





Approaching Sub-Nyquist Boundary: Optimized Compressed Spectrum Sensing Based on Multicoset Sampler for Multiband Signal

Zihang Song , Graduate Student Member, IEEE, Jian Yang , Student Member, IEEE, Han Zhang , Member, IEEE, and Yue Gao , Senior Member, IEEE

Abstract—Compressed spectrum sensing naturally pursues the use of fewer sampling resources to achieve spectrum support reconstruction and signal recovery. The theoretical lower boundary of averaging sampling rate to recover the multiband signal has been proved to be twice the Landau rate. However, it is still unreachable in practice. Based on the multicoset sampling architecture, this paper analyzes the influencing factors on perfect spectrum reconstruction from three aspects: data model, sampling pattern and greedy reconstruction algorithms, for which practical and feasible optimization schemes are proposed. To reduce redundant reconstruction for the multiple measurement vectors (MMV) signal model, a block MMV model is proposed to improve the accuracy in the spectrum support set reconstruction process. A sampling pattern selection algorithm is proposed to ensure a higher success rate of the spectrum support reconstruction to optimize the sensing matrix. We also deduced the representation of the signal and noise energy in the reconstructed spectrum based on the mathematical model of compressed sensing. We thus proposed a non-orthogonal double-threshold matching pursuit algorithm to avoid a high false-alarm rate due to the manually set converging conditions for matching pursuit algorithms. Numerical experiments on real-world wideband signals are carried out to demonstrate the feasibility and advantages of the proposed approaches. With integrated optimization, the required sampling density to ensure perfect reconstruction is approaching the sub-Nyquist sampling boundary.

Index Terms—Compressed spectrum sensing, multiple measurement vectors, matching pursuit algorithms, multicoset sampling.

I. INTRODUCTION

WITH the accelerated arrival of the Internet of Everything, the conflict between spectrum supply and demand continues to intensify. The high-density, high-intensity,

Manuscript received 27 October 2021; revised 25 April 2022 and 10 July 2022; accepted 1 August 2022. Date of publication 11 August 2022; date of current version 2 September 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chiara Ravazzi. This work was supported by the Engineering and Physical Sciences Research Council of United Kingdom under the Grant EP/R00711X/2. (Corresponding author: Yue Gao.)

Zihang Song, Han Zhang, and Yue Gao are with the Institute for Communication Systems, University of Surrey, GU2 7XH Guildford, U.K. (e-mail: zihang.song@surrey.ac.uk; han.zhang@surrey.ac.uk; yue.gao@ieee.org).

Jian Yang is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: hityangjian@stu.hit.edu.cn).

Digital Object Identifier 10.1109/TSP.2022.3198186

and multi-frequency electromagnetic waves released by various communication systems make the electromagnetic environment around us increasingly complex. Currently, the low-frequency spectrum below 6 GHz and the millimeter-wave frequency band above 6 GHz has also been allocated for 2 G~5 G use. The increase in communication demand has increased the shortage of spectrum resources, and 6 G will require more spectrum resources [1].

However, while spectrum resources are scarce, there is also a low utilization rate of spectrum in time and space [2]. The Dynamic Spectrum Access (DSA) technology allows Cognitive Radio (CR) devices to opportunistically select idle frequency bands unoccupied by the licensed Primary User (PU) for communication, thereby greatly improving spectrum utilization [3], [4].

DSA inherently requires CR devices to have the ability to passively sense a fairly wide range of the spectrum, so as to select an idle channel. However, the traditional wideband sampling method requires a high-rate Analog-to-Digital Converter (ADC), which is expensive and has high energy consumption. Therefore, this solution is not suitable for general CR equipment. Taking advantage of the sparsity feature of the spectrum, the Compressed Sensing (CS) theory is introduced into the field of spectrum sensing [5]. According to the CS theory, when certain sampling conditions are satisfied, sparse signals can be reconstructed with high probability from the compressed sampling points. This means that only sampling at a sub-Nyquist rate and the corresponding reconstruction algorithm are sufficient to complete the sensing of occupied channels in a wide spectrum range. This solution has lower power consumption and cost and is more suitable for general CR equipment [6].

As a famous sub-Nyquist sampling architecture, the multicoset sampler has been extensively studied for its feasible and comprehensible implementation [7], [8], [9], [10]. The multicoset sampler achieves periodical nonuniform sampling by multiple low-rate sampling cosets with random delays. At the theoretical level, the research on the minimum sampling rate required for the perfect reconstruction of sparse signals have been very mature [11], [12]. While, at the application level, the average sampling rate required by the existing compressed spectrum sensing scheme to reconstruct the signal still cannot reach the theoretical lower bound [13], [14], [15], [16].

This theory-to-practice gap mainly comes from three aspects: the limitations of the compressed sampling architecture, the mismatch of the signal model, and the ability of the reconstruction algorithm, detailed as follows:

Signal model mismatch: The signal model is complementary and interrelated with the compressed sampling architecture. A theoretical Infinite Measurement Vectors (IMV) model is commonly deduced from the multicoset structure and is proved to be equivalent to a finite-dimension Multiple Measurement Vectors (MMV) model via the continuous to finite (CTF) blocks proposed in [12]. Compared with the Single Measurement Vector (SMV) model, the MMV model is more efficient and robust to noise, so it has been extensively studied [17], [18], [19], [20], [21]. In MMV model it is usually assumed that all the columns of the solution matrix are joint sparse, while this assumption can be inaccurate, especially in the blind case when the real spectrum support is unknown. The fact that the actual signal support and the MMV row division may not completely match leads to extra sampling resources to recover redundant unoccupied bands around the real support in the MMV model, which is a significant factor that results in a necessary sampling rate higher than the theoretical value. To our knowledge, there has been no relative research on the model mismatch problem.

Compressed sampling architecture: In the premise of the signal sparsity, a measurement matrix that strictly follows Null-Space Property (NSP), Spark constraints, and Restricted Isometry Property (RIP) guarantees successful signal reconstruction, while those properties are unnecessary and require high computational complexity to validate [22], [23], [24]. For the multicoset sampler, the measurement matrix is formally-restricted and its property largely depends on the delay pattern of analog delays imposed on different sampling cosets. The delay pattern selection is stated as a combinatorial problem in [12] but the authors did not provide a method or criterion for the pattern design. In [25] an arithmetic progression form of the pattern design method is proposed for a full spark sensing matrix but lacks a rigorous argument. An obvious counterexample is that any all-odd or all-even pattern will naturally have its Kruskal rank as 1. Another research [26] analyzes the relationship between the condition number of the selected columns of the sensing matrix and the reconstruction error, and provides several searching strategies for multicoset patterns. However, this approach only considers the equation-solving stage with known spectrum support but does not mention the way to get the correct support in a blind case. [27] use an exhausted searching strategy but the computational cost tends to increase explosively as the channel number increases.

Reconstruction algorithm: Reconstructing signals from compressed sampling points is a key part of compressed spectrum sensing. The performance of the reconstruction algorithm directly affects the quality of the reconstruction result [23], [28], [29]. In consideration of computational efficiency, greedy algorithms, represented by the matching pursuit type of algorithms, are often employed in the reconstruction stage. Matching pursuit type of algorithms usually reconstructs the signal support set to determine the frequency band occupied by the PU's signal,

then apply the least-squares (LS) method to acquire the optimal estimation of the signal. Therefore, the most important factor to evaluate the algorithm is how accurate the original signal support set can be retrieved. Among the widely-used greedy algorithms Matching Pursuit (MP) and Simultaneous Orthogonal Matching Pursuit (SOMP) are most commonly referred to as the representative of non-orthogonal and orthogonal algorithms [30], [31], [32], [33], [34], [35]. The condition of convergence for those algorithms is usually chosen from three scheduled boundaries: a fixed iteration number, a given lower bound of the residual energy, or a given threshold of the channel power in the reconstructed signal [36]. The reconstructed support is sensitive to the manually-set conditions which can hardly be precisely predicted in the blind case. Ulteriorly, the LS result will be apart from the ground truth with inaccurate support as input.

A. Main Contributions

Without increasing hardware complexity and average sampling rate, this paper aims to fully explore the information contained in the sampling points, improve the performance of support set detection and signal reconstruction, and narrow the gap between the actual compressed sampling rate and the theoretical boundary. The major contributions of the paper are summarized as follows:

- We analyze the potential mismatch between the traditional MMV signal model for joint-block sparse signals and the real spectrum support and deduce the maximum sampling rate under the mismatch, and propose a block MMV model to reduce redundant reconstruction in the spectrum support set.
- We provide a general criterion to evaluate the property of the multicoset sampling pattern based on the column correlation of the sensing matrix, and hereby propose an optimization algorithm for delay pattern selection. Compared with the randomly selected pattern, the sampling pattern optimized by the proposed algorithm ensures a higher success rate of the spectrum support reconstruction.
- We also deduce a theoretical representation of the noise energy in the reconstructed signal in different channels based on the mathematical model of compressed sensing and propose a double-threshold matching pursuit algorithm to avoid high false-alarm rate as a result of manually-set converging conditions.

B. Outline

The rest of this paper is organized as follows. The mathematical model of multiband signal, the formulation of the problem, and a brief overview of the multicoset sampling scheme and the reconstruction algorithm are given in Section II. The proposed block recovery method and pattern selection algorithm is given in Section III and Section IV. We also developed a double-threshold matching pursuit algorithm, which is detailed in Section V. Numerical experiments are described in Section VI. Common notation, as summarized in Table I, is used throughout the paper.

TABLE I
NOTATION

$x(t)$	continuous-time baseband complex signal
$X(f)$	Fourier transform of $x(t)$
$s(t)$ or $\mathbf{S}(f)$	signal component in $x(f)$ or $X(f)$
$n(t)$ or $\mathbf{N}(f)$	signal component in $n(f)$ or $N(f)$
\mathcal{T}	uncountable support set of $X(f)$
\mathcal{T}_i	i^{th} support in \mathcal{T}
$\lambda(\mathcal{T})$	the Lebesgue measurement of \mathcal{T}
\mathcal{S}	the row indexes of $\mathbf{x}(f)$ that are non-identically zero
$x_{c_i}[n]$	discrete sampling points in the i^{th} coset
$\mathbf{y}(f)$ or \mathbf{Y}	measurement matrix in IMV or MMV model
$\mathbf{x}(f)$ or \mathbf{X}	original signal in IMV or MMV model
\mathbf{A}	sensing matrix
Θ	optimal approximation of \mathbf{X}
C	multicoset sampling pattern
\mathbf{Y}_i^d or \mathbf{X}_i^d	i^{th} block of \mathbf{Y} or \mathbf{X} in the block MMV model with block coefficient d
\mathcal{Y}, \mathcal{X} and \mathcal{A}	measurement matrix, signal matrix and sensing matrix in block MMV model
\cdot^H	conjugate-transpose of \cdot
$\cdot_{i,j}$	$(i, j)^{\text{th}}$ element of \cdot
$\cdot_{\mathcal{S}}$	submatrix of \cdot composed of columns indicated by \mathcal{S}
$\hat{\cdot}$	the estimation of \cdot

II. MATHEMATICAL MODEL

A. Signal Model

The multiband signal model is often referred to as a fundamental mathematical approach to the actual signal received by a wideband receiver [12]. In this paper, all the discussions is under a complex multiband signal model. The continuous signal at the receiver baseband can be expressed in time domain as

$$x(t) = s(t) + n(t) \quad (1)$$

where $s(t)$ is the signal component and $n(t)$ is Gaussian noise. In frequency domain, the received signal is

$$X(f) = S(f) + N(f), \quad (2)$$

where $S(f)$ and $N(f)$ denote the signal components and the noise components, respectively. $X(f)$ is band-limited to a frequency range $\mathcal{F} = [-f_{Nyq}/2, f_{Nyq}/2]$. In this range, N_{sig} disjoint intervals are occupied by $S(f)$ with support set $\mathcal{T} = \{\mathcal{T}_i | 1 \leq i \leq N_{sig}\}$ ($\mathcal{T} \subseteq \mathcal{F}$), and the maximum bandwidth occupied by single transmission support \mathcal{T}_i does not exceed B . Additive noise $N(f)$ spreads through \mathcal{F} .

B. Blind Compressed Spectrum Sensing and Minimum Average Sampling Rate

Without any prior knowledge of the band locations of \mathcal{T} (hereafter referred to as the ‘‘blind’’ case), the task of compressed spectrum sensing can be concluded as recovering the frequency support \mathcal{T} and reconstructing $S(f)$ from the discrete sampling points of $x(t)$ with average sampling rate under f_{Nyq} .

In consideration of the hardware, it is always a practical demand to achieve perfect reconstruction in a certain \mathcal{F} with an average sampling rate as low as possible. For years, there has been enormous research on the theoretical minimal sampling rate, among which two boundaries are notable:

- 1) In the non-blind case (known \mathcal{T}), Landau first proved that for known spectrum support $\mathcal{T} \subseteq \mathcal{F}$, the lower bound of the necessary sampling rate for exact signal recovery is $\lambda(\mathcal{T})$ (the Lebesgue measurement of \mathcal{T}) [11];
- 2) In the blind case, the lower bound of the average sampling rate is proved to be [12]

$$\min \{2\lambda(\mathcal{T}), f_{Nyq}\}, \quad (3)$$

that is, by defining the spectrum occupation rate $\Omega = \lambda(\mathcal{T})/f_{Nyq}$, the minimum average sampling rate is twice the Landau rate for $\Omega \leq 0.5$ or Nyquist rate for $\Omega > 0.5$.

For the multiband signal model in the blind case and $\Omega \leq 0.5$, the sampling rate should not be lower than $2N_{sig}B$ to ensure perfect reconstruction according to (3). However, the existing sampling schemes still require a rate much higher than $2N_{sig}B$ to achieve stable signal reconstruction [12]. Based on multicoset sampling architecture, this theory-practice gap is discussed.

C. Multicoset Sampling

By defining the Nyquist frequency as $f_{Nyq} = 1/T$, where T is the Nyquist sampling period, the structure of multicoset sampler can be described as p uniform sampling cosets at a sub-Nyquist rate $f_s = 1/(LT)$ with a unique time delay indicated by $C = \{c_i\}_{i=1}^p$. With common settings $p < L$ and $c_i \in \mathbb{Z}$ satisfying

$$0 \leq c_1 < c_2 < \dots < c_i < \dots < c_p \leq L - 1. \quad (4)$$

The Nyquist samples are divided into blocks of length L . In each block, only p ($p < L$) samples indexed by pattern C are actually sampled by the multicoset sampler. The sampled data in channel i can be expressed as

$$x_{c_i}[n] = x((nL + c_i)T), \quad n = 0, 1, 2, \dots \quad (5)$$

It is worth noting that no upsampling [12] is carried out on the subsamples as commonly done in this scheme. Thus, the influence of the sampling pattern, namely C , is not reflected in the sampled data. As compensation, a digital filter is applied in the digital domain on each coset of the signal, then the filtered frequency-domain signal in coset i is proved to be the linear combination of L samples from the original signal's Nyquist spectrum

$$\begin{aligned} &G_{c_i}(f)X_{c_i}(e^{j2\pi fLT}) \\ &= \frac{1}{LT} \sum_{n=-\lfloor \frac{L}{2} \rfloor}^{+\lfloor \frac{L}{2} \rfloor - 1} X\left(f - \frac{n}{LT}\right) \exp\left(-j2\pi \frac{nc_i}{L}\right), \\ &i \in \{1, 2, \dots, p\}, f \in [0, f_s], \end{aligned} \quad (6)$$

where the filter $G_{c_i}(f)$ applies phase delays on the signal based on c_i

$$G_{c_i}(f) = \begin{cases} e^{-j2\pi c_i T f}, & f \in [0, f_s] \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

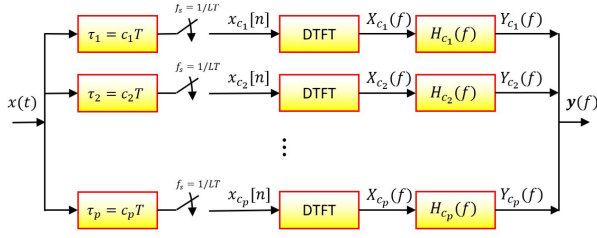


Fig. 1. Basic structure of multicoset sampler.

and

$$X_{c_i}(e^{j2\pi fLT}) = \sum_{n=-\infty}^{\infty} x_{c_i}[n]e^{-j2\pi fLTn} \quad (8)$$

is the discrete-time Fourier transform (DTFT) of the digitized signal in channel i . By denoting the frequency-domain compressed measurement $Y_{c_i}(f)$ as

$$Y_{c_i}(f) = LTG_{c_i}(f)X_{c_i}(e^{j2\pi fLT}), \quad (9)$$

the relationship between the original spectrum $X(f)$ and $Y_{c_i}(f)$ is clearly described by

$$Y_{c_i}(f) = \sum_{l=-\lfloor \frac{L}{2} \rfloor}^{+\lfloor \frac{L}{2} \rfloor - 1} X\left(f + \frac{l}{LT}\right) \exp\left(j2\pi \frac{lc_i}{L}\right), \quad (10)$$

indicating that the i^{th} -coset measurement $Y_{c_i}(f)$ can be represented by a linear combination of L different segments of the signal's original spectrum $X(f)$.

Stacking all the p cosets together bring the IMV form of (10)

$$\mathbf{y}(f) = \mathbf{A}\mathbf{x}(f), \quad f \in \left[0, \frac{1}{LT}\right), \quad (11)$$

where

$$\mathbf{y}(f) = [Y_{c_1}(f) \ Y_{c_2}(f) \ \cdots \ Y_{c_p}(f)]^T, \quad (12)$$

and

$$\mathbf{x}(f) = \begin{pmatrix} X\left(f + \frac{-L/2}{LT}\right) \\ X\left(f + \frac{-L/2+1}{LT}\right) \\ \vdots \\ X\left(f + \frac{L/2-1}{LT}\right) \end{pmatrix} \quad (13)$$

has L rows, corresponding to the L equal segments of the original spectrum $X(f)$ within \mathcal{F} . The coefficient matrix \mathbf{A} is a $p \times L$ matrix whose $(i, l)^{\text{th}}$ element is given by

$$A_{i,l} = e^{j2\pi \frac{c_i}{L}(-2/L+l-1)}. \quad (14)$$

Model (11) is a linear projection process from L dimension to p dimension. This projection can also be regarded as aliasing the L Sections with p groups of coefficients specified by the p rows in \mathbf{A} , as illustrated in Fig. 2.

The IMV model (11) is set up on the basis of infinite measurements. In practice, only a limited number of samples can be acquired in one frame of the observing window. Thus, the signal spectrum reconstructed by the above-mentioned method

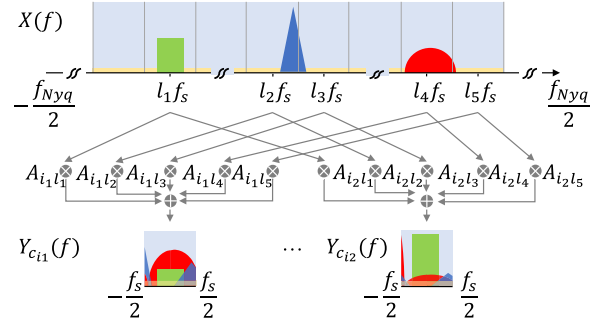


Fig. 2. The diagram illustrates the relationship between the compressed measurement $Y_{c_i}(f)$ and the original spectrum $X(f)$, where the green rectangular, the blue triangle, and the red arch represent the PU's transmissions, and the yellow stripes represent out-of-band noise. For multicoset sampling, the height changes of the transmissions depicted in this figure represent phase delays instead of amplitude gains.

is actually discrete. Denote the window length to be N sequential sampling points per coset, a practical multicoset sampler usually generates an MMV model

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \quad (15)$$

where \mathbf{Y} is a $p \times N$ measurement matrix whose $(i, n)^{\text{th}}$ element is

$$Y_{i,n} = Y_{c_i}\left(\frac{n-1}{NLT}\right) \quad (16)$$

and \mathbf{X} is a $L \times N$ matrix whose $(l, n)^{\text{th}}$ element is

$$X_{l,n} = X\left(\frac{-L/2 + l - 1 + (n-1)/N}{LT}\right). \quad (17)$$

D. Blind Spectrum Reconstruction

In the case $p < L$, model (15) is a set of underdetermined equations and has an infinite number of solutions. The low utilization rate of the spectrum leads to an underlying assumption that $X(f)$ is sparse. Thus, the compressed sensing theory can be applied to transfer (11) to a l_0 -norm minimization problem

$$\min_{\hat{\Theta} \in \mathbb{C}^{L \times N}} \|\hat{\Theta}\|_0 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{A}\hat{\Theta}\|_2 < \epsilon, \quad (18)$$

where $\hat{\Theta}$ is considered as the optimal estimation of \mathbf{X} .

Problem (18) is an NP-hard problem. A mathematical approximation is to transform it into a group l_1 -norm minimization problem by convex relaxation and then reconstruct the signal using convex optimization strategies, such as basis pursuit (BP). The l_1 -norm approach can guarantee the sparse exact solution of (18) under certain RIP requirement on the sensing matrix [28]. However, in the multicoset scenario, no known RIP guarantee is given for the deterministic construction of the sensing matrix \mathbf{A} , specifically, a partial inverse Fourier matrix by row selection [36]. Moreover, this kind of approach is characterized by high computational complexity thus is unsuitable for applications that requires real-time performance [37]. Another more efficient and widely-studied kind of approach is greedy algorithms, including MP, SOMP, etc [31], [32], [33], [34].

Greedy algorithms try to find a locally optimal solution to the l_0 -norm minimization problem by seeking the signal support set represented by the channel indexes, namely

$$\mathcal{S} = \left\{ \mathcal{S}_i \in \mathbb{Z} \mid 1 \leq \mathcal{S}_i \leq L, \right. \\ \left. -\frac{f_{Nyq}}{2} + \frac{\mathcal{S}_i - 1}{LT} + f \in \mathcal{T}, \exists f \in \left[0, \frac{1}{LT} \right) \right\}. \quad (19)$$

Both MP and SOMP algorithms seek to find the channel with the next strongest signal components in each iteration cycle. The only difference is that MP deducts only one maximum component from the residual but SOMP deducts all the selected components from the original measurement in one iteration cycle. Usually, there are two options for the cut-off conditions for MP and SOMP. One is setting the maximum number of iteration cycles to K , i.e. the number of channels that contain signal components. However, for the blind case, the transmissions number and their locations are unknown in advance. An irrational estimation of K may lead to incomplete or excessive reconstruction. Thus, an alternative cut-off condition ϵ_R is applied to judge whether all the signal components have been deducted from the residual. A better estimation of ϵ_R requires certain prior knowledge of the noise environment, which will be detailed in Section V.

The output of both algorithms is set as the reconstructed support $\hat{\mathcal{S}}$ as the optimal estimation of \mathcal{S} , as well as the LS signal reconstruction.

III. PROPOSED BLOCK MMV MODEL FOR ACCURATE SUPPORT RECONSTRUCTION

By taking advantage of the clustered distribution feature of the actual radio spectrum, the columns of \mathbf{X} are often considered as sharing the same sparse structure and can be simultaneously processed. The size requirement for \mathbf{Y} to perfectly reconstruct \mathbf{X} is deduced in Proposition 1.

Proposition 1: For the MMV problem (15), let \mathbf{X} be a row-sparse matrix with K non-zero rows, the multicostet sampler should satisfy

$$p \geq \min\{2K, L\} \quad (20)$$

for getting a unique solution of \mathbf{X} .

The proof of Proposition 1 is given in Appendix A.

The MMV model treats each column of \mathbf{X} as a vector with the same sparse structure based on the joint sparse character of $\mathbf{x}(f)$, which means that the elements in the same row of \mathbf{X} are simultaneously treated as zero or non-zero. However, the real band division is hardly like the same way as L equal-length intervals.

From CR's point of view, if both the occupied and the unoccupied frequency points are contained in a certain row of \mathbf{X} , then that row should be picked out to the recovered support and all the elements in that row are reconstructed in the recovery algorithm.

Based on the multiband model and $L \leq 1/(BT)$, there exists two possible cases of the relationship between real transmission support \mathcal{T}_i and the rows in \mathbf{X}

- 1) the corresponding frequency points in \mathcal{T}_i are completely located in a certain row of \mathbf{X} ;
- 2) the corresponding frequency points in \mathcal{T}_i is distributed in two adjacent rows of \mathbf{X} .

In an MMV problem, the row sparsity K is defined as the total number of rows that contains corresponding frequency points in the transmission support \mathcal{T} . Consider the worst case where all \mathcal{T}_i in \mathcal{T} belong to case 2) and the indexes of the selected rows all differ from each other, K takes the maximum value $2N_{sig}$. According to proposition 1, to ensure a unique K -sparse solution, the average sampling rate in the MMV problem should be no less than $4N_{sig}/(LT)$, which is no lower than twice the theoretical lower bound.

The reason for this gap is that the joint-sparse assumption in the MMV model causes "useless" frequency points that originally belongs to the noise band to have to be picked out as "signal" because they are model-compliantly distributed to the same channel with the frequency points that belongs to the real transmissions. For example, according to the band division in Fig. 2, a totally of 5 in L channels are occupied, which means 10 cosets are needed to ensure all the transmissions are perfectly detected and reconstructed, while the theoretical lower bound indicates a rate $3B$ that no more than 3 cosets can provide.

As we have noticed that such a sampling rate gap does not exist in the SMV model because in SMV the elements treated as non-zero in the sparse vector are truly occupied, while in MMV the row-sparse hypothesis enlarges the support set unnecessarily. In the multicostet sampling scheme, we can increase L to divide the whole spectrum into denser blocks, but setting $L \gg 1/(BT)$ is inefficient because the joint-sparse nature of the spectrum is not utilized. On the other hand, aimlessly increasing L will simultaneously require a larger coset number p to guarantee the necessary sampling rate, causing more complicated implementation.

We hereby propose a block MMV model to strive for a lower unnecessary sampling rate without performance reduction based on the existing $L \leq 1/(BT)$ multicostet scheme, in other words, to improve detection and reconstruction performance without increasing the average sampling rate. This approach decompose (15) into d independent MMV problems:

$$\mathbf{Y}_i^d = \mathbf{A}\mathbf{X}_i^d, \quad i = 1, 2, \dots, d. \quad (21)$$

where \mathbf{Y}_i^d is the i^{th} block acquired by evenly dividing \mathbf{Y} columnwisely into d individual $p \times N/d$ submatrixes, namely

$$\mathbf{Y}_i^d = \begin{pmatrix} Y_{c_1}[id-d] & Y_{c_1}[id-(d-1)] & \cdots & Y_{c_1}[id-1] \\ Y_{c_2}[id-d] & Y_{c_2}[id-(d-1)] & \cdots & Y_{c_2}[id-1] \\ \vdots & \vdots & \ddots & \vdots \\ Y_{c_p}[id-d] & Y_{c_p}[id-(d-1)] & \cdots & Y_{c_p}[id-1] \end{pmatrix}, \\ i = 1, 2, \dots, d. \quad (22)$$

From (15) it is easy to derive that \mathbf{X}_i^d is the corresponding submatrix of \mathbf{X} composed of its $((i-1)d+1)^{th}$ to the $(i, d)^{th}$ column. Parameter d is called the block coefficient.

The proposed approach is illustrated in Fig. 3. The signal number $N_{sig} = 3$ but they are divided into 5 different channels.

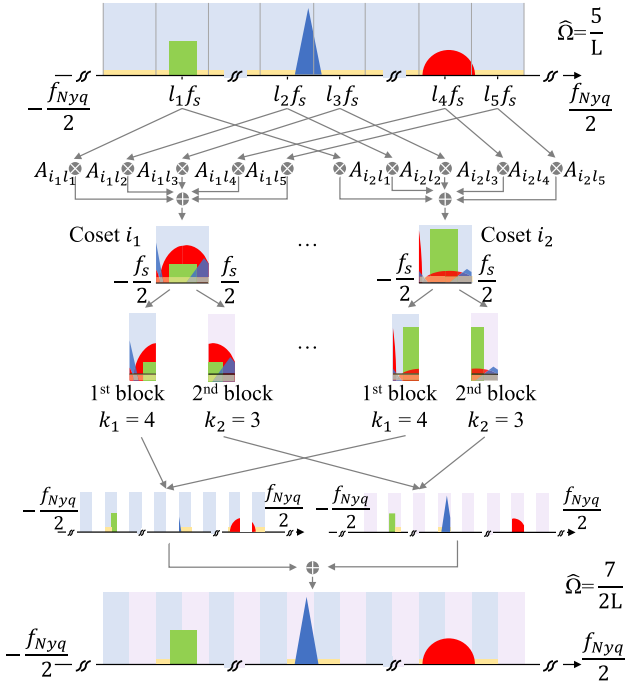


Fig. 3. A conceptual illustration of the spectrum recovery based on block MMV model with $d = 2$.

Therefore, each $Y_{c_i}(f)$ is an aliased spectrum of 5 non-zero signal components. If we directly reconstruct $\mathbf{x}(f)$ then the row sparsity $K = 5$, which means $p \geq 10$ is required for perfect reconstruction. In the case of constant L , we manually divide each row of $\mathbf{y}(f)$ evenly into 2 blocks. We can find that all the 1st blocks are aliased by 4 signal components and all the 2nd blocks are aliased by 3 components. Therefore $p \leq 8$ is enough to individually reconstruct from the 1st blocks and the 2nd blocks.

Proposition 2: Let $x(t)$ be a multiband signal. For the blind case, if $L \leq \frac{1}{BT}$, then when $d \rightarrow \infty$, the row sparsity of \mathbf{X}_i^d does not exceed N_{sig}

The proof of Proposition 2 is given in Appendix B.

Noted that the case $d = 1$ correspond to the original MMV model. Proposition 2 indicates that the row sparsity of individual matrix \mathbf{X}_i^d can be reduced as d increases. Based on Proposition 1 and 2, as the partition coefficient $d \rightarrow \infty$, the minimum sampling rate for perfect reconstruction approaches the theoretical lower bound $2N_{sig}/(LT)$ by solving each of the d matrixes \mathbf{X}_i^d independently.

In order to reduce the number of calls of the matching pursuit algorithm. We can regroup all the d models in a block MMV model

$$\mathbf{Y} = \mathcal{A}\mathbf{X}, \quad (23)$$

where the new measurement matrix \mathbf{Y} of size $dp \times N/d$ is composed by recombining \mathbf{Y}_i^d in rows, namely,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1^{dT} & \mathbf{Y}_2^{dT} & \dots & \mathbf{Y}_d^{dT} \end{bmatrix}^T. \quad (24)$$

The new sensing matrix \mathcal{A} is of size $dp \times dL$ composed by d matrix \mathbf{A} on the diagonal.

$$\mathcal{A} = \text{diag}(\underbrace{\mathbf{A}, \mathbf{A}, \dots, \mathbf{A}}_d) \quad (25)$$

And the structure of \mathcal{X} can be easily derived as

$$\mathcal{X} = \begin{bmatrix} \mathbf{X}_1^{dT} & \mathbf{X}_2^{dT} & \dots & \mathbf{X}_d^{dT} \end{bmatrix}^T. \quad (26)$$

Because the spanning spaces of all the d blocks in \mathcal{A} are orthogonal, it can be easily proved that $\text{spark}(\mathcal{A}) = \text{spark}(\mathbf{A})^1$ and $\mu(\mathcal{A}) = \mu(\mathbf{A})$ (the definition of $\mu(\cdot)$ will be given in (28) indicating the column correlation of the matrix). Thus, with model (23), the uniqueness of solution (in noise-free case) and the reconstruction performance (in noise case) can be guaranteed, as can be done with model (15).

The reconstruction algorithms mentioned in Section II-D can be directly applied on problem (23), then the original solution can be acquired by recombine \mathcal{X} as \mathbf{X} .

With model (23), the improvement of Fig. 3 can be understand from the spectrum occupation perspective. When $d = 1$, \mathcal{X} is identical to \mathbf{X} , which has L rows and 5 of them are non-identically zero. The estimated occupation rate (if correctly estimated) is $\hat{\Omega} = 5/L$. When $d = 2$, \mathcal{X} has totally $2L$ rows, in which 7 rows are non-identically zero. The estimated occupation rate is $\hat{\Omega} = 7/(2L) < 5/L$. As d increasing, $\hat{\Omega}$ will converge on the real occupation rate Ω . Thus, according to Proposition 1, the necessary sampling rate tend to converge on the theoretical lower bound (3). In other words, with the same sampling rate, the detection and reconstruction performance will be improved as d increasing.

When N is a finite number, d can be chosen as an integer divisor of N . The selection of d needs to be considered to seek a balance between detection and reconstruction performance and computing efficiency as will be detailed in Section VI. Generally, given a premise that $L \leq 1/(BT)$, a small value of d like 2 or 4 can bring obvious performance improvement compared with $d = 1$, while a larger value of d may improve the detection performance and reconstruction accuracy to some extent, but the computational burden can be unworthy.

As is concerned, the computational burden for using the block MMV model is analyzed. The time complexity of the SOMP and MP algorithms for the traditional MMV model are both $\mathcal{O}(pLN)$ [31]. By replace model (15) with model (23), the parameters of the problem scale are changed as $p \sim dp$, $L \sim dL$ and $N \sim N/d$. Thus the time complexity for SOMP and MP algorithms for block MMV model is $\mathcal{O}(dpLN)$, which is only linearly increased compared to the traditional MMV model with any $d > 1$. By limiting d in a reasonable range ($p \leq 10$ for instance), the computation time for the greedy algorithms can be controlled within the same order.

The proposed block MMV model alleviates the waste of resources caused by the traditional MMV model when dealing with multiband signals in blind cases. Specifically, for the

¹The spark of a matrix is the smallest number of columns that are linearly dependent.

traditional MMV model, N_{sig} transmissions with bandwidth B have to be reconstructed under an average sampling rate no less than $4N_{sig}B$ (four times the Landau Rate). For the block MMV model, the requirement on sampling rate converges to $2N_{sig}B$ (twice the Landau Rate, i.e. the theoretical lower bound) with block coefficient d increasing. Note that an ideal condition, where there is no restriction on computational cost, and the actual sampling rate f_s of a single coset is arbitrarily valued, is given as a premise here for the convergence.

IV. DELAY PATTERN SELECTION FOR SENSING MATRIX OPTIMIZATION

Proposition 1 is only a necessary condition but not a sufficient one for perfectly reconstructing \mathbf{X} . The property of the sensing matrix \mathbf{A} can have significant effect the performance of reconstruction. For a multicoset sampler with fixed parameter L and p , \mathbf{A} is determined by the delay pattern C . In [12], a “universal” pattern refers to a pattern C that yields a full-spark sensing matrix \mathbf{A} , namely

$$\text{spark}(\mathbf{A}) = p + 1. \quad (27)$$

A universal pattern can guarantee the existence of a unique K -row-sparse solution of (15) for $2K \leq p$ in noise-free cases, and most patterns used in the multicoset sampling scheme can meet this constraint except for a few exemptions. Furthermore, a stronger requirement on the regularity of \mathbf{A} is always desired for the approximation approaches to problem (18), especially for the myopic greedy matching pursuit algorithms. A sensing matrix \mathbf{A} with atoms of stronger orthogonality can help the matching pursuit algorithms converge faster to the correct support and improve performance in noised cases [38].

One well-known property to depict the similarity between \mathbf{A} and a standard independent bases is RIP, but it is hard to calculate. Research [39] proved that examining the column correlation coefficient is a feasible criterion for the reconstruction condition. When the columns of \mathbf{A} meet certain uncorrelation conditions, the RIP property can be satisfied with high probability. The column correlation coefficient $\mu(\mathbf{A})$ of the sensing matrix \mathbf{A} is defined as

$$\mu(\mathbf{A}) = \max_{1 \leq i < j \leq L} \frac{|\langle \mathbf{A}_{:,i}, \mathbf{A}_{:,j} \rangle|}{\|\mathbf{A}_{:,i}\|_2 \|\mathbf{A}_{:,j}\|_2}. \quad (28)$$

The smaller $\mu(\mathbf{A})$ is, the more likely \mathbf{A} is to meet the RIP property, then the requirement to signal sparsity is less strict, heralding a more robust CS system. In the compressed sensing case, $p < L$ is a fundamental assumption to ensure signal compression. In the case of constant L , the expectation of $\mu(\mathbf{A})$ by randomly selecting a pattern increases as p being reduced [40]. However, considering the hardware complexity, it is impractical to implement a multicoset sampler with a large coset number. Thus, it is quite important to preliminary screen the pattern C to avoid a bad pattern with high $\mu(\mathbf{A})$.

An intuitive idea is to go through every possible pattern among the C_L^p combinations to get a pattern that makes $\mu(\mathbf{A})$ minimum, but this may not be an efficient method. Although a pattern with the minimum $\mu(\mathbf{A})$ is always desired for reconstruction,

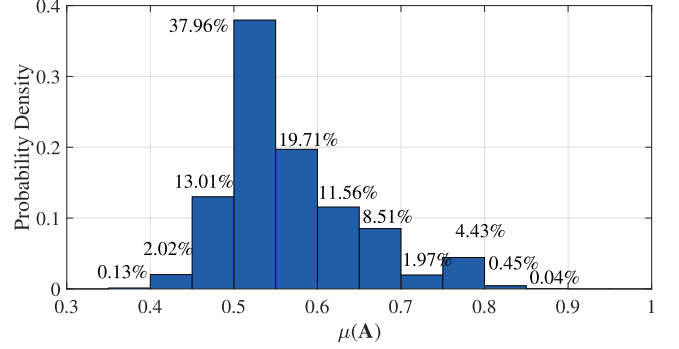


Fig. 4. The histogram of the distribution $\mu(\mathbf{A})$ corresponding to all the patterns given $L = 30$ and $p = 8$, sorted in descending order.

Algorithm 1: Pattern Selection Algorithm.

Input: $L; p$
Output: C

- 1: $t \leftarrow 1, \mu_0 \leftarrow 1, C_0 \leftarrow \emptyset$
- 2: **while** $t \leq \tau$ **do**
- 3: randomly choose a pattern $C_t \neq C_{t-1}$ from all the C_L^p possible patterns
- 4: calculate \mathbf{A}
- 5: $\mathbf{G} \leftarrow \text{abs}\{\mathbf{A}^H \mathbf{A}\}$
- 6: $\mu_t \leftarrow \min_{2 \leq i \leq L} \|\mathbf{G}_{1,i}\|$
- 7: **if** $\mu_t < \mu_{t-1}$ **then**
- 8: $C \leftarrow C_t$
- 9: **end if**
- 10: **end while**

we show that a relatively small value of $\mu(\mathbf{A})$ can serve as a good approximation to significantly improve the reconstruction performance.

Fig. 4 shows the distribution of $\mu(\mathbf{A})$ of all the possible patterns given $L = 30$ and $p = 8$. About 2.15% percent of the values are distributed in the range $[0.35, 0.45]$, and about 0.13% percent of the values are distributed in the range $[0.35, 0.4]$. By 100 times of random selection we have 88.6% chance to get a pattern yielding $\mu(\mathbf{A}) < 0.45$; by 1000 times of random selection we have 99.99999996% chance to get $\mu(\mathbf{A}) < 0.45$ and 72.77% chance to get $\mu(\mathbf{A}) < 0.4$. Those probabilities indicate that a pattern with a approximately minimum $\mu(\mathbf{A})$ can be acquired within rather few times of random selections compared to $C_{30}^8 = 5852925$ (exhaustive search). Similar results goes for other combinations of L and p since their $\mu(\mathbf{A})$ shares the similar distribution. We hereby propose a pattern selection algorithm for the multicoset sampler, as shown in Algorithm 1.

For the multicoset sampler, the module of the inner product of any two columns of \mathbf{A} is

$$|\langle \mathbf{A}_{:,i}, \mathbf{A}_{:,j} \rangle| = \left| \sum_{m=1}^p e^{j2\pi c_m(i-j)/L} \right|, \quad (29)$$

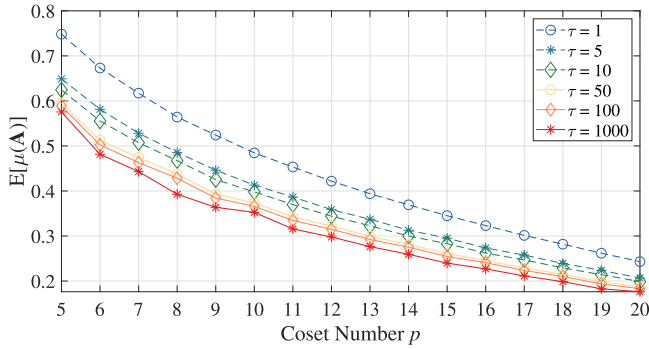


Fig. 5. The mathematical expectation of minimum $\mu(\mathbf{A})$ by randomly selecting a pattern from all the possible combinations for τ times for different coset number p from 5 to 20 ($L = 30$).

which is only related to the difference $|i - j|$. Note that $|\sum_{m=1}^p e^{j2\pi c_m(i-j)/L}| = |\sum_{m=1}^p e^{j2\pi c_m(i-j+L)/L}|$. Thus,

$$\mathbf{G} = \text{abs}\{\mathbf{A}^H \mathbf{A}\} \quad (30)$$

is a circulant matrix, where $\text{abs}\{\cdot\}$ denotes modulus operation to each element in the matrix. For any column in \mathbf{A} , we can always find another column so that their correlation coefficient equals $\mu(\mathbf{A})$. Because $\|\mathbf{A}_{:,i}\|_2 = \sqrt{p}$ is constant for all $1 \leq i \leq p$, we only need to examine the first row of \mathbf{G} to determine $\mu(\mathbf{A})$ of the corresponding pattern.

In Algorithm 1 the number of iterations $\tau \leq C_L^p$ can be manually set for a balance between computing efficiency and reconstruction accuracy for general signal. Usually a relatively small τ can bring significant reduction on $\mu(\mathbf{A})$, while the gain on reduction on $\mu(\mathbf{A})$ decreases as τ increasing (an example is shown in Fig. 5). With $\tau \rightarrow C_L^p$, $\mathbf{E}[\mu(\mathbf{A})]$ converges to the minimum value.

V. PROPOSED DOUBLE-THRESHOLD MATCHING PURSUIT ALGORITHM

A. Accumulative Error in SOMP Algorithm

The flow of MP and SOMP algorithms are detailed together in Algorithm 2 for convenience of discussion. The main difference between MP and SOMP is the choice of atoms (i.e. the columns of \mathbf{A}) for the update of residuals, shown in step 5 and step 6. For MP, only one atom is selected to be subtracted from the residual in each iteration, while SOMP updates the residual using all the selected atoms in each iteration.

Assume that both algorithms are applied to a same dataset. During the first iteration, a same atom indicated by $\hat{\alpha}_1$ should be selected by both algorithms so that the residual \mathbf{R}_1 should be identical. For the second iteration, we can get $\hat{\alpha}_2^{(\text{MP})} = \hat{\alpha}_2^{(\text{SOMP})} = \hat{\alpha}_2$ from $\mathbf{R}_1^{(\text{MP})} = \mathbf{R}_1^{(\text{SOMP})} = \mathbf{R}_1$. The second residuals are calculated as

$$\mathbf{R}_2^{(\text{MP})} = \mathbf{R}_1 - \mathbf{A}_{:, \hat{\alpha}_2} (\mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{A}_{:, \hat{\alpha}_2})^{-1} \mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{R}_1 \quad (31)$$

and

$$\begin{aligned} \mathbf{R}_2^{(\text{SOMP})} &= \mathbf{R}_0 - \mathbf{A}_{\hat{\mathcal{S}}_2} (\mathbf{A}_{\hat{\mathcal{S}}_2}^H \mathbf{A}_{\hat{\mathcal{S}}_2})^{-1} \mathbf{A}_{\hat{\mathcal{S}}_2}^H \mathbf{R}_0 \\ &= \mathbf{R}_1 - \mathbf{A}_{\hat{\mathcal{S}}_2} (\mathbf{A}_{\hat{\mathcal{S}}_2}^H \mathbf{A}_{\hat{\mathcal{S}}_2})^{-1} \mathbf{A}_{\hat{\mathcal{S}}_2}^H \mathbf{R}_1, \end{aligned} \quad (32)$$

Algorithm 2: MP/SOMP Algorithm.

Input: \mathbf{Y} ; \mathbf{A} (of size $p \times L$); ϵ_R

Output: $\hat{\mathcal{S}}$, $\hat{\Theta}$

- 1: $t \leftarrow 1$; $\mathbf{R}_0 \leftarrow \mathbf{Y}$; $\hat{\mathcal{S}}_0 \leftarrow \emptyset$; $\mathbf{A}_0 \leftarrow \emptyset$; $E_R \leftarrow \infty$;
 $\hat{\Theta} \leftarrow \mathbf{0}$
- 2: **while** $t \leq p \vee E_R > \epsilon_R$ **do**
- 3: $\hat{\alpha}_t \leftarrow \arg \max_i \|\mathbf{A}_{:,i}^H \mathbf{R}_{t-1}\|_2$, $i = 0, 1, \dots, L - 1$
- 4: $\hat{\mathcal{S}}_t \leftarrow \hat{\mathcal{S}}_{t-1} \cup \hat{\alpha}_t$
- 5: $\mathbf{A}_t \leftarrow \mathbf{A}_{:, \hat{\alpha}_t}$ (MP) or $\mathbf{A}_t \leftarrow \mathbf{A}_{\hat{\mathcal{S}}_t}$ (SOMP)
- 6: $\hat{\Theta}_t \leftarrow \arg \min_{\Theta} \|\mathbf{Y} - \mathbf{A}_t \Theta\|_2$
- 7: $\mathbf{R}_t \leftarrow \mathbf{R}_{t-1} - \mathbf{A}_t \hat{\Theta}_t$ (MP) or $\mathbf{R}_t \leftarrow \mathbf{Y} - \mathbf{A}_t \hat{\Theta}_t$ (SOMP)
- 8: $E_R \leftarrow \|\mathbf{R}_t\|_F^2$
- 9: $t \leftarrow t + 1$
- 10: **end while**
- 11: $\hat{\mathcal{S}} \leftarrow \hat{\mathcal{S}}_t$
- 12: fill the rows of $\hat{\Theta}$ indexed by $\hat{\mathcal{S}}$ with $\hat{\Theta}_t$

where $\hat{\mathcal{S}}_2 = \{\hat{\alpha}_1, \hat{\alpha}_2\}$. Then we apply Schimidt orthogonalization to $\mathbf{A}_{\hat{\mathcal{S}}_2}$:

$$\mathbf{b}_{\hat{\alpha}_1} = \mathbf{A}_{:, \hat{\alpha}_1} - \mathbf{A}_{:, \hat{\alpha}_2} (\mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{A}_{:, \hat{\alpha}_2})^{-1} \mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{A}_{:, \hat{\alpha}_1}, \quad (33)$$

where $\mathbf{b}_{\hat{\alpha}_1}$ is a vector orthogonal to $\mathbf{A}_{:, \hat{\alpha}_2}$ in the column space of $\mathbf{A}_{\hat{\mathcal{S}}_2}$, then $\mathbf{R}_2^{(\text{SOMP})}$ can be rewritten as

$$\begin{aligned} \mathbf{R}_2^{(\text{SOMP})} &= \mathbf{R}_1 - \mathbf{A}_{:, \hat{\alpha}_2} (\mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{A}_{:, \hat{\alpha}_2})^{-1} \mathbf{A}_{:, \hat{\alpha}_2}^H \mathbf{R}_1 \\ &\quad - \mathbf{b}_{\hat{\alpha}_1} (\mathbf{b}_{\hat{\alpha}_1}^H \mathbf{b}_{\hat{\alpha}_1})^{-1} \mathbf{b}_{\hat{\alpha}_1}^H \mathbf{R}_1. \end{aligned} \quad (34)$$

Compared to the residual (31) of MP algorithm, a term $\mathbf{b}_{\hat{\alpha}_1} (\mathbf{b}_{\hat{\alpha}_1}^H \mathbf{b}_{\hat{\alpha}_1})^{-1} \mathbf{b}_{\hat{\alpha}_1}^H \mathbf{R}_1$ is subtracted in the residual of SOMP in the second iteration. The term indicates the projection of \mathbf{R}_1 on $\mathbf{b}_{\hat{\alpha}_1}$ direction. Because \mathbf{A} is an over-complete dictionary in which most atoms are not orthogonal to each other, the atoms in \mathbf{A} are not likely to be orthogonal to $\mathbf{b}_{\hat{\alpha}_1}$, which lead to excessive subtraction of the components of other signals on $\mathbf{b}_{\hat{\alpha}_1}$ direction. Some undetected signals may be weakened by the excessive subtraction in the residual and become hard to be found in the subsequent iterations. This phenomenon tends to deteriorate along with iterations.

For MP, the excessive subtraction problem does not exist. Thus, MP usually achieves a higher detection probability than SOMP under the same iteration conditions. However, the non-orthogonal character slows up the convergence of MP, causing an increased number of iterations and thereby higher false-alarm probability.

B. Reconstruction Noise Analysis and DTMP Algorithm

To achieve both high detection probability and low false-alarm probability in the estimated support set, we propose an energy-based approach to distinguish real transmission support from the reconstructed signal, detailed as follows.

Recall the signal representation (2) which divide the frequency-domain received signal $X(f)$ into signal component $S(f)$ and noise component $N(f)$. We write in the similar way

the discrete matrix form of (2) as

$$\mathbf{X} = \mathbf{S} + \mathbf{N}, \quad (35)$$

where \mathbf{S} and \mathbf{N} are matriculated PUs' noise-free transmissions and noise, respectively. Both \mathbf{S} and \mathbf{N} are of size $L \times N$ entered like (17). \mathbf{S} has only K non-zero rows corresponding to the channel occupation, and \mathbf{N} is considered white here with Gaussian-distributed elements.

Similarly, the reconstructed signal $\hat{\mathbf{\Theta}}$ (defined in (18)) can be divided as

$$\hat{\mathbf{\Theta}} = \hat{\mathbf{S}} + \hat{\mathbf{N}}. \quad (36)$$

$\hat{\mathbf{\Theta}}$ is the $L \times N$ CS estimation of \mathbf{X} and only contains $\text{card}(\hat{\mathbf{S}})$ non-zero rows. Thus, $\hat{\mathbf{S}}$ and $\hat{\mathbf{N}}$ are also $\text{card}(\hat{\mathbf{S}})$ -sparse. Note that excluding the rows indexed by $\hat{\mathbf{S}}$, all the other rows of $\hat{\mathbf{\Theta}}$ will be set as zero by the CS reconstruction algorithms. Thus, its noise component $\hat{\mathbf{N}}$ in the reconstructed signal $\hat{\mathbf{\Theta}}$ is not white on all the L channels anymore. For convenience, we can remove all the zero rows in (36) and only discuss upon the submatrices $\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}}$, $\hat{\mathbf{S}}_{\hat{\mathcal{S}}}$ and $\hat{\mathbf{N}}_{\hat{\mathcal{S}}}$ corresponding to the non-zero rows in $\hat{\mathbf{\Theta}}$, $\hat{\mathbf{S}}$ and $\hat{\mathbf{N}}$. Similarly we have

$$\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}} = \hat{\mathbf{S}}_{\hat{\mathcal{S}}} + \hat{\mathbf{N}}_{\hat{\mathcal{S}}}. \quad (37)$$

In (37), $\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}}$ is the submatrix composed of the non-zero rows of $\hat{\mathbf{\Theta}}$, namely

$$\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A} \mathbf{X}. \quad (38)$$

By plugging (35) and (37) in (38), the reconstructed signal $\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}}$ is considered in portions of $\hat{\mathbf{S}}_{\hat{\mathcal{S}}}$ and $\hat{\mathbf{N}}_{\hat{\mathcal{S}}}$. For the signal component, we have

$$\hat{\mathbf{S}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A} \mathbf{S}. \quad (39)$$

Because the non-zero elements in \mathbf{S} only exists in the channels occupied by the PUs, (39) can be further deduced as

$$\hat{\mathbf{S}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}} \mathbf{I}_{\hat{\mathcal{S}}} \mathbf{S}, \quad (40)$$

where $\mathbf{I}_{\hat{\mathcal{S}}}$ is an index matrix of size $K \times L$ that picks out the non-zero rows in \mathbf{S} .

When the MP algorithm iteration stops, if $\hat{\mathcal{S}} = \mathcal{S}$, the reconstructed signal

$$\hat{\mathbf{S}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}} \mathbf{I}_{\hat{\mathcal{S}}} \mathbf{S} = \mathbf{I}_{\hat{\mathcal{S}}} \mathbf{S} \quad (41)$$

is the correct estimation of PUs' signal; if $\mathcal{S} \subset \hat{\mathcal{S}}$, namely, the true support is completely included in the estimation by MP, the reconstructed signal can be deduced as

$$\hat{\mathbf{S}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}} \mathbf{I}_{\hat{\mathcal{S}}} \mathbf{S} = \mathbf{I}_{\hat{\mathcal{S}}} \mathbf{S}. \quad (42)$$

In this case, $\hat{\mathbf{S}}_{\hat{\mathcal{S}}}$ is still the correct estimation of \mathbf{S} because the signal components are zero for the unoccupied channels.

Now we move on to the reconstructed noise components $\hat{\mathbf{N}}_{\hat{\mathcal{S}}}$. According to the the previous analysis, $\hat{\mathbf{S}}$ is the correct estimation of \mathbf{S} as long as $\mathcal{S} \subseteq \hat{\mathcal{S}}$. In this case, the noise component in the reconstructed signal can be calculated as

$$\hat{\mathbf{N}}_{\hat{\mathcal{S}}} = (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A} \mathbf{N}. \quad (43)$$

Algorithm 3: DTMP Algorithm.

Input: \mathbf{Y} ; \mathbf{A} (of size $p \times L$); ϵ_R ; σ^2

Output: $\hat{\mathcal{S}}$, $\hat{\mathbf{\Theta}}$

- 1: $t \leftarrow 1$; $\mathbf{R}_0 \leftarrow \mathbf{Y}$; $\hat{\mathcal{S}}_0 \leftarrow \emptyset$; $\mathbf{A}_0 \leftarrow \emptyset$; $E_R \leftarrow \infty$,
 $\hat{\mathbf{\Theta}} \leftarrow \mathbf{0}$
 - 2: **while** $t \leq p \vee E_R > \epsilon_R$ **do**
 - 3: $\hat{\alpha}_t \leftarrow \arg \max_i \|\mathbf{A}_{:,i}^H \mathbf{R}_{t-1}\|$, $i = 0, 1, \dots, L-1$
 - 4: $\hat{\mathcal{S}}_t \leftarrow \hat{\mathcal{S}}_{t-1} \cup \hat{\alpha}_t$
 - 5: $\mathbf{A}_t \leftarrow \mathbf{A}_{:, \hat{\alpha}_t}$
 - 6: $\hat{\mathbf{\Theta}}_t \leftarrow \arg \min_{\Theta} \|\mathbf{Y} - \mathbf{A}_t \Theta\|$
 - 7: $\mathbf{R}_t \leftarrow \mathbf{R}_{t-1} - \mathbf{A}_t \hat{\mathbf{\Theta}}_t$
 - 8: $E_R \leftarrow \|\mathbf{R}_t\|_F^2$
 - 9: $t \leftarrow t + 1$
 - 10: **end while**
 - 11: $\Sigma \leftarrow \sigma^2 L (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1}$
 - 12: $\hat{\mathcal{S}} \leftarrow \{\hat{\mathcal{S}}_i \in \hat{\mathcal{S}}_t | \hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}_i}^H \hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}_i}^H > C \Sigma_{ii}\}$
 - 13: fill the rows of $\hat{\mathbf{\Theta}}$ indexed by $\hat{\mathcal{S}}$ with $\hat{\mathbf{\Theta}}_t$
-

The covariance matrix of $\hat{\mathbf{N}}_{\hat{\mathcal{S}}}$ is calculated as

$$\begin{aligned} \Sigma &= \mathbf{E}[\hat{\mathbf{N}}_{\hat{\mathcal{S}}} \hat{\mathbf{N}}_{\hat{\mathcal{S}}}^H] \\ &= \mathbf{E}[(\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A} \mathbf{N} \mathbf{N}^H \mathbf{A}^H \mathbf{A}_{\hat{\mathcal{S}}} (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1}]. \end{aligned} \quad (44)$$

The original noise \mathbf{N} follows Gaussian distribution as the previous assumption. Thus, we have

$$\mathbf{E}[\mathbf{N} \mathbf{N}^H] = \sigma^2 \mathbf{I}, \quad (45)$$

where σ denotes the standard deviation of the noise, then (44) can be further simplified as

$$\begin{aligned} \Sigma &= \sigma^2 (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A} \mathbf{A}^H \mathbf{A}_{\hat{\mathcal{S}}} (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \\ &= \sigma^2 L (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}} (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1} \\ &= \sigma^2 L (\mathbf{A}_{\hat{\mathcal{S}}}^H \mathbf{A}_{\hat{\mathcal{S}}})^{-1}. \end{aligned} \quad (46)$$

According to (46), the power of the reconstructed noise is still constant in each channel indicated by $\hat{\mathcal{S}}$. The power of $\hat{\mathbf{N}}_{\hat{\mathcal{S}}}$ in a certain channel is indicated by the value on the corresponding position of the diagonal of Σ .

Denoting the power matrix of the reconstructed signal $\hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}}$ as

$$\mathbf{\Pi} = \hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}} \hat{\mathbf{\Theta}}_{\hat{\mathcal{S}}}^H. \quad (47)$$

Based on (47), we apply a secondary filter to the non-orthogonally reconstructed support $\hat{\mathcal{S}}$ in Algorithm 2 by comparing $\mathbf{\Pi}_{ii}$, the power of the reconstructed signal in the channel indicated by the i^{th} element of the support, with its corresponding noise power Σ_{ii} . When $\mathbf{\Pi}_{ii} \leq C \Sigma_{ii}$ (C is a constant hyperparameter), channel $\hat{\mathcal{S}}_i$ is considered to contain only noise and should be removed from the final decision. When $\mathbf{\Pi}_{ii} > C \Sigma_{ii}$, channel $\hat{\mathcal{S}}_i$ is considered as containing PU's signal and should be kept. We hereby propose a double-threshold matching pursuit (DTMP) algorithm, detailed as Algorithm 3.

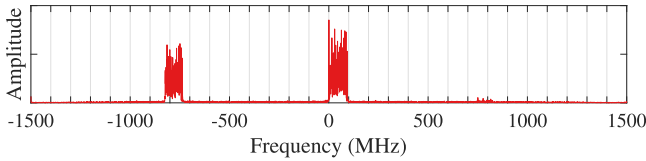


Fig. 6. An example baseband signal spectrum received by the MTS receiver. The whole sensing range is $[-1.536, 1.536]$ GHz, among which $[-850, -750]$ MHz and $[0, 100]$ MHz are occupied. The multicostet sampler evenly divides the spectrum into 30 channels with bandwidth of 102.4 MHz, shown as the grids.

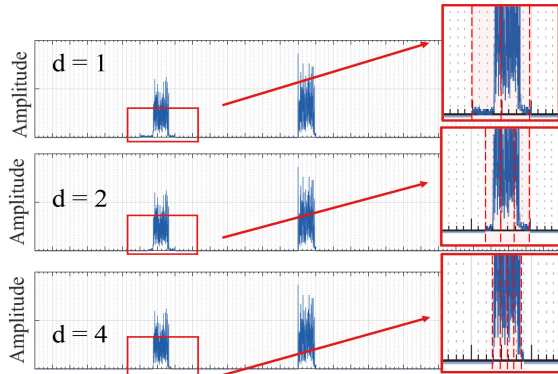


Fig. 7. The reconstruction result of the spectrum in Fig. 6 by SOMP algorithm for different block coefficients d .

VI. NUMERICAL EXPERIMENTS

A. Signal Generation and Sampling Parameters

The signal used for simulation is generated by the National Instruments millimeter wave (mmWave) transceiver system (MTS) [41]. The transmitter can generate up to 8 signals of 100 MHz bandwidth among the $[-1, 1]$ GHz baseband, and transmit on the 28.5 GHz mmWave band. The receiver receives the signal by a mmWave antenna and down-converts the signal to the baseband, then a 3.072GSps Nyquist ADC is used to sample the in-phase and quadrature components. By discarding a certain part of the Nyquist samples, a simulated multicostet sampler is implemented in software.

An example Nyquist spectrum of the received signal in baseband is shown in Fig. 6, of which the number of transmissions $N_{sig} = 2$ with 100 MHz bandwidth. The whole sensing band, from the blind CR receiver aspect, is $[-1.536, 1.536]$ GHz. The frame length of the sampling window is set as 30000 Nyquist samples. The whole spectrum is divided into $L = 30$ channels, namely, the bandwidth for a single multicostet channel is 102.4 MHz, which satisfies the condition $L \leq 1/(BT)$.

The Nyquist samples are first down-sampled by the software multicostet sampler, of which the coset number p is set among $5 \sim 12$ and the sampling pattern C is generated in random. The compressed measurement and sensing matrix is processed by the block method with the block coefficient $d = 1$ (classic) $d = 2$ and $d = 4$, respectively, and the signal is reconstructed by MP and SOMP algorithms.

By comparing the reconstructed signal support $\hat{\mathcal{S}}$ with the ground truth \mathcal{S} , the detection probability P_d and false-alarm probability P_f are defined as

$$P_d = E \left[\frac{|\mathcal{S} \cap \hat{\mathcal{S}}|}{|\hat{\mathcal{S}}|} \right], \quad (48)$$

and

$$P_f = E \left[\frac{|\hat{\mathcal{S}} \setminus \mathcal{S}|}{L - |\mathcal{S}|} \right]. \quad (49)$$

where \setminus denotes the set difference operation.

B. Performance Gain Delivered by Block MMV Model

MP and SOMP algorithms with $d = 1$, $d = 2$ and $d = 4$ are used for the support recovery in the simulation. The relationship between P_d , P_f and the coset number p are tested with channel number $L = 30$, as shown in Fig. 8(a) and (b), respectively. From the result, when $p > 8$, both cases ($d = 1$ and $d = 2$) can guarantee nearly 100% P_d and P_f under 1%. When $p \leq 8$, the case $d = 2$ brings improvement on P_d and reduction on P_f for both MP and SOMP algorithms. From $d = 2$ to $d = 4$, the recovered support is further narrowed to exclude more redundant frequency points that only includes noise. The signal reconstruction is thereby more precise. An example of the recovered spectrum with different d is given in Fig. 7. When $d = 1$, two consecutive 102.4 MHz channels are reconstructed as occupied by the 1st transmission located astride the channel border of 100 MHz bandwidth, causing redundant reconstruction of 104.8 MHz spectrum which only contains noise. With $d = 2$, three sub-channels of width 51.2 MHz are reconstructed and the redundant reconstruction are reduced to 53.6 MHz. With $d = 4$, the 1st transmission band almost fully located in 4 consecutive sub-channels with 25.6 MHz bandwidth and only 2.4 MHz redundant spectrum is reconstructed. Restricted by the finite sampling points in an observing window, d cannot be set too large in our experiment.

By properly increasing d the detection and recovery performance can be improved to a certain extent. Fig. 9(a) and (b) shows the trends of P_d and P_f against d . The performance tends to grow more slowly as d increases from 1 to 6. One of the reasons is that the $\hat{\Omega}$ is converging on Ω . On the other hand, because the number of samples is inversely proportional to d , the reduction of columns in model (23) will weaken the noise resistance during the support recovery process. Thus, choosing an appropriate d before the reconstruction to balance the detection performance improvement and computing efficiency is significant.

From Fig. 6, the signal support $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2\}$ is distributed in 3 channels, where \mathcal{T}_1 is divided to channel 7 and 8, and \mathcal{T}_2 is all contained in channel 16. Thus, when we set $d = 1$, the row sparsity of \mathbf{X} is $K = 3$, and the subsequent algorithm output proves that as well. According to proposition 1, the signal should be perfectly reconstructed with $p \geq 6$. However, from Fig. 8(a) and (b) we can tell that the P_d does not approaches 100% until $p > 8$. The reason for the gap is that $p = 6$ is too

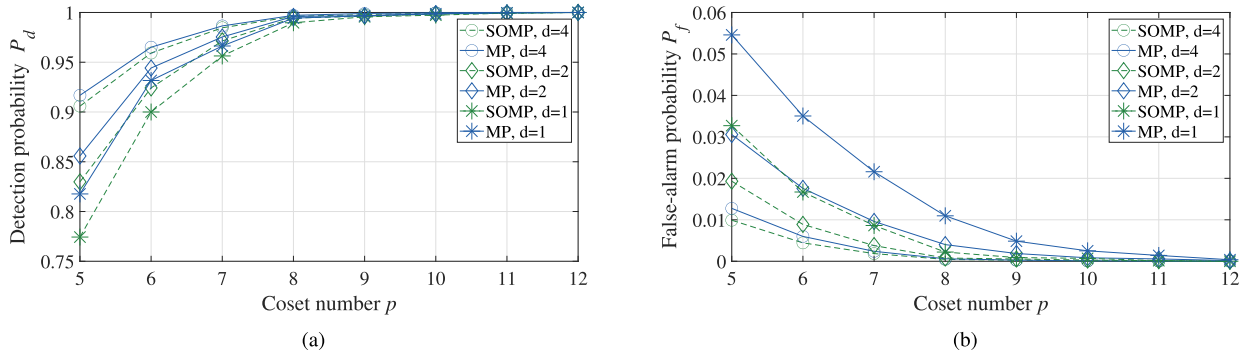


Fig. 8. P_d (a) and P_f (b) of MP and SOMP against d from 1 to 6 with $p = 6, p = 8$ and $p = 10$ ($L = 30$).

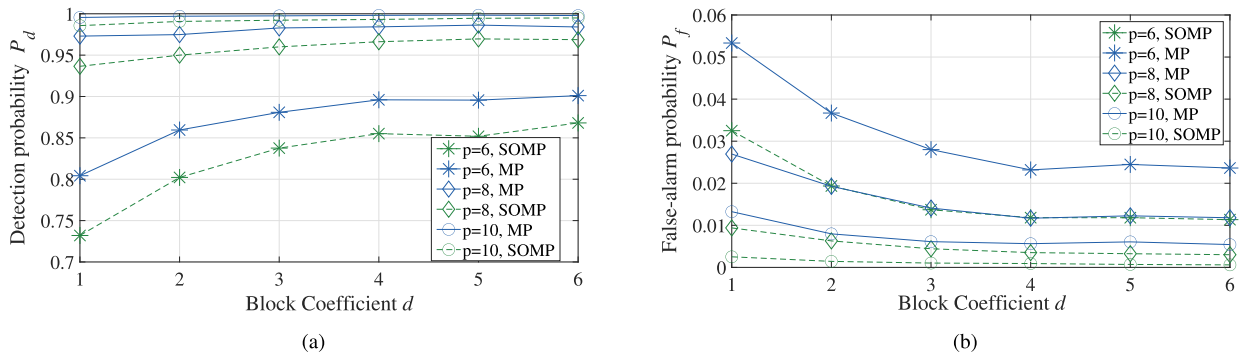


Fig. 9. P_d (a) and P_f (b) of MP and SOMP against coset number p from 5 to 14 with $d = 1, d = 2$ and $d = 4$ ($L = 30$).

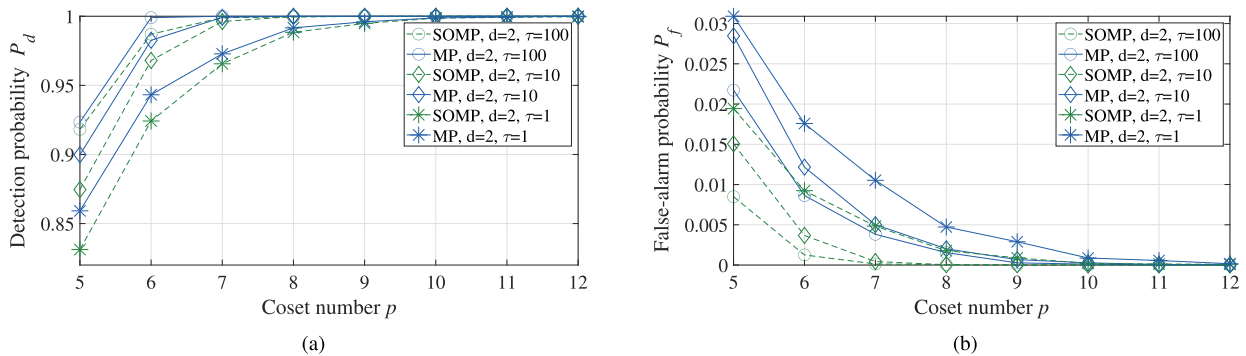


Fig. 10. P_d (a) and P_f (b) of MP and SOMP against coset number p from 5 to 14 with pattern selection iteration number $\tau = 1, \tau = 10$ and $\tau = 100$ ($L = 30, d = 2$).

small to ensure a relatively low column correlation of matrix \mathbf{A} . By applying pattern selection before multicohort sampling, $\mu(\mathbf{A})$ can be effectively reduced, then P_d can be improved.

C. Performance Gain Delivered by Pattern Selection Algorithm

Fig. 10 shows the detection performance and false-alarm probability using MP and SOMP algorithms with random patterns ($\tau = 1$) and optimized patterns selected by Algorithm 1 with $\tau = 10$ and $\tau = 100$. The average detection probability of both algorithms shows an evident increase and the average false-alarm drops with the increase of τ . That is because the

optimized pattern generates a sensing matrix \mathbf{A} with smaller maximum column correlation coefficient $\mu(\mathbf{A})$. The reconstruction performance can be further improved by continuously increasing τ , but the gain on performance converges to a certain level according to Section IV. The trade-off between recovery accuracy and computational efficiency should be considered before the application stage.

D. Performance Gain Delivered by DTMP Algorithm

The proposed DTMP algorithm is also tested with the MTS data, and the result is compared with those by MP and SOMP, shown in Fig. 11. In the recovery process, both the block

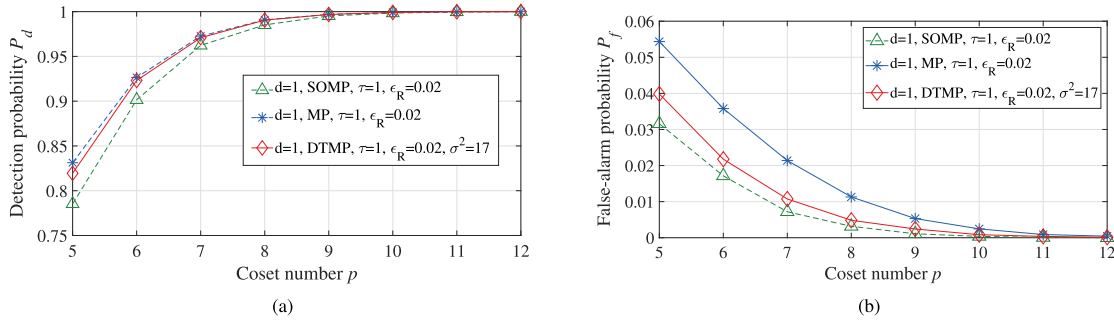


Fig. 11. P_d (a) and P_f (b) of MP, SOMP and proposed DTMP against coset number p from 5 to 16 with same $\epsilon_R = 0.02$ ($L = 30$), for DTMP $\sigma^2 = 16.8$.

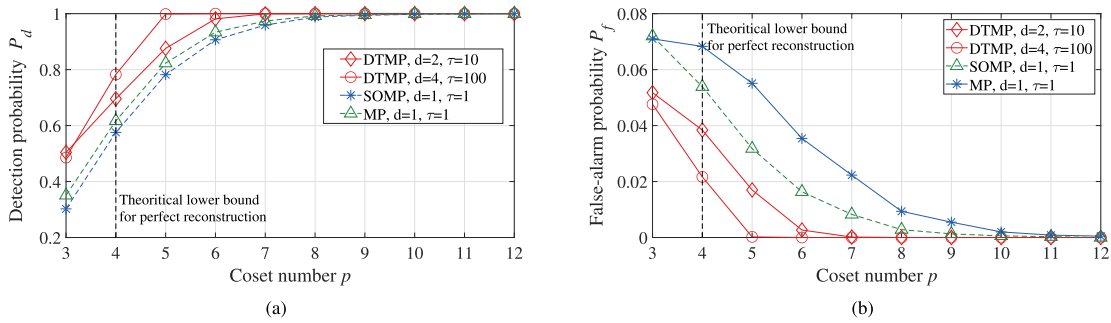


Fig. 12. P_d (a) and P_f (b) of DTMP with and pattern selection on block MMV model compared with those by MP and SOMP without pattern selection on traditional MMV model against coset number p from 3 to 12. ($L = 30$).

coefficient b and the pattern iteration number τ are set as 1 for all three algorithms. The stop condition of the iteration for both Algorithms 2 and 3 are set as a small value $\epsilon_R = 0.02$ to ensure that $\mathcal{S} \subset \mathcal{S}_i$ after all iterations. That is to say that the MP and the DTMP will stop their iteration with the same \mathcal{S}_i but DTMP will apply a secondary screening to obtain the final $\hat{\mathcal{S}}$. The second threshold of DTMP is set as $\sigma^2 = 16.8$ according to the noise estimation of the radio environment ($\sigma^2 = 16.8$ can be approached by detecting a certain band that is known as idle). From Fig. 11 we can tell that the P_d of DTMP is as high as that of MP, while the P_f shows evident decrease after the secondary screening. Given a small ϵ_R and an appropriate $\sigma^2 = 16.8$, MP and DTMP share identical P_d higher than that by SOMP. This is because non-orthogonal algorithms can better distinguish two occupied channels corresponding to the two columns in \mathbf{A} with relatively high correlation. The P_f of DTMP is a little higher than that of SOMP and lower than that of MP, which indicates the redundant noise components selected by mistake in the preliminary result \mathcal{S}_i have been largely filtered out during the secondary screening in DTMP. Using the filtered support for spectrum reconstruction will bring a more accurate signal estimation.

E. Integrated Optimization Based on the Proposed Scheme

The above-mentioned three aspects of optimization, namely the block MMV model, pattern selection algorithm and DTMP algorithm can be used both individually and integrally because each of them acts on a different step of compressed spectrum

sensing. By integrating the three optimization methods together, the recovery performance is tested with a practical and reasonable range of parameters (namely $d \leq 4$ and $\tau \leq 1000$). The result of P_d and P_f are shown in Fig. 12. Compared with the result produced by SOMP and MP without pattern selection on the traditional MMV model, the DTMP algorithm with pattern selection on the block MMV model shows a great improvement in reconstruction performance. For ease of view, the range of coset number p is set as 3~12 to include the theoretical lower bound $p = 4$ for the test data. The SOMP and MP algorithms with traditional setting $d = 1$ and $\tau = 1$ achieve perfect support reconstruction (the detection probability using DTMP reaches nearly 100% and the false-alarm reaches nearly 0%) at $p \geq 10$ and $p \geq 12$, respectively, which are higher than 200% of the theoretical boundary. The DTMP algorithm, together with block MMV model $d = 2$ and pattern selection $\tau = 10$, achieves perfect support reconstruction at $p = 7$ (160% sampling density of the theoretical boundary). With $d = 4$ and $\tau = 100$, which only lightly increase the computation resources, DTMP achieves perfect support reconstruction at only $p = 5$ (120% sampling density of the theoretical boundary). It is certain that the theoretical boundary can be further approached by continuously increasing d and τ , by taking the above-mentioned accuracy-efficiency trade-off into consideration.

VII. CONCLUSION

This paper has proposed an optimization framework for achieving perfect blind multiband signal reconstruction based on

the multicoset sampling architecture with an average sampling rate closer to the theoretical boundary. In this framework, three factors that affect the performance of spectrum reconstruction have been analyzed and optimized, e.g., the MMV model for joint sparse data, the selection of sampling delay pattern, and the accuracy of the reconstruction algorithm.

A block MMV data model is proposed for the first time. In this model, an original multicoset channel is divided into multiple narrower blocks. The support set of the spectrum can be reconstructed more accurately. We proved that the block MMV model can reduce the minimum sampling rate required for a multicoset sampler. Through numerical experiments, we have verified that block MMV can achieve better reconstruction performance compared with the traditional MMV model with the same sampling rate. The block coefficient can be adjusted according to the actual situation to seek a balance between improving reconstruction performance and computational cost.

The delay pattern of the multicoset sampler can affect the performance of the sensing matrix. We have proposed a delay pattern selection algorithm to reduce the column correlation of the sensing matrix. Compared with a randomly selected pattern, the pattern selected by the proposed algorithm yields a better sensing matrix with great probability, thereby improving the spectrum reconstruction performance.

This paper has also derived the representation method of the signal and Gaussian noise in the reconstructed signal based on the compressed sensing theory and thus has established a double-threshold matching pursuit algorithm that can perform secondary screening on the reconstructed support set to eliminate noise components. Numerical experiments have proven that this algorithm can achieve a lower false alarm rate than a non-orthogonal algorithm.

The three optimization methods can be applied integrally. With integrated optimization, the sampling resources required for perfect reconstruction have been proved to be reduced from above 200% to 120% or less with computational-friendly optimization parameters.

APPENDIX A PROOF OF PROPOSITION 1

We first consider problem (15) as N individual SMV problems, stated as

$$\mathbf{Y}_{:,n} = \mathbf{A}\mathbf{X}_{:,n}, \quad 1 \leq n \leq N. \quad (50)$$

Assuming a full row-rank \mathbf{A} , for $K \geq L/2$, $p \geq L$ measurements can guarantee the unique solution of $\mathbf{X}_{:,n}$, regardless of its sparsity.

For $K < L/2$, $\mathbf{X}_{:,n}$ is K -sparse at the most since \mathbf{X} has totally K non-zero rows. According to the well-known theorem in CS theory [42], the number of measurements to reconstruct $\mathbf{X}_{:,n}$ satisfies

$$p \geq 2K. \quad (51)$$

QED.

APPENDIX B PROOF OF PROPOSITION 2

We first deduce the IMV form (11) of model (21) as

$$\mathbf{y}_i^d(f) = \mathbf{A}\mathbf{x}_i^d(f), \quad f \in \left[\frac{i-1}{dLT}, \frac{i}{dLT} \right), \quad i \in \{1, 2, \dots, d\}, \quad (52)$$

where $\mathbf{y}_i^d(f)$ and $\mathbf{x}_i^d(f)$ are defined in the same way as $\mathbf{y}(f)$ and $\mathbf{x}(f)$ in (12) and (13) except for the boundary of f . Specifically, we emphasize

$$\mathbf{x}_i^d(f) = \begin{pmatrix} X(f + \frac{-L/2}{LT}) \\ X(f + \frac{-L/2+1}{LT}) \\ \vdots \\ X(f + \frac{L/2-1}{LT}) \end{pmatrix}. \quad (53)$$

When $d \rightarrow \infty$, we have

$$f_i \triangleq \lim_{d \rightarrow \infty} \frac{i-1}{dLT} = \lim_{d \rightarrow \infty} \frac{i}{dLT}.$$

Model (52) becomes a SMV model where

$$\mathbf{x}_i^\infty = \begin{pmatrix} X(f_i + \frac{-L/2}{LT}) \\ X(f_i + \frac{-L/2+1}{LT}) \\ \vdots \\ X(f_i + \frac{L/2-1}{LT}) \end{pmatrix}. \quad (54)$$

In other word, when $d \rightarrow \infty$ we got \mathbf{x}_i^∞ as a set of discrete samples on the original spectrum $X(f)$, where the sampling interval is $1/(LT)$. Since $L \leq 1/(BT)$, the interval $1/(LT) \geq B$. Thus, at most N_{sig} elements in \mathbf{x}_i^∞ occupied by singles. QED.

REFERENCES

- [1] in Spectrum needs for the terrestrial component of IMT in the frequency range between 24.25 GHz and 86 GHz, 2016-09-22, ITU-R WP5D Contribution 259.
- [2] R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *Proc. Int. Conf. Wireless Netw., Commun. Mobile Comput.*, 2005, pp. 464–469.
- [3] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.
- [4] R. Liu, Y. Ma, X. Zhang, and Y. Gao, "Deep learning-based spectrum sensing in space-air-ground integrated networks," *J. Commun. Inf. Netw.*, vol. 6, no. 1, pp. 82–90, 2021.
- [5] Z. Song, Y. Gao, and R. Tafazolli, "A survey on spectrum sensing and learning technologies for 6G," *IEICE Trans. Commun.*, vol. E104.B, pp. 1207–1216, 2021.
- [6] D. Cohen, S. Tsiper, and Y. C. Eldar, "Analog-to-digital cognitive radio: Sampling, detection, and hardware," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 137–166, Jan. 2018.
- [7] R. Venkataramani and Y. Bresler, "Optimal sub-nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, Oct. 2001.
- [8] T. Moon, H. W. Choi, N. Tzou, and A. Chatterjee, "Wideband sparse signal acquisition with dual-rate time-interleaved undersampling hardware and multicoset signal reconstruction algorithms," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6486–6497, Dec. 2015.
- [9] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.
- [10] C.-P. Yen, Y. Tsai, and X. Wang, "Wideband spectrum sensing based on sub-nyquist sampling," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3028–3040, Jun. 2013.

- [11] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, 1967.
- [12] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [13] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
- [14] Y. Gao et al., "Sub-nyquist spectrum sensing and learning challenge," *Front. Comput. Sci.*, vol. 15, no. 4, pp. 1–5, 2021.
- [15] J. Yang, Z. Song, Y. Gao, X. Gu, and Z. Feng, "Adaptive compressed spectrum sensing for multiband signals," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7642–7654, Nov. 2021.
- [16] H. Zhang, J. Yang, and Y. Gao, "Machine learning empowered spectrum sensing under a sub-sampling framework," *IEEE Trans. Wireless Commun.*, early access, doi: [10.1109/TWC.2022.3164800](https://doi.org/10.1109/TWC.2022.3164800).
- [17] M. Shekaramiz, T. K. Moon, and J. H. Gunther, "On the block-sparsity of multiple-measurement vectors," in *Proc. IEEE Signal Process. Signal Process. Educ. Workshop*, 2015, pp. 220–225.
- [18] M. Shekaramiz, T. K. Moon, and J. H. Gunther, "Hierarchical Bayesian approach for jointly-sparse solution of multiple-measurement vectors," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, 2014, pp. 1962–1966.
- [19] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [20] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-nyquist rates," *Circuits Devices Syst. Lett.*, vol. 5, no. 1, pp. 8–20, 2009.
- [21] H. Kwon and B. D. Rao, "On the benefits of the block-sparsity structure in sparse signal recovery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 3685–3688.
- [22] M. Fornasier and H. Rauhut, "Compressive sensing," *Handbook of Math. Methods Imag.*, vol. 1, pp. 187–229, 2015.
- [23] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [24] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [25] M. E. Domínguez-Jiménez, N. González-Prelcic, G. Vazquez-Vilar, and R. López-Valcarce, "Design of universal multicostet sampling patterns for compressed sensing of multiband sparse signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 3337–3340.
- [26] H. B. Celebi, L. Durak-Ata, and H. Celebi, "Multi-coreset sampling and reconstruction of signals: Exploiting sparsity in spectrum monitoring," in *Proc. 21st Eur. Signal Process. Conf.*, 2013, pp. 1–5.
- [27] B. Tausiesakul and N. González-Prelcic, "Power spectrum blind sampling using optimal multicostet sampling patterns in the MSE sense," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 1055–1059.
- [28] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.: A J. Issued by Courant Inst. Math. Sci.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [29] S. Liu, Y. D. Zhang, T. Shan, and R. Tao, "Structure-aware bayesian compressive sensing for frequency-hopping spectrum estimation with missing observations," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2153–2166, Apr. 2018.
- [30] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [31] H. Qi, X. Zhang, and Y. Gao, "Low-complexity subspace-aided compressive spectrum sensing over wideband whitespace," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11762–11777, Dec. 2019.
- [32] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [33] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. v/721–v/724.
- [34] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [35] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [36] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [37] X. Zhang, Y. Ma, Y. Gao, and W. Zhang, "Autonomous compressive-sensing-augmented spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6970–6980, Aug. 2018.
- [38] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [39] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [40] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
- [41] Z. Song, H. Qi, and Y. Gao, "Real-time multi-gigahertz sub-nyquist spectrum sensing system for mmWave," in *Proc. 3rd ACM Workshop Millimeter-wave Netw. Sens. Syst.*, 2019, pp. 33–38.
- [42] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.



Zihang Song (Graduate Student Member, IEEE) received the bachelor's and master's degrees in applied physics from Beihang University, Beijing, China. He started the Ph.D. degree with the Queen Mary University of London, London, U.K., in 2019. He is currently working with Prof. Yue Gao and Prof. Rahim Tafazolli, Institute for Communication Systems, University of Surrey, Guildford, U.K. His research interests include millimeter-wave spectrum sensing and sub-Nyquist signal processing.



Jian Yang (Student Member, IEEE) received the B.S. degree in electronic information engineering under-water acoustics from Harbin Engineering University, Harbin, China, in 2015. He is currently working toward the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China. His research interests include cognitive radio, compressed sensing, and sub-Nyquist signal processing.



Han Zhang (Member, IEEE) received the Ph.D. degree in electrical and electronics engineering from the University of California, California, CA, USA, in 2019. He is currently working as a Research Assistant with the University of Surrey, Guildford, U.K. His research interests include utilizing data-driven methods in telecommunication scenarios, such as transceiver design, and compressive sensing.



Yue Gao (Senior Member, IEEE) received the Ph.D. degree from the Queen Mary University of London, London, U.K., in 2007. He is currently a Professor and Chair of wireless communications with the School of Computer Science and Electronic Engineering, University of Surrey, Guildford, U.K. He also leads Antennas and Signal Processing Lab developing fundamental research into practice of the interdisciplinary area of smart antennas, signal processing, spectrum sharing, millimetre-wave, and Internet of Things technologies in mobile and satellite systems.

He has authored or coauthored more than 200 peer-reviewed journal and conference papers, one book and five book chapters. He was the co-recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and elected as an Engineering and Physical Sciences Research Council Fellow in 2017. He is also a Member of the Board of Governors and Distinguished Lecturer of the IEEE Vehicular Technology Society (VTS), Vice-Chair of the IEEE ComSoc Wireless Communications Technical Committee, past Chair of the IEEE ComSoc Technical Committee on Cognitive Networks. He is an Editor of several IEEE Transactions and Journals, and Symposia Chair, Track Chair, and other roles in the organising committee of several IEEE ComSoc, VTS, and other conferences.