

Exploiting Self-Similarity for Under-Determined MIMO Modulation Recognition

Wei Xiong, Lin Zhang, Maxwell McNeil, Petko Bogdanov, Mariya Zheleva

Department of Computer Science, University at Albany – SUNY

{wxiong, lzhang22, mmcneil2, pbogdanov, mzheleva}@albany.edu

Abstract—Modulation recognition (modrec) is an essential functional component of future wireless networks with critical applications in dynamic spectrum access. While predominantly studied in single-input single-output (SISO) systems, practical modrec for multiple-input multiple-output (MIMO) communications requires more research attention. Existing MIMO modrec impose stringent requirements of fully- or over-determined sensing front-end, i.e. the number of sensor antennas should exceed that at the transmitter. This poses a prohibitive sensor cost even for simple 2x2 MIMO systems and will severely hamper progress in flexible spectrum access with advanced higher-order MIMO.

We design a MIMO modrec framework that enables efficient and cost-effective modulation classification for under-determined settings characterized by fewer sensor antennas than those used for transmission. Our key idea is to exploit the inherent multi-scale self-similarity of MIMO modulation IQ constellations, which persists in under-determined settings. Our framework called **SYMMeTRY** (Self-similarITy for MIMO ModulaTion Recognition) designs domain-aware classification features with high discriminative potential by summarizing regularities of symbol co-location in the MIMO constellation. To this end, we summarize the fractal geometry of observed samples to extract discriminative features for supervised MIMO modrec. We evaluate **SYMMeTRY** in a realistic simulation and in a small-scale MIMO testbed. We demonstrate that it maintains high and consistent performance across various noise regimes, channel fading conditions and with increasing MIMO transmitter complexity. Our efforts highlight **SYMMeTRY**'s high potential to enable efficient and practical MIMO modrec.

I. INTRODUCTION

Dynamic Spectrum Access (DSA) is projected as a key capability in 5G mobile networks, seeking to address the shortage of radio resources by opportunistic reuse of frequencies. While DSA promises improved network performance, it hinges on robust and affordable measurement capabilities in support of spectrum technology, policy and enforcement. This has brought the problem of modulation recognition (modrec) in the spotlight of research [5], [30]. The goal of a modrec algorithm is to automatically classify a transmitter's modulation. While traditionally it has been tackled in the SISO context [2], [6], [10], [11], [20], [26], [35], [41], the ubiquity of MIMO technology requires the design of robust and cost-efficient MIMO modrec with practical applicability to emerging spectrum sensing platforms [1], [22], [34].

The problem of MIMO modrec is more challenging than the SISO case since (i) the number of symbols in the IQ constellation grows exponentially with the number of transmit antennas and (ii) the channel state parameters grow quadratically with the number of transmit/receive antennas. To overcome these

	Cost, (\$)			Data, (GB)		
	U	F	O	U	F	O
SISO	1,216	1,216	2,432	10	10	20
MIMO 2x2	1,216	2,426	6,791	10	20	40
MIMO 4x4	1,216	6,791	11,655	10	40	80

TABLE I: Monetary cost and data footprint of a 10-second MIMO trace with under- (U), fully- (F) and over-determined (O) sensing using a USRP B210 at a sampling rate of 1MSps and an Octoclock-G.

challenges, MIMO modrec approaches require the sensing infrastructure to be either fully- or over-determined, that is the number of antennas on the sensor should be equal or double that of the target transmitter [4], [16], [24], [44]. These requirements pose prohibitive cost for an individual sensor and present a major road-block to affordable and ubiquitous spectrum sensing. In addition, fully- and over-determined sensing generates large volumes of spectrum data posing high bandwidth, storage and computation requirements and further hampering spectrum analytics at scale.

To put this into perspective, Tbl. I presents a breakdown of the cost and data footprint of spectrum sensing when using a USRP B210 sensor at \$1,216 per board, collecting a short 10-second scan at 1MSps. We note that a USRP B210 can support up to 2x2 MIMO operation with its embedded capabilities and assume the use of an Octoclock-G CDA-2990 at \$1,927 for higher order MIMO setups. It is evident that both the monetary cost and the data footprint scale super-linearly with the MIMO complexity. For the worst case of over-determined sensing of a 4x4 MIMO setup the cost for a single sensor is \$11,655 generating 80GB of data for 10 seconds. These costs are just for the radio hardware and would further be amplified if the host computer is accounted for.

In order to address this prohibitive cost we propose to utilize *under-determined spectrum sensing*, whereby a multi-antenna transmitter is scanned by a sensor with fewer than the transmitter's antennas. While this has the potential to keep the monetary and data cost of MIMO sensing in check, it raises fundamental challenges for characterization. Existing MIMO modrec [4], [16], [24], [44] requires fully- or over-determined sensing as it typically relies on invariant statistical properties of the sample constellations. Thus, standard MIMO modrec cannot be readily-applied in the under-determined scenario. To address this, we propose a MIMO modrec framework called **SYMMeTRY** (Self-similarITy for MIMO ModulaTion Recognition) that enables robust recognition for under-determined sensing. **SYMMeTRY** explores the

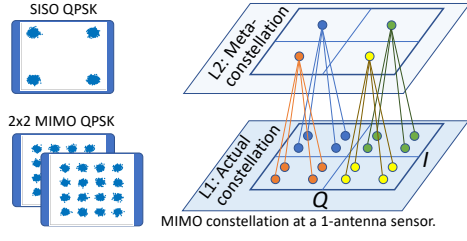


Fig. 1: Illustration of MIMO self-similarity for 2x2 MIMO transmission with QPSK. self-similar geometric patterns in the MIMO constellation to extract domain-informed features for classification. The key insight behind SYMMeTRY is that the MIMO constellation exhibits self-similarity at different scales which is well-preserved in under-determined settings and can be employed for modrec. Fig. 1 illustrates this concept of self-similarity via an example 2x2 QPSK modulated signal. A SISO QPSK constellation (top left corner of Fig. 1) contains a group of four clusters, one for each of the QPSK symbols. In the 2x2 MIMO QPSK case (bottom left), four new QPSK symbol groups “hatch” around each of the existing SISO positions. Therefore, for 2x2 QPSK MIMO there are at most 4^2 constellation symbols. Such templated replication behavior leads to a fractal-like self-similar organization of the MIMO constellation corresponding to a given modulation. The 2x2 QPSK MIMO constellation can, thus, be thought of as a 2-tier hierarchical representation (Fig. 1(right)), whereby at the first level, we observe the actual constellation composed of four SISO QPSK groups, while at the second level, we observe a meta-constellation that is determined by the centroids of each group at layer 1 and resembles a scaled version of a SISO QPSK constellation. This self-similar multi-scale organization can be observed for higher order modulations, whereby, the number of meta-constellation levels increases to $\log(M) - 1$ for increasing modulation order up to M . Furthermore, this self-similarity persists even when a scan is under-determined and is at the center of our design of modulation classification features. To capture the relative co-location of symbols within the constellation we extend tools from fractal geometry [9], [19]. We combine features based on the above patterns with higher order cumulants [12], [35] in our overall feature design.

Our paper makes the following key contributions:

- We conceptualize under-determined MIMO spectrum sensing and modulation classification.
- We are the first to formalize and utilize the self-similarity of MIMO constellations into an adaptive framework called SYMMeTRY, for robust feature design in support of under-determined MIMO modrec.
- We investigate the effects of the MIMO channel on the constellation self-similarity and recognition accuracy.
- Using simulation and a testbed, we show that SYMMeTRY maintains high and consistent performance across various channel conditions and with increasing MIMO complexity.

II. BACKGROUND

In this section we introduce the general MIMO signal model and under-determined MIMO sensing. We detail the geometric

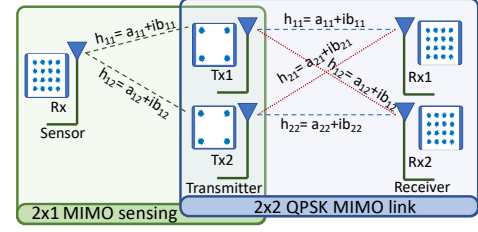


Fig. 2: System model for under-determined sensing.

self-similarity of MIMO constellations to provide intuition behind our methodology. We also analyze the relationship between MIMO channel properties such as (i) signal to noise ratio (SNR), (ii) coherence and (iii) pairwise antenna gain, and the geometric properties of the observed constellation.

A. MIMO signal and system models

The MIMO system is composed of N_t transmitting antennas and N_r receiving antennas. Fig. 2 (right) illustrates an example of a 2x2 MIMO link (i.e. $N_r = 2$ and $N_t = 2$) using QPSK modulation. At any given instant, the signals $r(n) \in \mathbb{C}^{N_r}$ on the receiving side can be represented as a linear combination of the transmitted baseband signals $s(n) \in \mathbb{C}^{N_t}$ and instantaneous additive Gaussian noise $w(n) \in \mathbb{C}^{N_r}$:

$$r(n) = H s(n) + w(n), \quad (1)$$

where $H = [h_{ij}] \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix specifying the pairwise channel response h_{ij} for each transmit-receive antenna pair (i, j) . The channel noise is modeled as sampled from a 0-mean normal distribution $w(n) \in \mathcal{N}(0, \sigma^2 I_{N_r})$ with 0 co-variance terms, i.e., $E[w(n)w(n)^H] = \sigma^2 I_{N_r}$, where I_{N_r} is the identity matrix of size N_r and $w(n)^H$ denotes the conjugate transpose of the row vector $w(n)$.

Depending on how application data is divided over the individual transmitter streams N_t and how each stream is modulated, there are several different ways to realize a MIMO transmitter [31]. In terms of application data assignment to transmit streams we differentiate between (i) MIMO with *spatial diversity* and (ii) MIMO with *spatial multiplexing*. The former passes redundant application data on each transmitter antenna allowing the receiver to decode the minimal-error stream. This redundancy leads to lower throughput in comparison with spatial multiplexing, but enables high probability of successful data decoding at the receiver. In *spatial multiplexing* MIMO each transmitter stream handles a unique portion of the application data resulting in increased throughput at the cost of higher sensitivity to poor channel conditions.

We also differentiate between *direct-mapped* and *precoded* MIMO links, based on the power and modulation assignment approach across transmitter streams. The former allocates the same power level and modulation across transmitter streams without considering the channel conditions, while the latter performs channel estimation and adaptively assigns power and modulation to each stream to maximize throughput. Thus, the choice between direct-mapping and precoding offers a trade-off between throughput and implementation complexity.

In this paper, we focus on MIMO with spatial multiplexing, which has a greater potential for improved wireless network throughput compared to spatial diversity. We begin our exploration on a system with direct-mapping and leave modrec of MIMO with precoding for future work.

B. Limitations of existing MIMO modrec approaches

Modulation recognition can be viewed as a classification task: given a set of N IQ samples (an “instance”) the goal is to determine the modulation from which the observations are sampled. Classifiers used in such tasks are trained in a supervised manner, i.e. they require annotated instances in order to learn to recognize modulations (called classes). Most existing MIMO modrec algorithms require prior knowledge of the channel conditions [4], [14], [16], [24], [27]. In the context of the MIMO signal model (Eq. (1)), this means that existing approaches require prior knowledge of the channel response matrix H . Once this information is available, the paths are considered individually and common features (e.g. cumulants) are employed for classification. There are two critical limitations that hamper the applicability of these approaches in under-determined MIMO modrec: (i) cumulants dispersion, which deteriorates their discriminative power and (ii) inability for channel estimation, as H is under-determined.

Cumulants dispersion. Fig. 3 illustrates the effects of under-determined sensing on state-of-the-art MIMO modrec that uses cumulants [25]. We consider 200 instances of QPSK and 200 of 16-QAM, and for each instance we calculate a feature vector comprised of the seven cumulants typically used in prior work: $[C_{40}, C_{41}, C_{42}, C_{60}, C_{61}, C_{62}, C_{63}]$. We then adopt principle component analysis (PCA) [15] to reduce the dimensionality of instances to two dimensions and plot them in Fig. 3 for increasing determination of the sensing system. The figure presents scatter plots of these projections for QPSK (blue) and 16-QAM (red) and the respective density “iso”-lines for each of the classes. From right to left we have over-, fully- and under-determined signal. *The increasing overlap of the classes demonstrates qualitatively that the discriminative power of cumulants deteriorates in under-determined settings.* We further quantify the effect of under-determined sensing on modrec accuracy in the experimental section (Sec. IV).

Inability for channel estimation. If the channel information is not apriori known, which often occurs in a non-cooperative cognitive radio sensing, existing modrec algorithms adopt an additional step to estimate H via Independent Component Analysis (ICA) [4] or Expectation Maximization (EM) [44]. Such approaches, however, *require over-determined sensing*, and thus, impose prohibitive cost on the number of receiving antennas. Furthermore, these methods are *not directly applicable with under-determined MIMO modrec*, as they would trigger channel estimation with fewer receiving than transmitting antennas, which is unfeasible.

C. Under-determined MIMO sensing

To alleviate the problem of high cost, we consider under-determined MIMO sensing, whereby the number of receiver

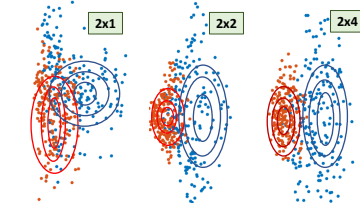


Fig. 3: Discriminative power of cumulants extracted from under- (2x1), fully- (2x2) and over-determined (2x4) sensing. Red presents 16-QAM, while blue, QPSK. The discriminative power of cumulants deteriorates with under-determined sensing.

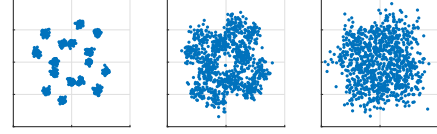


Fig. 4: 2x1 QPSK constellations at SNR 20dB (left), 10 (middle) and 5dB (right). As SNR decreases the geometry of the constellation is less-pronounced.

antennas is lower than that of the transmitter. Fig. 2 (left) depicts schematically the system model for under-determined MIMO sensing, where the example 2x2 MIMO link is sensed with a single-antenna sensor. We call this 2x1 sensing. Intuitively, a 2x1 sensing of a 2x2 MIMO system will gain a single observation of the MIMO constellation.

Our signal model for under-determined sensing follows the same definition as in Eq. (1), where $H = [h_{ij}] \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix with $i \in [1, N_r]$, $j \in [1, N_t]$ and $N_r < N_t$. For example, as illustrated in Fig. 2 (left) for the 2x1 sensing case, the H matrix is a 2x1 vector of the form $H = [h_{11} h_{12}]$, where each h_{ij} is a complex number: $h_{ij} = a_{ij} + ib_{ij}$,

A MIMO constellation is comprised of $2^{M \times N_t}$ individual symbols, where M is the modulation order (e.g. $M=2$ for QPSK, $M=3$ for 8-PSK, etc.) and N_t is the number of transmitter antennas. Since we will use the geometry of the constellation as a predictive “fingerprint” of the modulation, the spread and overlap of symbols and the tightness of a symbol’s cluster will play an important role. The spread and overlap of constellation symbols depend on the channel gain and fading, whereas the tightness of individual clusters depends on the SNR. In what follows, we explore each of these in the context of our signal model and illustrate their effects on the constellation geometry and self-similarity.

1) *Impact of noise on the constellation geometry:* The SNR of a MIMO channel affects how tightly-clustered are the IQ samples around each constellation symbol. The lower the SNR, the more dispersed the samples and the harder it is to recognize the modulation or demodulate the signal. This effect is illustrated in Fig. 4 which depicts the constellation shape of 2x1 sensing of a QPSK signal at an SNR of 20dB (left), 10dB (middle) and 5dB (right).

2) *Impact of channel gain on constellation geometry:* The gain $|h_{ij}|$ of a MIMO path (i, j) is defined as:

$$|h_{ij}| = \sqrt{a_{ij}^2 + b_{ij}^2}, \quad (2)$$

where a_{ij} and b_{ij} are the channel coefficients. We introduce the *channel gain sum* (G_S) and the *channel gain ratio* (G_R)

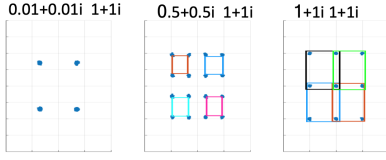


Fig. 5: Effects of G_R^p on the MIMO constellation geometry.

which jointly control the MIMO constellation shape, where

$$G_S = \sum_{ij} |h_{ij}|, \quad (3)$$

controls the *spread* of the MIMO constellation symbols. High G_S allows easier MIMO decoding and modrec. The channel gain ratio G_R , defined as:

$$G_R^p = \frac{\mathcal{L}(p)}{\mathcal{L}(p+1)} = \frac{|h_{ij}^{p-1}|}{|h_{ij}^p|}, 0 \leq G_R^p \leq 1, \quad (4)$$

controls the *overlap* of constellation symbols. Here the individual paths $|h_{ij}|$ are considered in decreasing order $\mathcal{L} = [|h_{ij}^p|]$, where $p \in [1, P-1]$ and $P = N_t \times N_r$ is the total number of individual paths (i, j) .

To put the notion of G_R^p into context, let us consider our running example of 2x1 sensing from Fig. 2. Fig. 5 demonstrates the effects of G_R^p on the geometry of the constellation. In the left of the figure, we see an example of 2x1 QPSK sensing with $G_R^p = 0.1$. This means that the channel gain $|h_{12}|$ of the second link is much smaller than that of the first link. As a result the centroids of each sub-constellation are far apart (due to the high gain on the first link $|h_{11}|$), whereas the symbols in each sub-constellation are very close (due to the low gain on the second link $|h_{12}|$). Thus, the geometry of the overall constellation following 2x1 sensing of a MIMO channel with $G_R^p = 0.1$ resembles a QPSK SISO constellation. Next, we consider the constellation of 2x1 sensing with $G_R^p = 0.5$, which is illustrated in the middle pane of Fig. 5. $G_R^p = 0.5$ means that the gain of the stronger channel is twice that of the weaker (i.e. $|h_{11}| = 2 \times |h_{12}|$). For this setting, we get a “fully-unfolded” MIMO constellation with the maximum of $2^{M \times N_t}$ non-overlapping symbols, which in the case of 2x1 sensing is $2^{2 \times 2} = 16$. Finally, for $G_R^p = 1$, $|h_{11}| = |h_{12}|$, the distance between sub-constellation symbols is the same as that between the sub-constellation centroids. As a result, some of the symbols in the sub-constellations overlap, as illustrated in the right pane of Fig. 5. This results in a partially-unfolded MIMO constellation with a total of nine non-overlapping symbols. Based on this qualitative analysis, we expect that the channel sum and ratio will affect the informativeness of geometric features for modrec; a hypothesis we confirm experimentally in §IV.

3) *Impact of channel fading on under-determined modrec:* Channel fading is a random process that models the change of the channel gain over time. In the context of modrec, of specific interest are the *channel model* and the *coherence time*, i.e. to what extent and how often does the channel change. While channel fading does not directly affect the

shape of a MIMO constellation, the time-variance it introduces in the sensed data may affect the accuracy of supervised classification, such as the one performed by SYMMeTRY.

Several well-established models capture the fading of wireless channels [31], including the Rayleigh model, typical for multi-path environments; and the Rician model, for channels with strong Line-Of-Sight path. $|h_{ij}| = \sqrt{a_{ij}^2 + b_{ij}^2}$ is said to be a *Rayleigh random variable*, if $a_{ij} \in \mathcal{N}(\mu_{a_{ij}}, \sigma_H)$ and $b_{ij} \in \mathcal{N}(\mu_{b_{ij}}, \sigma_H)$, such that $\mu_{a_{ij}} = \mu_{b_{ij}} = \mu_{ij}$ [31]. Similarly, $|h_{ij}|$ is said to be a *Ricean random variable* if $a_{ij} \in \mathcal{N}(\mu_{a_{ij}}, \sigma_H)$ and $b_{ij} \in \mathcal{N}(\mu_{b_{ij}}, \sigma_H)$, such that $\mu_{a_{ij}} \neq \mu_{b_{ij}}$ [31]. Our evaluation (§IV) adopts these realistic models, whereby we control the channel gain by setting $\mu_{a_{ij}}$ and $\mu_{b_{ij}}$, the severity of the fading by controlling σ_H and the rate of the fading by controlling the number of consecutive samples K , for which the channel remains unchanged.

III. METHODOLOGY

The geometry of MIMO constellations exhibit nested self-similarity and this pattern is the key intuition behind our methodology. We summarize this self-similar structure into a discriminative fingerprint which can then be employed for supervised feature-based modulation recognition. To this end, we design our novel features based on the notion of Minkowski-Bouligand fractal dimension [17]. We combine these features with the more traditional higher order cumulants [35] and evaluate their individual and joint accuracy for under-determined MIMO modrec. We demonstrate experimentally that the combined features are robust to various signal conditions and enable improvements over cumulants alone.

A. Preliminaries

We first present a short introduction to higher order cumulants [35] and fractal dimensions [17]. While cumulants have been previously employed in modrec, fractal analysis, which helps summarize self-similarity in IQ sample constellations, has not been considered for modrec before.

Higher Order Cumulants are a successful family of features used in the modrec literature [12], [35]. Cumulants summarize the statistical properties of IQ samples, arising from a complex-valued stationary random process $x(n)$ [35]. The k -th order cumulant is a polynomial function of moments:

$$C_{kv} = \sum_{(\cup_{p=1}^q I_p) = I} [(-1)^{q-1} (q-1)! \prod_{p=1}^q M_{\hat{k}\hat{v}}] \quad (5)$$

where $M_{kv} = E[x(n)^{k-v} x^*(n)^v]$ are the empirical estimates of the moments associated with the stationary process from which the IQ samples are drawn, and $x^*(n)$ denotes the complex conjugation of $x(n)$. The summation extends over all partitions $\{I_1, \dots, I_q\}$, $q \in \{1, \dots, k\}$. Each partition consists of q sets and $\hat{k}$ and $\hat{v}$ are the number of complex and conjugation terms in set p of partition I_p . Some commonly used cumulants in the literature [35] are defined as follows in terms of moments: $C_{21} = M_{21}$ and $C_{42} = M_{41} - (M_{20})^2 - 2(M_{21})^2$. In practice, fourth- and sixth-order cumulants have received

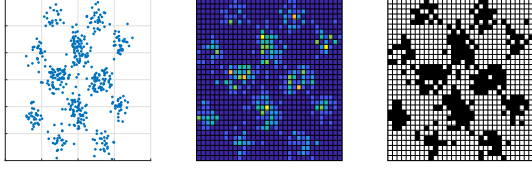


Fig. 6: Converting IQ samples to a binary matrix for box-counting.

most attention, and to remove the effect of the signal scale on cumulants, they are typically normalized by C_{21} [35]. In addition, since some cumulants are complex numbers, their L_2 norm is adopted as a real-valued feature in classification. **Fractal Dimensions.** The intrinsic (or fractal) dimension of a finite set of points is a widely adopted descriptor to quantify fractals' self-similarity [17]. Fractal dimensions have been employed in dimensionality reduction and feature selection in machine learning [37], to optimize the utility of spatial index structures [7] and to analyze biological images [21]. While there are different kinds of dimensions that characterize fractals [33], we focus on the *Minkowski-Bouligand (AKA box-counting) dimension* which has been successfully employed in data analysis applications due to efficient implementations and relatively low-computational complexity [9], [19].

For a set of points in Euclidean space $X = \{x_i\}, x_i \in \mathcal{R}^d$ and a space partitioning in a grid of resolution ϵ (i.e. ϵ is the size of a hyper-cube voxel in the grid), let the *box count* $N(X, \epsilon)$ denote the number of voxels required to cover the points in X . The box counting dimension is defined as the logarithmic rate of increase of the box count $N(X, \epsilon)$ as a function of the log of the resolution $1/\epsilon$ in the limit:

$$D_{\text{box}}(X) := \lim_{\epsilon \rightarrow 0} \frac{\log N(X, \epsilon)}{\log(1/\epsilon)}. \quad (6)$$

This quantity is typically estimated numerically by varying ϵ starting from a fixed maximal resolution and estimating a linear fit for the log-log plot of $N(X, \epsilon)$ as a function of $1/\epsilon$. In this work we are not interested in the actual dimension size, but instead in the discriminative power of the cover set growth functions as features characterizing the constellations of observed IQ samples.

B. Extracting Fractal Features

Our goal is to employ fractal dimensions to extract signatures of modulation-specific IQ samples. Thus, we summarize the self-similar hierarchical structure of MIMO constellations using the box-counting dimension approximation. Informally, *our key idea is that samples from the same modulation will exhibit discriminative box counting growth patterns which differ from those corresponding to other modulations.*

We represent the complex IQ samples within an instance as 2-dimensional real points X based on their real and imaginary parts. Since we will consider the box counts $N(\epsilon)$ starting from a fixed $\epsilon_{\min}$ resolution, we first pre-compute the number of points in each box at the highest resolution. This aggregation helps us (i) keep the complexity of subsequent steps fixed regardless of the number of IQ samples and (ii) allows us to filter noise in low-occupancy bins.

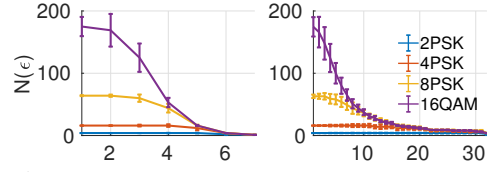


Fig. 7: A comparison of the fractal feature vectors for four modulations using exponential (left) and linear (right) resolution schedules T . X-axis is the box size ϵ .

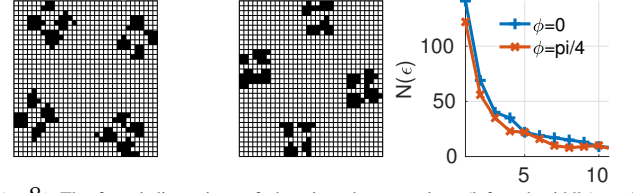


Fig. 8: The fractal dimensions of changing phase rotations (left and middle) could appear different (right). In the right figure, x-axis indicates the index of all grid size enumerations (using exponential growth) and the y-axis indicates the extracted slopes.

This process is demonstrated in Fig. 6. We first normalize the IQ samples. We then impose a square grid at resolution $\epsilon_{\min}$ over the IQ constellation and compute a histogram of point counts in each box (Fig. 6 middle). Finally, we binarize the histogram into full/empty (i.e. 1/0) boxes to obtain a set of coordinates containing samples (Fig. 6 right). The simplest approach to this binarization is to declare any cell with at least one point as full. In order to minimize the effect of noise, we consider a more general frequency thresholding which effectively declares “near-empty” cells as empty. We adopt Otsu’s method commonly used to binarize histograms [29]. The result of the above pre-processing is a binary matrix $Y^{k \times k}, y_{ij} \in \{0, 1\}$ used as an input to the box-counting.

Representing an observed IQ constellation by its estimated fractal dimension D_{box} may result in loss of discriminative power. Instead we compute the empirical decrease of $N(Y, \epsilon)$ as we increase ϵ from $\epsilon_{\min}$ to a value $\epsilon_{\max}$ which covers the complete Y matrix with a single box. Let $T \in \mathcal{R}^m$ be a schedule of increasing box sizes and $N(Y, T_i)$ denote the number of boxes at resolution T_i needed to cover all non-empty cells of our binarized matrix representation of the IQ samples Y . Our fractal features are comprised of the m -dimensional vector $f_{\text{FRA}} = [N(Y, T_1), \dots, N(Y, T_m)]$.

Resolution schedule T. In methods approximating the fractal dimension via box-counting, consecutive ϵ values are typically increased exponentially as the goal is to estimate the slope of an exponential fit of $N(X, \epsilon)$. Assuming a fixed $\epsilon_{\min}$ corresponding to a size- k quantization in Y , the consecutive sizes of boxes in an exponential schedule have the following form $T = (c^i)$ for an integer $c > 1$. In order to extract a more detailed, and potentially more discriminative, shape of the $N(Y, \epsilon)$, we also consider linear schedules of the form $T = (ci)$ for integer values of $c > 1$. A comparison of the average extracted shapes using exponential $T = (2^i)$ (left) and linear $T = (i)$ (right) schedules is presented in Fig. 7. This illustrates (i) the discriminative power of the box-counting curves and (ii) the increased detail due to linear box counting.

Handling phase variation. While the fractal features are invariant to variations in the amplitude within the same modulation (we re-scale IQ samples to have a fixed maximal

norm), variations in the phase, which effectively “rotate” the constellation, may result in significantly different f_{FRA} from the same modulation. We demonstrate this sensitivity to rotation in Fig. 8. The left and middle pane show the same constellation with a phase offset of $\pi/4$, while the right pane shows the corresponding $N(\epsilon)$ profiles. The reason for the differences is that covering boxes are axis-aligned, while clusters in the constellation may have irregular shapes which under rotation may require different number of boxes to be covered at the same resolution. To overcome this challenge, we rotate an observed sample at different phase offset and summarize offset-specific feature variants: f_{FRA}^ω .

It is important to note that noise-free (theoretical) MIMO constellations have a rotational symmetry of order 4 [38], and hence, considering phase offsets outside of $[0, \pi/4]$ would not yield additional discriminative power. Hence, it is reasonable to consider rotations in this interval, where more rotations will potentially add discriminative power at the cost of more features. Our evaluation revealed that simply adding a $\pi/4$ rotation provides an accuracy boost while further rotations add negligible improvements. Thus, our fractal features in all experiments are the concatenation of these two angles $f_{FRA} = (f_{FRA}^0, f_{FRA}^{\pi/4})$.

C. Sparsity Regularized Classification

Our feature vector is constructed by concatenating cumulant and fractal features $f = [f_{CUM}, f_{FRA}]$. The features obtained may have redundant information over multiple dimensions due to multi-perspective projection during the geometric feature extraction. Therefore we utilize an 1-norm linear SVM for classification [43]. It replaces the standard ridge penalty with a lasso penalty, which enforces sparsity of coordinates for the the separation hyperplane, and thus, performs feature selection along with classification. Note that this simple classifier is not a necessary component of our framework and can be replaced by standard feature selectors followed by any classification approach, including deep learning models.

IV. EXPERIMENTS AND RESULTS

We evaluate SYMMeTRY in both realistic simulation and a small-scale MIMO testbed realized with USRPs. Our results show that CUM and FRA features exhibit complimentary performance. Their combination, however, maintains high and consistent performance across all noise regimes, fading conditions and with increasing MIMO transmitter complexity. These trends are retained both in simulation and with over-the-air experiments and demonstrate the high potential of SYMMeTRY to enable low-cost under-determined MIMO modrec.

A. Experimental setup

Implementation. Our box counting pipeline is implemented in MATLAB and executed on Ubuntu 14 PC. We employ liblinear’s [8] implementation of SVMs and one-vs-rest [3] training and evaluation for multi-class classification.

Data. We use data from a realistic simulation and from an over-the-air experiment in a 2x1 MIMO USRP testbed. Our

synthetic datasets are generated with the MATLAB Communications System Toolbox and include four modulations: BPSK, QPSK, 8PSK and 16QAM, which are typically considered in prior modrec literature [36], [44]. Training and testing instances contain 512 IQ samples each.

Evaluation strategy. Our goal is to evaluate SYMMeTRY across varying constellation shapes and channel conditions. To control the constellation shapes, we vary the channel gain sum G_S (Eq. (3)) and the channel gain ratio G_R (Eq. (4)). To control the rate and severity of fading, we set the number of consecutive samples K that experience constant channel and the variance σ_H of each individual path gain $|h_{ij}| \in \mathcal{N}(\mu_{ij}, \sigma_H)$. We also consider varying SNR levels. Unless otherwise noted, we use 2x1 MIMO sensing. In §IV-E we also perform evaluation in more complex settings.

In experiments with fixed (G_S, G_R) combinations, we obtain individual channel coefficients a_{ij} and b_{ij} (see Eq. (II-C)) and (2)) by solving a linear system of equations (3) and (4), which for a 2x1 sensing setup with $|h_{11}| > |h_{12}|$ gives us

$$|h_{11}| = G_S * G_R / (1 + G_R) \text{ and } |h_{12}| = G_S / (1 + G_R) \quad (7)$$

Since this system is under-determined w.r.t. a_{ij} and b_{ij} , we further assume equal real and imaginary components: $a_{ij} = b_{ij}$. Thus, as per Eq. (2), $a_{ij} = |h_{ij}|/\sqrt{2}$. Finally, to evaluate SYMMeTRY with time-variant channels while controlling G_R and G_S , we draw the path gains $|h_{ij}|$ from a normal distribution $\mathcal{N}(\overline{h_{ij}}, \sigma_H)$. For a 2x1 setup, we set $\overline{h_{11}}$ and $\overline{h_{12}}$ according to Eq.(7) and vary σ_H .

In all experiments we compare the accuracy of SYMMeTRY, defined as the fraction of correctly-predicted instances over all instances. We compare the classification accuracy across three different feature configurations: cumulants (CUM) which are employed in all classification-based prior modrec work, fractals (FRA) and their combination (CUM+FRA).

B. SYMMeTRY on a time-invariant channel

We first evaluate the performance of SYMMeTRY in time-invariant channels, i.e. H is fixed in all instances. Our results indicate that CUM and FRA alone have complementary advantages in different gain/noise regimes. The combination of the two features maintains consistently advantageous performance across all experimental regimes.

1) *Effects of channel gain ratio G_R :* We first evaluate the effects of symbol overlap controlled by G_R . As detailed in §II, with a small G_R the MIMO constellation is folded, converging to its SISO equivalent. Mid-range G_R produces a fully-unfolded MIMO constellation, while G_R close to 1 produces a partially-unfolded constellation, whereby some constellation symbols overlap. These effects of G_R on the constellation geometry directly affect the discriminative power of our features. To quantify this, we fix the channel noise $w(n)$ (Eq. (1)) and scale G_R from 0 to 1 in increments of 0.1 in a 2x1 sensing setup. Figs. 9a, 9b present our results comparing the accuracy in mid-noise (10dB) and high noise (5dB) settings. The performance of cumulants deteriorates as G_R increases. This is expected, as cumulants are robust with

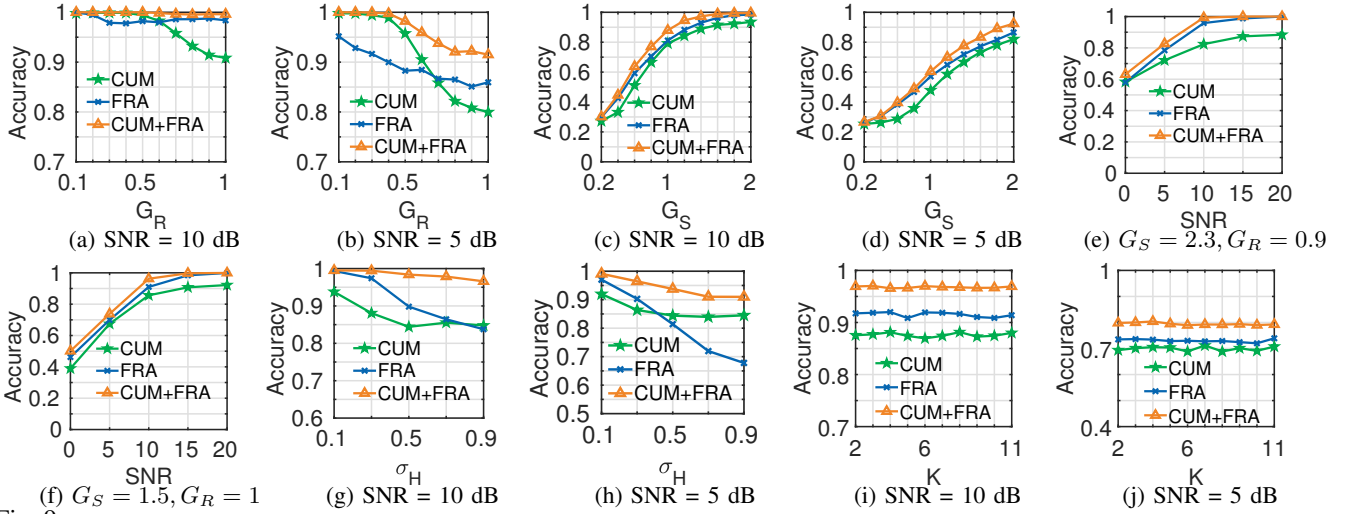


Fig. 9: (a), (b): Accuracy over varying channel gain ratio G_R . Cumulants are robust in SISO-like cases, whereas fractals perform better in MIMO-like cases. Thus, the combination of the two features leads to robust performance classification across constellation shapes; (c), (d): Accuracy over varying channel gain sum G_S . FRA and CUM+FRA features outperform CUM across all G_S . Performance is consistent across SNR regimes.; (e), (f): Accuracy over varying SNR for two channels. FRA and FRA+CUM outperform CUM across all SNR regimes. High-gain channels ($H1$), further emphasize the advantage of FRA and CUM+FRA over CUM; (g), (h): Accuracy across increasing fading variations as controlled by σ_H . FRA+CUM consistently outperforms CUM/FRA alone across σ_H and SNR. (i), (j): Accuracy across channel coherence time. Performance is consistent across K and SNR, indicating that SYMMeTRY is robust in both fast- and slow-fading channels.

SISO modrec (i.e. when G_R is low) and quickly deteriorate as the MIMO constellations unfold with increasing G_R . The fractal features retain stable and high performance across all G_R settings for SNR=10dB, which demonstrate the robustness of the fractal features to channel variations. With lower SNR of 5dB, cumulants outperform fractals for low values of G_R and switch at high G_R . The CUM+FRA combination retains a stable and high performance across ratios even when the SNR is 5dB. These results indicate the robustness of the combined feature CUM+FRA to noise and constellation shape changes.

2) *Effects of channel gain sum G_S* : The sum of individual path gains affects the spread of the MIMO constellation: the higher the gain, the more spread the symbols are, the easier will be to classify a signal's modulation. Thus, we set out to evaluate the effects of G_S on SYMMeTRY's performance. We vary G_S from 0.2 to 2 in increments of 0.2 for 2x1 sensing. Figs. 9c, 9d present our results for mid-noise (10dB) and high noise (5dB) regimes. Across all regimes, FRA and the combined CUM+FRA features outperform cumulants alone. For low gain regimes all counterparts suffer deteriorated performance, which rebounds as the sum gain increases beyond 1. These trends are consistent across SNR regimes.

3) *Effects of SNR*: In this experiment we seek to evaluate the effects of channel noise on the MIMO modrec performance. Specifically, we vary the SNR from 0 to 20dB in increments of 5dB for two channels $H1$ and $H2$ ($G_S^{H1} > G_S^{H2}$). The channels were generated with the MATLAB `comm.MIMOchannel` block and their respective channel responses were $H1 = [1.17 - 0.32i, 1.03 + 0.36i]$ and $H2 = [0.74 + 0.17i, 0.75 - 0.13i]$. Our results are presented in Figs. 9e 9f. FRA and FRA+CUM outperform CUM across all SNR regimes for both channel conditions. With high-gain channels, (i.e. $H1$ in Fig. 9e), the performance of FRA and CUM+FRA over CUM is further emphasized.

C. SYMMeTRY on a time-varying channel

In reality, channels vary with time depending on changes in a link's environment. Hence, we next evaluate the effects of time-varying channels on SYMMeTRY's performance. In our model (§II-C3), we control fading by setting the variance σ_H of the path gain, and the speed at which the channel varies, by setting the number of consecutive samples K for which the channel is time-invariant. We begin by evaluating the effects of σ_H and K on SYMMeTRY's performance. We then evaluate performance with changing constellation geometry (i.e. G_R and G_S). Finally, we examine the effects of SNR. Our findings indicate that the performance of both FRA and CUM alone deteriorate in the face of time-varying channel, however, the combined feature maintains high performance across varying constellation shapes, fading rates and noise level.

1) *Effects of channel fading*: In this experiment, we set G_R to 0.8, G_S to 3 and K to 1 and evaluate accuracy while increasing the path gain variance σ_H from 0.1 to 1 in increments of 0.1. Figs. 9g, 9h shows our results for SNR 10dB(left) and 5dB(right). FRA outperforms CUM at lower channel variance, whereas the contrary holds high variance. The combined feature consistently outperforms either of the features alone across both SNR regimes.

2) *Effects of channel coherence time*: The channel coherence time captures the duration for which the channel response remains unchanged. In our simulation, we control this by setting K , which determines the number of consecutive instances during which the channel remains constant. We define an instance as a vector of 512 IQ samples. Figs. 9i, 9j present SYMMeTRY's accuracy at SNR 10 and 5dB as K increases from 1 to 11 in increments of 1. Here G_R is 0.8, G_S is 1.5 and σ_H is 0.1. The combined feature FRA+CUM outperforms the other two features consistently and across both SNR regimes. The performance does not change significantly, as K increases,

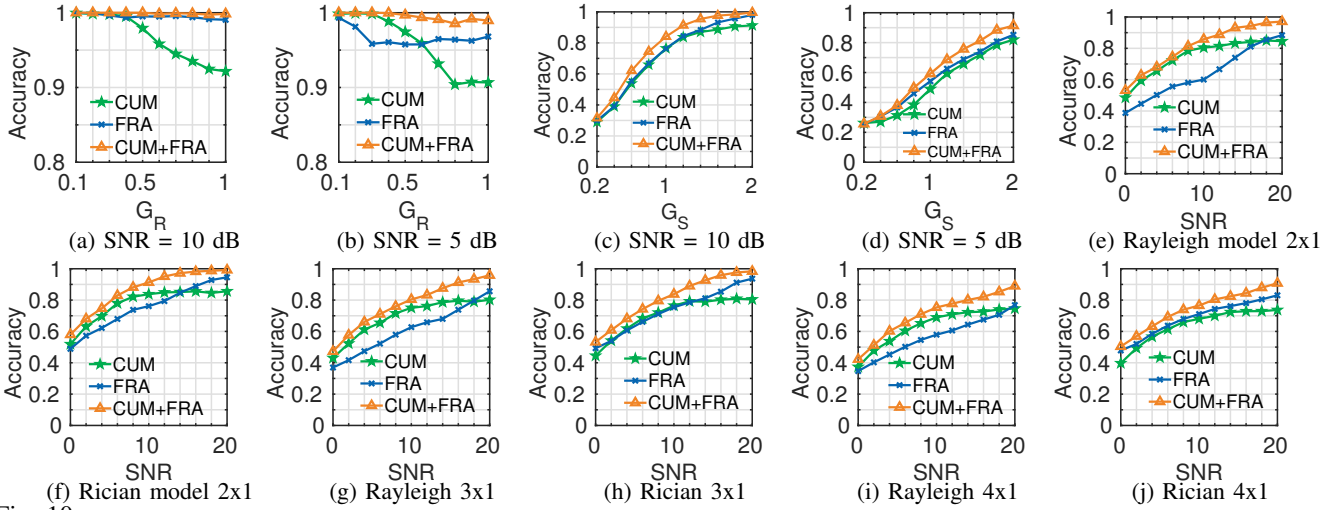


Fig. 10: (a), (b): Accuracy as a function of the constellation symbol overlap; and (c), (d): as a function of the constellation symbol spread; (e), (f): Accuracy with increasing SNR over a Rayleigh and Rician channel; (g)-(j): Accuracy across SNR and channel models with increasingly-complex MIMO transmitter. The combined CUM+FRA feature is able to retain high performance even when a 4-antenna transmitter is sensed with a single-antenna sensor.

indicating that SYMMeTRY will perform robustly in both fast- and slow-fading channels.

3) *Effects of the constellation geometry*: We now evaluate the effects of the constellation shape. As before, we control the overlap of constellation symbols with G_R and the spread of the constellation with G_S . Figs. 10a, 10b present our results for varying G_R with $G_S = 3$, $K = 1$ and $\sigma_H = 0.1$. FRA outperforms CUM at SNR=10dB, while CUM outperform FRA for G_R lower than 0.5 at SNR=5dB. The combined feature leads to high and consistent performance across both SNR regimes and all G_R settings. In addition, Figs. 10c, 10d evaluate the effects of constellation spread (as controlled by G_S) on performance. We vary G_S from 0.2 to 2 in steps of 0.2. G_R is 0.8, K is 1 and σ_H is 0.1. The combined CUM+FRA feature marginally outperforms the other two. The accuracy is affected by both the channel gain sum and the SNR, however, for a gain higher than 1 across both SNR regimes SYMMeTRY achieves high classification performance.

D. SYMMeTRY under realistic unconstrained channel models

In this section, we remove any constraints on the constellation geometry and fading (i.e. all of G_R , G_S , σ_H and K are uncontrolled) and evaluate the performance of SYMMeTRY. We adopt two common channel models: Rician and Rayleigh, and evaluate the performance as a function of the channel SNR. We use the MATLAB `comm.MIMOchannel` function for our implementation with the default parameter setting of the function. In brief, we assume a flat fading channel and also do not consider the Doppler effect (i.e. the sensor and transmitter were considered stationary). We assume the path delay is zero and the averaged power of the path gains is normalized to 0 dB across realizations. We generate one dataset using Rayleigh fading distribution model and another one using Rician model. The channel for each instance is independently generated using the aforementioned MATLAB function. Figs. 10e, 10f show the results of this analysis. CUM and FRA alone have distinct regimes of high performance

TABLE II: Evaluation on 2x1 MIMO USRP testbed

	Rx Gain	CUM	FRA	CUM+FRA
20 in.	40 dB	0.5128	0.5523	0.6240
	50 dB	0.6240	0.7385	0.7973
60 in.	40 dB	0.5128	0.6717	0.7342
	50 dB	0.4800	0.7103	0.7738

across both channel models. The combination of the two leads to high and consistent performance across both channel models and all SNR regimes.

E. SYMMeTRY with increasing transmitter antennas

We now set out to examine SYMMeTRY's performance with increasing transmitter complexity. For this experiment, we use the same data generation setup as discussed in the previous section (§IV-D). We increase the number of transmitter antennas N_t to 3 and 4, while maintaining a single antenna on the sensor. Figs. 10g-10j demonstrate our results. Across all SNR levels and both channel models, the combined feature CUM+FRA maintains the best performance. The performance improves in high SNR regimes. Finally, even though the complexity of the transmitter affects the modrec accuracy, SYMMeTRY with CUM+FRA feature is able to achieve an accuracy of over 0.8 at SNR 10dB and 0.9 at SNR 20dB for the most challenging case of 4x1 MIMO sensing. This demonstrates SYMMeTRY's performance to support modrec of complex MIMO systems even when the sensor is only equipped with a single antenna.

F. SYMMeTRY on real-world traces

We evaluate the performance of our method in real over-the-air transmissions from a USRP-based testbed. We use a transmitter comprised of a USRP B210 attached to an Intel i7-5600U CPU host, and a receiver comprised of a USRP B210 with an Intel i7-6700 CPU host. Both hosts are running on low-latency Linux kernel. We establish a 2x1 sensing setup by transmitting on both Tx channels and receiving on only

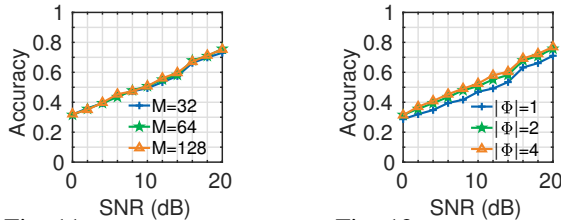


Fig. 11: Impact of M on performance. Fig. 12: Impact of Φ on performance.

on Rx channel. Using GNURadio, the transmitter generates a MIMO signal modulated with BPSK, QPSK, 8PSK and 16QAM. We record 3000 samples for each modulation. We use one third of the samples as the testing set and the rest for classifier training. The two USRP devices are located in a line of sight. We evaluate two difference scenarios: one where the transmitter/sensor are 20in apart and another where they are 60in apart. For each distance, we collect two traces while setting the sensor's Rx gain to 40 and 50dB.

Table II shows the results. For all settings, FRA outperforms CUM by a margin of 0.04 (at 20in/40dB) to 0.23 (at 60in/50dB). The combined features maintain the maximum performance across all measurement scenarios with the highest of 0.797 at 20in and 50dB gain.

G. Effects of parameter selection

In this section, we evaluate the impact of input parameters on performance. Two parameters are worth noting: the number of grid steps at each edge M and the phase projections Φ (§III-A). Figure 11 and 12 present accuracy results for SYMMeTRY using the FRA feature across changing M and Φ , respectively. We use three settings for M ($M = \{32, 64, 128\}$) and three settings for Φ ($\Phi_1 = \{0\}$, $\Phi_2 = \{0, \frac{\pi}{4}\}$ and $\Phi_3 = \{0, \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{6}\}$). As the figures show, SYMMeTRY's performance is not influenced by the selection of M and Φ .

V. RELATED WORK

Modrec without channel estimation. Most previous MIMO modrec work requires channel estimation [4], [14], [16], [24], [27], and hence the constraint of high number of receivers to ensure that the linear system at the core of channel estimation techniques is not under-determined. [12] stand out from the above, as they evaluated the utility of high order statistics as features without channel estimation and established that these features are not robust to noise and the channel mixing effect. Tian et Al. [36] employed a shape-constrained clustering approach within a likelihood based framework. This work, however, incurs a very high computational cost which further grows with the constellation order. In addition, different from us, all above methods require over-determined sensing, and thus, high overhead on the sensing hardware.

Fractal Geometry. Fractal dimension analysis, which is at the core of our novel features, has been successfully applied to the field of digital image processing and applications of medical image analysis. [13] derives fractal dimension as features from pathological images and finds it exhibits advantageous performance in classifying cancer images. In [18], an automatic scar quantification approach based on segmentation-based fractal texture analysis has been presented that provides accurate and

consistent results for MRI scan sequences. Different from the above, our fractal features framework goes beyond the fractal dimension (D_{box}) and employs the detailed rate of decrease of the box counts $N(\epsilon)$.

Sparsity and classification. Our fractal features are in a sense exhaustive and in order to focus on the most discriminate ones we employ sparsity-promoting classifiers. The merits of sparsity on features selection have been studied extensively [39]. We adopt an L1-norm regularized linear SVM [42]. Other modrec approaches have also recently employed feature selection for exhaustive features such as order statistics [40]. The recent success of deep learning techniques across application domains have also inspired approaches for modrec in the communications domain [23], [28], [32]. Our work is complementary with advanced classification approaches as it introduces geometry-aware discriminative features which can be employed in any classification scheme, including deep learning approaches.

VI. DISCUSSION AND OPEN PROBLEMS

Our analysis focuses on under-determined sensing of a direct-mapped MIMO link, i.e., a link in which each stream is transmitted with the same modulation. We expect that our methodology will have advantage in both over-determined settings and heterogeneous-modulation MIMO links: directions we plan to explore as future research. In addition, to handle phase variation we concatenate the FRA features from two specific angles. An important open question is how the accuracy depends on the number of considered angles.

VII. CONCLUSION

In this paper we were the first to consider the challenging problem of supervised MIMO modulation recognition for *under-determined* sensing infrastructures. Our key idea is to exploit the inherent multi-scale self-similarity of MIMO modulation IQ constellations, which persists in under-determined settings. We set out with a careful characterization of the various effects of the channel on the the constellation organization which informed our feature-extraction framework exploiting self-similarity patterns in MIMO constellations. Our framework is rooted in the rich methodology of fractal geometry. We performed an extensive evaluation of our framework in a realistic simulation and in a USRP testbed and demonstrated high and persistent performance across various SNR regimes, channel fading conditions and with increasing complexity of the MIMO transmitter. These results were also confirmed through under-determined MIMO sensing in a small-scale USRP testbed. Our exploration and analysis paints a clear and feasible path to practical and cost-efficient sensing with few receiver antennas—an essential component of future Dynamic Spectrum Access technology, policy and enforcement.

VIII. ACKNOWLEDGEMENTS

This work was supported by NSF CISE Research Initiation Initiative (CRII) grant CNS-1657476, NSF Smart and Connected Communities (SCC) grant CMMI-1831547 and NSF CAREER grant CNS-1845858.

REFERENCES

- [1] Microsoft's Spectrum Observatory. <https://observatory.microsoftspectrum.com/>.
- [2] H. Abuella and M. K. Ozdemir. Automatic modulation classification based on kernel density estimation. *Canadian Journal of Electrical and Computer Engineering*, 39(3):203–209, 2016.
- [3] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [4] V. Choqueuse, S. Azou, K. Yao, L. Collin, and G. Burel. Blind modulation recognition for mimo systems. *MTA Review*, 19(2):183–196, 2009.
- [5] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET communications*, 1(2):137–156, 2007.
- [6] O. A. Dobre, Y. Bar-Ness, and W. Su. Higher-order cyclic cumulants for high order modulation classification. In *IEEE MILCOM*, Boston, MA, 2003.
- [7] C. Faloutsos and I. Kamel. Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension. In *Proceedings of the thirteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 4–13. ACM, 1994.
- [8] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *Journal of machine learning research*, 9(Aug):1871–1874, 2008.
- [9] K. Foroutan-pour, P. Dutilleul, and D. L. Smith. Advances in the implementation of the box-counting method of fractal dimension estimation. *Applied mathematics and computation*, 105(2-3):195–210, 1999.
- [10] H. Gang, L. Jiandong, and L. Donghua. Study of modulation recognition based on HOCs and SVM. In *IEEE VTC Spring*, Milan, Italy, 2004.
- [11] L. Han, F. Gao, Z. Li, and O. A. Dobre. Low complexity automatic modulation classification based on order-statistics. *IEEE Trans. on Wireless Communications*, 16(1):400–411, 2017.
- [12] K. Hassan, I. Dayoub, W. Hamouda, C. N. Nzeza, and M. Berbineau. Blind digital modulation identification for spatially-correlated mimo systems. *IEEE Trans. on Wireless Communications*, 11(2):683–693, 2012.
- [13] P.-W. Huang and C.-H. Lee. Automatic classification for pathological prostate images based on fractal analysis. *IEEE Trans. on Medical Imaging*, 28(7):1037–1050, 2009.
- [14] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang. Automatic modulation classification of overlapped sources using multiple cumulants. *IEEE Trans. on Vehicular Technology*, 66(7):6089–6101, 2017.
- [15] I. Jolliffe. *Principal component analysis*. Springer, 2011.
- [16] E. Kanterakis and W. Su. Modulation classification in mimo systems. In *Proc. IEEE MILCOM*, pages 35–39, 2013.
- [17] K. Kumaraswamy. Fractal dimension for data mining.
- [18] T. Kurzenhofer, K. Breininger, S. Steidl, A. Brost, C. Forman, and A. Maier. Myocardial scar segmentation in lge-mri using fractal analysis and random forest classification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3168–3173. IEEE, 2018.
- [19] L. S. Liebovitch and T. Toth. A fast algorithm to determine fractal dimensions by box counting. *Physics Letters A*, 141(8-9):386–390, 1989.
- [20] G. Lu, K. Zhang, S. Huang, Y. Zhang, and Z. Feng. Modulation recognition for incomplete signals through dictionary learning. In *IEEE WCNC*, pages 1–6, San Francisco, CA, 2017.
- [21] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen. Applications of artificial intelligence in ophthalmology: General overview. *Journal of ophthalmology*, 2018, 2018.
- [22] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood. Chicago Spectrum Occupancy Measurements and Analysis and a Long-term Studies Proposal. In *ACM TAPAS*, Boston, MA, 2006.
- [23] F. Meng, P. Chen, L. Wu, and X. Wang. Automatic modulation classification: A deep learning enabled approach. *IEEE Trans. on Vehicular Technology*, 67(11):10760–10772, 2018.
- [24] M. S. Muhlhaus, M. Oner, O. A. Dobre, and F. K. Jondral. A low complexity modulation classification algorithm for mimo systems. *IEEE Communications Letters*, 17(10):1881–1884, 2013.
- [25] M. S. Muhlhaus, M. Öner, O. A. Dobre, H. U. Jkel, and F. K. Jondral. Automatic modulation classification for mimo systems using fourth-order cumulants. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5. IEEE, 2012.
- [26] A. K. Nandi and E. E. Azzouz. Modulation recognition using artificial neural networks. *Signal processing*, 56(2):165–175, 1997.
- [27] Y. Nie, X. Shen, S. Huang, Y. Zhang, and Z. Feng. Automatic modulation classification based multiple cumulants and quasi-newton method for mimo system. In *Proc. IEEE WCNC*, pages 1–5, 2017.
- [28] T. J. O'Shea, T. Roy, and T. C. Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018.
- [29] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [30] P. Panagiotou, A. Anastasopoulos, and A. Polydoros. Likelihood ratio tests for modulation classification. In *IEEE MILCOM*, Los Angeles, CA, 2000.
- [31] J. G. Proakis and D. G. Manolakis. *Digital signal processing* 4th edition, 2006.
- [32] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. on Cognitive Communications and Networking*, 2018.
- [33] S. N. Rasband. *Chaotic dynamics of nonlinear systems*. Courier Dover Publications, 2015.
- [34] S. Roy, K. Shin, A. Ashok, M. McHenry, G. Vigil, S. Kannam, and D. Aragon. Cityscape: A metro-area spectrum observatory. In *IEEE ICCCN*, Vancouver, B.C., Canada, 2017.
- [35] A. Swami and B. M. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Trans. on Communications*, 48(3):416–429, 2000.
- [36] J. Tian, Y. Pei, Y.-D. Huang, and Y.-C. Liang. Modulation-constrained clustering approach to blind modulation classification for mimo systems. *IEEE Trans. on Cognitive Communications and Networking*, 4(4):894–907, 2018.
- [37] C. Traina Jr, A. Traina, L. Wu, and C. Faloutsos. Fast feature selection using fractal dimension. *Journal of Information and data Management*, 1(1):3, 2010.
- [38] N. Wang, W. Liu, H. Men, M. Jin, and H. Xu. Further complexity reduction using rotational symmetry for edas in spatial modulation. *IEEE Communications Letters*, 18(10):1835–1838, 2014.
- [39] L. Wolf and A. Shashua. Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach. *Journal of Machine Learning Research*, 6(Nov):1855–1887, 2005.
- [40] Z. Wu, S. Zhou, Z. Yin, B. Ma, and Z. Yang. Robust automatic modulation classification under varying noise conditions. *IEEE Access*, 5:19733–19741, 2017.
- [41] W. Xiong, P. Bogdanov, and M. Zheleva. Robust and efficient modulation recognition based on local sequential iq features. In *IEEE International Conference on Computer Communications (IEEE INFOCOM 2019)*, 2019.
- [42] L. Zhang and W. Zhou. On the sparseness of 1-norm support vector machines. *Neural Networks*, 23(3):373–385, 2010.
- [43] J. Zhu, S. Rosset, R. Tibshirani, and T. J. Hastie. 1-norm support vector machines. In *Advances in neural information processing systems (NIPS)*, pages 49–56, 2004.
- [44] Z. Zhu and A. K. Nandi. Blind modulation classification for mimo systems using expectation maximization. In *Proc. IEEE MILCOM*, pages 754–759, 2014.