{d-1}\} \),

\[
\hat{X} = \sum_{i=0}^{d-1} e_{i+1} e_i^T \quad \text{and} \quad \hat{Z} = \sum_{i=0}^{d-1} \omega_d^{i} e_i e_i^T,
\]

where \( \omega_d = \exp(\frac{2\pi i}{d}) \) is a complex primitive \( d \)-th root of unity and addition is modulo \( d \). Without loss of generality, we can take \( \{e_0, e_1, \ldots, e_{d-1}\} \) to be the standard basis. The elements of the WH group, \( X^a \hat{Z}^b \), can be identified with pairs of integers, \( (a, b) \in \mathbb{Z}_d \times \mathbb{Z}_d \). The displacement operator is defined as \( \hat{D}_{(a,b)} = \tau_{ab} X^a \hat{Z}^b \) where the phase factor \( \tau = -\exp(\frac{2\pi i}{d}) \). Now, the vectors in SIC POVM are constructed as \( a_{(a,b)} = \hat{D}_{(a,b)} a_{(0,0)} \), where \( (a,b) \in \mathbb{Z}_d^2 \).

where \( a_{(0,0)} \) is referred to as the fiducial vector, which is available for some specific dimensions \([22],[23]\).

C. Mutually Unbiased Bases (MUB)

In a Hilbert space of dimension \( d \), MUBs are a set of \( d + 1 \) unitary matrices (orthonormal bases), i.e., \( \mathcal{M}_d = \{U_k = [a_k^0, a_k^1, \ldots, a_k^{d-1}] : 0 \leq k \leq d\} \), such that the following properties are satisfied.

\[
|\langle a_k^q, a_r^q \rangle|^2 = \frac{1}{d}, \quad \forall k \neq q \in \{0,1,\ldots,d\},
\]

\[
|\langle a_k^q, a_r^q \rangle|^2 = \frac{1}{d}, \quad \forall l, r \in \{0,1,\ldots,d-1\}
\]
where \( a_k^l \) is the \( l \)-th column (basis vector) of \( k \)-th unitary matrix \( U_k \) and \( a_r^q \) is the \( r \)-th column (basis vector) of \( q \)-th unitary matrix \( U_q \) in \( \mathcal{M}_d \).

We can design a sensing matrix \( A_{M \times N} \) with mutual coherence \( \alpha = \frac{1}{|\mathcal{M}_d|} \) by using MUBs generated for dimension \( d = M \). Let \( |\mathcal{M}_d| = r \). We take \( r \) columns from every unitary matrix \( \{U_i \in \mathcal{M}_d : 0 \leq i \leq d\} \) and the remaining \( N - r(M + 1) \) columns are chosen, one each from the first \( N - r(M + 1) \) unitary (MUB) matrices. In this construction, the columns of the sensing matrix are uniformly distributed across all the MUB matrices so that, the magnitude of the inner product between any two randomly chosen columns is equal to \( \alpha \) with a high probability. This construction of sensing matrix using MUB try to mimic the equi-angular effect of SIC POVM. Construction of MUBs is available for dimensions \( d = p^n \) where \( p \) is a prime number and \( n \) is a non-negative integer. The procedure to construct MUBs is detailed in [24].

D. Approximately Mutually Unbiased Bases (AMUB)

To overcome the constraint on dimension for the construction of MUBs, we can use AMUBs to design sensing matrices as they can be generated for all dimensions. For any non prime dimension, \( \mathbb{C}^d \), AMUB is a set of \( d + 1 \) unitary matrices (orthonormal bases), i.e., \( A_d = \{V_k = [a_k^0, \ldots, a_k^{d-1}] : 0 \leq k \leq d\} \), but at the cost of relaxing the condition,

\[
|\langle a_k^l, a_r^q \rangle|^2 = \begin{cases} \frac{1 + \alpha(0)}{d} & \text{or} \\ \frac{1 + \alpha(l \log d)}{d} & \end{cases} \quad \forall k \neq q \in \{0, 1, \ldots, d\}, \quad \forall l, r \in \{0, 1, \ldots, d-1\},
\]

where \( a_k^l \) is the \( l \)-th column (basis vector) of \( k \)-th unitary matrix \( V_k \) and \( a_r^q \) is the \( r \)-th column (basis vector) of \( q \)-th unitary matrix \( V_q \) in \( A_d \). Sensing matrices can be constructed from AMUBs in the same manner as that from MUBs. Detailed procedure to construct AMUBs is presented in [25].

It is worth noting that the mutual coherence of the deterministic sensing matrix \( A_{M \times N} \) constructed from SIC POVM, MUB or AMUB is inversely proportional to \( \sqrt{M} \).

E. Random Sensing Matrices

We can construct \( M \times N \) sensing matrices by choosing the \( M \) rows randomly from an \( N \)-dimensional discrete Fourier transform (DFT) matrix. Similarly we can also construct sensing matrices by generating random Bernoulli or complex Gaussian ensembles [9]. It is important that the columns of these sensing matrices should be normalized to unit norm. Using randomly generated sensing matrices yields approximate distributions for all decision statistics.

V. SIMULATION RESULTS

In this section, we present the results obtained from Monte Carlo simulations of the algorithms described in Section III and our inferences thereof, by comparing their performance based on worst case decision delay \( D_w \) in (3) and average run length \( T_r \) in (2). To find \( D_w \), we fix the change point to be \( \nu = 20 \) for all the simulations. We compare the performance with SNR defined as

\[
\text{SNR (dB)} = 10 \log_{10} \frac{E \|x\|_2^2}{E \|n\|_2^2} = 10 \log_{10} \frac{\sum_{i \in S} \sigma_i^2}{M \sigma_n^2}. \tag{27}
\]

In the simulations, we set \( \sigma_n^2 = 1 \).

A. Effect of Sensing Matrix

The dimensions of the sensing matrix \( A \) are fixed as \( M = 124, N = 200 \). We define the compression ratio of the sensing matrix as \( c_r = \frac{M}{N} = 0.62 \). We set \( K = 5 \) and the support indices of \( x[t] \) are randomly selected from \( \{1, \ldots, N\} \). The non-zero entries of \( x[t] \) are drawn randomly from \( \mathcal{CN}(0, \sigma_x) \), where \( C_x = \sigma_x^2 I \) and \( \sigma_x^2 \) is chosen according to (27) for a particular value of SNR. We consider deterministic sensing matrices constructed using SIC POVM, MUB and AMUB, using the procedure detailed in Section IV. For sensing matrix designed using MUB, we choose the prime power closest to 124 and take \( M = 125 \). A truncated DFT matrix with randomly chosen rows is also considered. We also construct random sensing matrices with i.i.d. complex Gaussian (CN) entries and i.i.d. Bernoulli (BER) entries.

Figure 1 shows the performance of various sensing matrices in terms of \( T_r \) versus \( D_w \), at SNR = \(-10\) dB. Unitary (\( M \)) denotes \( M \times M \) unitary matrix that has the same number of measurements as \( A \) where as Unitary (\( N \)) denotes \( N \times N \) unitary matrix that retains the same number of columns as \( A \). We infer that deterministic sensing matrices constructed from SIC POVM, MUB and AMUB give better performance as compared to random sensing matrices since they have a lower value of mutual coherence. In addition, these deterministic matrices yield near-exact post-change distributions for various decision statistics. Thus, as evident from Figure 1, the performances of SIC POVM and MUB sensing matrices, are closest to that of Unitary (\( M \)) matrix, followed by AMUB, DFT, complex Gaussian and Bernoulli sensing matrices. Figure 2 shows similar trends when the number of columns in \( A \) are increased to \( N = 124^2 \) and the compression ratio decreases to \( c_r = 0.008 \). However, the detection delay \( D_w \) for a specific \( T_r \) increases for Aggregate and Correlator CUSUM. Also, Correlator CUSUM fails for random complex Gaussian and Bernoulli sensing matrices as the independence assumption on the entries of \( g[t] \) does not hold true due to the high mutual coherence of these random sensing matrices when \( c_r < 0.5 \).

B. Comparison of Various Decision Statistics

We fix the dimensions of \( A \) to be \( M = 124, N = 200 \) and compression ratio \( c_r = 0.62 \). The columns of \( A \) are SIC POVMs generated for dimension, \( M = 124 \). The sparsity level \( K = 5 \) and signal covariance \( C_x = \sigma_x^2 I \), where \( \sigma_x^2 \) is chosen according to (27).

Figure 3 illustrates the \( T_r \) versus \( D_w \) plot for various algorithms at SNRs \(-20\) dB, \(-10\) dB and \( 0 \) dB. Aggregate CUSUM algorithm performs better than other support-oblivious algorithms at all SNRs, but with additional computational complexity. We also observe that when SNR is \(-10\) dB and above, Correlator CUSUM performs better than Energy CUSUM. On the other hand, when SNR is very low at \(-20\) dB, the performance of Energy CUSUM becomes better than Correlator CUSUM, since entries in the correlation vector \( g[t] \) are highly corrupted by noise. The dashed lines in Figure 3 plot the performance of the parallel CUSUM
rule in (26) for the aforesaid algorithms when the signal covariance is known but the support and sparsity level are unknown. We fix the maximum sparsity level \(K_{\text{max}} = 10\). A slight deterioration in performance of all decision statistics is observed as compared to the case when the value of \(K\) is perfectly known. The performance of PSE CUSUM is poorer than other algorithms because accurate sparse support recovery \(x[t]\) becomes difficult to achieve with OMP algorithm at SNRs below \(0\) dB. For PSE, the size of the recovered support set \(K_p\) is kept identical to the true sparsity level \(K\), which gives the best performance compared to any other value of \(K_p\).

We also consider the SIC POVM sensing matrix of size \(M = 124\), \(N = 124^2\), with compression ratio \(c_r = 0.008\). All the other parameters are kept same as above. In Figure 3(a), at SNR 0 dB, the detection performance of various algorithms follows the same trend as that shown in Figure 3(a), but the overall detection delay \(D_w\) is larger for a given \(T_r\). In Figure 3(b), we see that Aggregate CUSUM algorithm performs close to Ideal CUSUM at SNR\(= -10\) dB. At SNR\(= -20\) dB, Aggregate CUSUM performs better than Energy CUSUM for higher values of the \(T_r\) and poorer than Energy CUSUM for relatively smaller values of the \(T_r\). Both Aggregate and Correlator CUSUM inherently assume/approximate that the entries in the correlation vector \(g[t]\) are independent. However, when \(N\) is very large as compared to \(M\), this independence approximation becomes inaccurate and the performance of these algorithms suffers.

C. Unknown Support, Signal Variance and Sparsity level

Now, we address various cases regarding the knowledge of support, sparsity level and signal variance and present the corresponding simulation results. When the signal variance is unknown, we generate the non-zero entries of \(x[t]\) with (unequal) variances which are uniformly distributed in the interval \([\sigma_{\text{min}}^2, \sigma_{\text{max}}^2]\). When the signal variances are unknown, SGD CUSUM algorithm (which tries to estimate the unknown parameters) as discussed in Section III-C can be used in addition to the detection techniques that are based on approx-
imating the post-change pdf. For the SGD CUSUM, we set step size \( a = 0.01 \) and window length \( c = 0.05 \) in \((24)\), in all the simulations.

When the sparsity level \( K \) is assumed to be known a priori but the support \( S \) and signal covariance \( C_x \) are unknown, we employ the techniques given in Section III-C and plot their performance in Figure 5.

When the support \( S \), signal covariance \( C_x \) and the sparsity level \( K \) of the signal are unknown, the methods outlined in Section III-D are used and the results are shown in Figure 6.

Some important observations are highlighted below.

1) If the signal variance is high \( \sigma^2_{\text{max}} = \sigma^2 \) or the compression ratio is large \( c_r > 0.5 \), Aggregate CUSUM performs better than Energy and Correlator CUSUM.

2) If the signal variance is small \( \sigma^2_{\text{max}} < \sigma^2 \) or the compression ratio is small \( c_r \ll 0.5 \), Energy CUSUM based on the signal energy \( \| y[t] \|_2^2 \) performs, in general, better than Correlator and Aggregate CUSUM, which use the correlation statistics \( g[t] \).

3) In general, SGD CUSUM performs better than the corresponding pdf approximation based counterparts.

4) In most cases, Aggregate-SGD-CUSUM gives the best performance.

D. Percentage of Sparse Recovery

In addition to detecting change, we are also interested in recovering the support of the signal \( x[t] \), when the change is detected. Note that, Aggregate CUSUM has an inherent mechanism to find the support by selecting the locations corresponding to the \( K \)-largest CUSUM metrics \((15)\), when the change is detected. On the other hand, for Energy and Correlator CUSUM, once the change is detected, we run OMP algorithm to find the support. We define percentage of support recovery to be the fraction of the support that is recovered correctly. From Figure 7 for an average run length of \( 5 \times 10^3 \), Aggregate CUSUM gives higher percentage of recovery than...
to propagation delays) in the reception of signals transmitted from different users. We assume that the timing offsets of all the users in the network are upper bounded by \( \Delta \in \mathbb{Z}^+ \).

Suppose the \( i^{th} \) user has a delay of \( \delta_i \in \{0, \cdots, \Delta\} \), the code sequence received at the central node from the \( i^{th} \) user will be \( \mathbf{a}_i^T \cdot \delta_i = [0^T \mathbf{a}_i^T \ 0^T \delta_i \ 0^T \chi_i - \delta_i] \). The timing offset values of users are not usually available at the central node. Let \( P \) be the number of users in the system. We form an augmented sensing matrix \( \mathbf{A}_\Delta \) of size \( (M + \Delta) \times P(\Delta + 1) \) which contains all sequences (including all the possible timing offsets) of the form \( \{\mathbf{a}_i, \chi_i\} \), \( 1 \leq i \leq P, 0 \leq \delta_i \leq \Delta \). After the change point, a subset of users \( S \) become active so that the observations are \( \mathbf{y}[t] = \mathbf{A}_\Delta \mathbf{x}_\Delta[t] + \mathbf{n}[t] \), where the locations of the non-zero entries in \( \mathbf{x}_\Delta[t] \) indicate the active users and their corresponding timing offsets.

We consider code constructions based on SIC POVM and compare it with Gold codes. For SIC-POVM of length \( M \), there are a total \( M^2 \) sequences. From the construction in Section \( \text{IV-B} \), all the cyclic shifts of a SIC POVM sequence are also SIC POVM sequences. However, since the augmented matrix \( \mathbf{A}_\Delta \) contains time shifts up to \( \Delta \) for each sequence, we can use only \( M \left\lfloor \frac{M}{\Delta + 1} \right\rfloor \) sequences as valid codes (in order to avoid the scenario where code of one user is highly correlated with the time delayed code of another user). Hence, the maximum number of users that can be accommodated with \( M \) length SIC POVM codes is \( P = M \left\lfloor \frac{M}{\Delta + 1} \right\rfloor \). In a very similar manner, we can construct \( \mathbf{A}_\Delta \) from the cyclic shifts of bipolar Gold codes. Gold codes exist for \( M \) values up to 1564 which contains \( M + \Delta \) sequences.

\( \mathbf{b}_i \) denoting the cyclic shift of \( \mathbf{b}_i \) by \( \tau \), it has been shown in [29] that \( |\langle \mathbf{b}_{i,\tau_1}, \mathbf{b}_{k,\tau_2} \rangle| \leq r(n) \) where

\[
r(n) = \begin{cases} 
\frac{2^{\frac{n+1}{2}}}{M^{n+1}}, & \text{if } n \text{ is odd,} \\
\frac{2^\frac{n}{2} + 1}{M^n}, & \text{if } n \text{ is even.}
\end{cases}
\]

In our simulations, we fix the number of users to be \( P = 1500 \) and the maximum admissible timing offset as \( \Delta = 8 \). We set \( M = 124 \) for SIC POVM and \( M = 127 \) for Gold codes. For this scenario, the mutual coherence of the augmented sensing matrix \( \mathbf{A}_\Delta \) for SIC POVM and Gold code based constructions is \( \alpha = 0.1564 \) and \( \alpha = 0.1969 \), respectively. Figure 8(a) shows that the user activity detection delay versus average run length is nearly the same for both constructions, with Energy and Aggregate CUSUM. In
we define \( A \) as composition of signal are correlated in time may be of interest for specific decision statistics at different SNR levels. The problem of the signal are unknown. Using deterministic sensing matrices when the support, signal variance and sparsity level use the pdf-approximation and parameter-estimation based approaches when the support, signal variance and sparsity level are of rank \( S \). Let \( \Phi, \Phi^* \) be non-zero eigenvalues.

VI. CONCLUSION

In this paper, we address the change detection problem with sparse signals by combining the techniques from compressive sensing with asymptotically optimal CUSUM algorithm. We use the pdf-approximation and parameter-estimation based approaches when the support, signal variance and sparsity level of the signal are known. Using deterministic sensing matrices with low mutual coherence further enhances the detection performance. We also analyze the detection performance of various decision statistics at different SNR levels. The problem of change detection when the non-zero entries of the sparse signal are correlated in time may be of interest for specific applications and may serve as a future scope of this work. Also, further research may be taken up to develop alternate techniques for detection when the distribution of the non-zero entries of the sparse signal, after the change point, is not known a priori.

APPENDIX A

APPROXIMATE PDF FOR SIGNAL ENERGY

The covariance matrix of post change observation \([9]\) is 
\[ C_Y = A_S C_X A_S^* + \sigma_n^2 I_M. \]
Let \( \bar{U} \bar{\Phi} \bar{U}^* \) denote the eigen decomposition of \( A_S C_X A_S^* \). Note that, when \( A_S \) is of rank \( K \), then \( \Phi \) has \( K \) real, non-zero eigenvalues. For convenience, let first \( K \) entries \( \{ \phi_i, 1 \leq i \leq K \} \) in the diagonal be non-zero.

We define \( z[t] = \bar{U}^* y[t] \) and note that \( \|z[t]\|^2 = \|y[t]\|^2 \). Since \( \bar{U} \) is unitary. Now, \( z[t] \) is also a zero mean complex Gaussian random vector with covariance matrix \( C_z = \bar{\Phi} + \sigma_n^2 I_M \). Since the off diagonal elements of \( C_z \) are zero, the entries of \( z[t] \) are uncorrelated and hence, independent. Each entry of \( z[t] \) is distributed as 
\[ z_i \sim \begin{cases} CN(0, \sigma_i^2 + \phi_i), & 1 \leq i \leq K, \\ CN(0, \sigma_i^2), & i > K. \end{cases} \]

We have \( \|z\|^2 = \sum_{i=1}^M |z_i|^2 \), with each \( |z_i|^2 \) being distributed as 
\[ |z_i|^2 \sim \begin{cases} \exp \left( \frac{1}{\sigma_i^2 + \phi_i} \right), & 1 \leq i \leq K, \\ \exp \left( \frac{1}{\sigma_i^2} \right), & i > K. \end{cases} \]

The Gramian matrices, \( A_S C_X A_S^* = A_S C_X A_S^* \) have the same \( K \) non-zero eigenvalues.

Due to CLT, we use a Gaussian approximation for the post change pdf of \( e[t] \), i.e., \( f_1^E = N(\mu_E, \sigma_E^2) \), where,
\[ \mu_E = E(E[t]) = E(\|z\|^2) = \sum_{i=1}^K E(|z_i|^2) \]
\[ = \sum_{i=1}^K E(|z_i|^2) + \sum_{i>K} E(|z_i|^2) \]
\[ = \sum_{i=1}^K (\phi_i + \sigma_i^2) + (M - K)(\sigma_n^2) \]
\[ = \sum_{i=1}^K \phi_i + M\sigma_n^2. \]

\[ \sigma_E^2 = \text{Var}(E[t]) = \text{Var}(\|z\|^2) = \sum_{i=1}^M \text{Var}(|z_i|^2) \]
\[ = \sum_{i=1}^K \text{Var}(|z_i|^2) + \sum_{i>K} \text{Var}(|z_i|^2) \]
\[ = \sum_{i=1}^K (\phi_i + \sigma_i^2)^2 + (M - K)(\sigma_n^2)^2 \]
\[ = \sum_{i=1}^K \phi_i^2 + 2\sigma_n^2 \sum_{i=1}^K \phi_i + M(\sigma_n^2)^2. \]

Next, we consider the following special cases:

Case 1: The unknown signal covariance matrix \( C_X \) is of the form \( \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2) \). Since, \( \sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_K^2 \), the worst case KL distance between post and pre-change pdf occurs when \( \sigma_i^2 = \sigma_{i+1}^2, 1 \leq i < K \). Assuming this and using the bound in (28), we approximate the post change pdf with the smallest possible value for \( \mu \) in (29) and variance in (30), which are given in (29) and (31) respectively.

Case 2: The signal covariance is of the form \( C_X = \sigma_n^2 I_K \). In (29), the sum of eigenvalues of the matrix \( \sigma_n^2 A_S A_S^* \) is \( \sum_{i=1}^K \phi_i = \text{tr}(\sigma_n^2 A_S A_S^*) = \text{tr}(\sigma_n^2 A_S A_S^*) = K \sigma_n^2 \), which
where \( n \) is approximated as

\[
\text{post-change pdf of } \text{diag}(g(31), \alpha) = \text{approximate post-change pdf of } c
\]

modified values of rate parameters \( \sigma \) distance between post and pre-change pdf occurs when \( \sigma \).

Case 2: The signal covariance matrix is of the form

\[
C = \sigma_{\text{diag}(g)}^2 I
\]

We substitute \( \sigma_{\text{sum}}^2 = K \sigma_{\text{min}}^2 \) and \( \sigma_i^2 = \sigma_{\text{min}}^2 \) in (31) and (32) to obtain the approximate post-change pdfs of \( g_i[t] \) and \( c[t] \) respectively, as given in (16) and (18).

\[\text{REFERENCES}\]

[1] G. Lorden, “Procedures for reacting to a change in distribution,” Ann. Math. Stat., vol. 42, pp. 1897–1908, Dec. 1971.

[2] M. Pollak, “Optimal detection of a change in distribution,” Ann. Stat., vol. 13, pp. 206–227, Mar. 1985.

[3] V. V. Veeravalli and T. Banerjee, “Chapter 6 - quickest change detection,” in Academic Press Library in Signal Processing: Volume 3 Array and Statistical Signal Processing, pp. 209 – 255, Elsevier, 2014.

[4] E. S. Page, “Continuous inspection schemes,” Biometrika, vol. 41, pp. 100–115, June 1954.

[5] Y. Mei, “Sequential change-point detection when unknown parameters are present in the pre-change distribution,” Ann. Stat., vol. 34, no. 1, pp. 92–122, 2006.

[6] I. V. Nikiforov, “A suboptimal quadratic change detection scheme,” IEEE Trans. Info. Theory, vol. 46, pp. 2095–2107, Sept. 2000.

[7] C. Li, H. Dai, and H. Li, “Adaptive quickest change detection with unknown parameter,” in Proc. ICASSP, pp. 3241–3244, Apr. 2009.

[8] V. Singamsetty, N. Nair, S. Bhashyam, and A. P. Kamu, “Change detection with unknown post-change parameter using Kiefer-Wolfowitz method,” in Proc. ICASSP, pp. 3919–3923, Mar. 2017.

[9] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Proc. Mag., vol. 25, pp. 21–30, Mar. 2008.

[10] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” IEEE Trans. Info. Theory, vol. 50, pp. 2231–2242, Oct. 2004.

[11] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, “Sparse signal detection from incoherent projections,” in Proc. ICASSP, vol. 3, May 2006.

[12] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, “Signal processing with compressive measurements,” IEEE J. Sel. Topics Signal Process., vol. 4, pp. 445–460, Apr. 2010.

[13] T. Wimalajeewa and P. K. Varshney, “Sparse signal detection with compressive measurements via partial support set estimation,” IEEE Trans. Info. Process. over Networks, vol. 3, pp. 46–60, Mar. 2017.

[14] J. Haupt and R. Nowak, “Compressive sampling for signal detection,” in Proc. ICASSP, vol. 3, pp. 1599–1512, Apr. 2007.

[15] Z. Wang, G. R. Arce, and B. M. Sadler, “Subspace compressive detection for sparse signals,” in Proc. ICASSP, pp. 3873–3876, Mar. 2008.

[16] H. Zheng, S. Xiao, and X. Wang, “Sequential compressive target detection in wireless sensor networks,” in Proc. IEEE Int. Conf. Commun., pp. 1–5, June 2011.

[17] C. Lei, J. Zhang, and Q. Gao, “Unknown and arbitrary sparse signal detection against background noise,” in Proc. IEEE Int. Conf. Signal Process. Process., pp. 46–49, Oct. 2010.

[18] C. Alippi, G. Boracchi, and B. Wohlfert, “Change detection in streams of signals with sparse representations,” in Proc. ICASSP, pp. 5252–5256, May 2014.

[19] G. Fellouris, G. V. Moustakides, and V. V. Veeravalli, “Multistream quickest change detection: Asymptotic optimality under a sparse signal,” in Proc. ICASSP, pp. 6444–6447, Mar. 2017.

[20] J. Kiefer and J. Wolfowitz, “Stochastic estimation of the maximum of a regression function,” Ann. Math. Stat., pp. 462–466, 1952.

[21] S. Li and G. Ge, “Deterministic sensing matrices arising from near orthogonal systems,” IEEE Trans. Info. Theory, vol. 60, pp. 2291–2302, Apr. 2014.

[22] A. J. Scott and M. Gras, “SIC-POVMs: A new computer study,” J. Math. Phys., vol. 51, 2010.

[23] M. Grassl and A. J. Scott, “Fibonacci-Lucas SIC-POVMs,” J. Math. Phys., vol. 58, Dec. 2017.

[24] A. Klappenecker and M. Rütteler, Constructions of Mutually Unbiased Bases, pp. 137–144, Berlin, Heidelberg: Springer, 2004.

[25] I. E. Shparlinski and A. Winterhof, Constructions of Approximately Mutually Unbiased Bases, pp. 793–799, Berlin, Heidelberg: Springer, 2006.

[26] T. Oskiper and H. V. Poor, “Online activity detection in a multiuser environment using the matrix CUSUM algorithm,” IEEE Trans. Info. Theory, vol. 48, pp. 477–493, Feb. 2002.

[27] V. Boljanovi, D. Vukobratovi, P. Popovski, and . Stefanov, “User activity detection in massive random access: Compressed sensing vs. coded slotted ALOHA,” in Proc. IEEE Workshop Signal Process. Adv. Wireless Commun., pp. 1–6, July 2017.

[28] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Info. Theory, pp. 4655–4666, Dec. 2007.

[29] R. Gold, “Optimal binary sequences for spread spectrum multiplexing (cosr.),” IEEE Trans. Info. Theory, vol. 13, pp. 619–621, Oct. 1967.