Beam Discovery Using Linear Block Codes for Millimeter Wave Communication Networks

Yahia Shabara, *Student Member, IEEE*, C. Emre Koksal, *Senior Member, IEEE*, and Eylem Ekici, *Fellow, IEEE*

***Abstract*— The surge in mobile broadband data demands is expected to surpass the available spectrum capacity below 6 GHz. This expectation has prompted the exploration of millimeter wave (mm-wave) frequency bands as a candidate technology for next generation wireless networks. However, numerous challenges to deploying mm-wave communication systems, including channel estimation, need to be met before practical deployments are possible. This paper addresses the mm-wave channel estimation problem and treats it as a beam discovery problem in which locating beams with strong path reflectors is analogous to locating errors in linear block codes. We show that a significantly small number of measurements (compared to the original dimensions of the channel matrix) is sufficient to reliably estimate the channel. We also show that this can be achieved using a simple and energy-efficient transceiver architecture.**

***Index Terms*— Millimeter wave, channel estimation, beam discovery, linear block codes, sparse recovery.**

I. Introduction

We investigate the problem of channel estimation in millimeter wave (mm-wave) wireless communication networks. Mm-wave refers to the wavelength of electromagnetic signals at 30-300 GHz frequency bands. At these high frequencies, channel measurement campaigns revealed that wireless communication channels exhibit very limited number of scattering clusters in the angular domain [2]–[4]. A *cluster* refers to a propagation path or continuum of paths that span a small interval of transmit Angles of Departure (AoD) and receive Angles of Arrival (AoA). Moreover, signal attenuation is very significant at mm-wave frequencies. This motivates the use of large antenna arrays at the transmitter (TX) and receiver (RX) to provide high antenna gains that compensate for high path losses [5]. Nevertheless, due to the high power consumption of mixed signal components, e.g., Analog to Digital Converters (ADCs) [6], conventional digital transceiver architectures that employ a complete RF chain per antenna is not practical. Hence, alternate architectures have been proposed for mm-wave radios with the objective of maintaining close performance to channel capacity. Among the proposed solutions are the use of hybrid analog/digital beamforming [7]–[9] and fully digital beamforming with low resolution ADCs [10]–[12].

Manuscript received June 2, 2018; revised January 24, 2019; accepted May 23, 2019; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor X. Zhou. Date of publication June 28, 2019; date of current version August 16, 2019. This work was supported in part by the NSF CNS under Grant 1618566, Grant 1514260, Grant 1421576, and Grant 1731698. A preliminary version of this work has appeared in the proceedings of IEEE INFOCOM 2018 [1]. *(Corresponding author: Yahia Shabara.)*

The authors are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: shabara.1@osu.edu; koksal.2@osu.edu; ekici.2@osu.edu).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNET.2019.2923395

For all proposed solutions, *channel estimation* remains one of the most critical determinants of performance in communication. Due to the large number of antennas at TX and RX, estimating the full channel gain matrix may require a large number of measurements. For instance, classical analog beamforming transceivers with directional beam patterns at TX and RX demands a number of measurements equal to the product of the number of transmit and receive antennas (ruling out the availability of side information as in [13] and [14]). This imposes a great burden on the estimation process. To address this issue, various methods have been used, the most prevalent among them, is compressed sensing [7], [12], [15]–[17], which leverages channel sparsity. Performance of compressed sensing based approaches is heavily dependent on the design of sensing matrices. For instance, while random sensing matrices are known to perform well, in practice, sensing matrices involve the design of transmit and receive beamforming vectors and the choice of dictionary matrices.[1] Hence, purely random matrices have not been used in practice [18]. On the other hand, no design that involves deterministic sensing matrices has been considered for sparse channel estimation.

Despite the efforts, we do not have a full understanding of the dependence of channel estimation performance on the channel parameters and number of measurements. In an effort to understand this relationship, the study in [19] proposed a multi-user mm-wave downlink framework based on compressed sensing in which the authors evaluate the achievable rate performance against the number of measurements.

In this work, we follow a different approach. We propose a systematic method in which we use sequences of error correction codes chosen in a way to control the channel estimation performance. To demonstrate our approach, consider the following simple example. Let a point to point communication channel be such that, there exists 3 possible receive AoA directions, only one of which may have a strong path to TX. We need to obtain the correct AoA at RX, if it exists. Instead of exhaustively searching all 3 possible AoA directions, we alternatively measure signals from combined directions. For instance, by combining directions 1&2 in one measurement and 2&3 in the next measurement, we can find the AoA in just two measurements. Specifically, four different scenarios might occur, namely, i) only the 1st, or ii) only the 2nd measurement contains a strong path, iii) both 1st and 2nd measurements contain a strong path, and finally, iv) neither measurement reveals a strong path. Interpretation of those cases is: AoA is in i) direction 1, ii) direction 3, iii) direction 2, and iv) none exists. Therefore, only 2 measurements are sufficient for beam detection instead of 3 that are needed for exhaustive search.

[1] A dictionary matrix is used to express the channel in a sparse form.

We will generalize this idea to develop a systematic method for beam detection, inspired by linear block coding. Specifically, we show that linear block error correcting codes (LBC) possess favorable properties that fit in with the desirable behavior of sparse channel estimation. As a result, we are able to **i) provide rigorous criteria for solving the channel estimation problem**, **ii) significantly decrease the number of required measurements**, and **iii) utilize a fairly simple and energy-efficient transceiver architecture.** We design the system using LBCs that leverage the fact that transmission errors are typically sparse in transmitted data streams, and hence, only a few number of erroneous bits need to be corrected per transmitted codeword. Similarly, mm-wave channels are also sparse, i.e., only a small number of AoAs/AoDs carry strong signals. LBCs can correct sparse transmission errors by identifying their location in a transmitted sequence (followed by flipping them). We are inspired by LBC's ability to locate erroneous bits and exploit it to identify the AoAs/AoDs that carry strong signals (and their path gains) among all possible AoA/AoD values. To this end, we exploit hard decision decoding of LBCs, in which the receiver obtains an *error syndrome* that maps to one of the correctable error patterns. An obtained error pattern determines the positions where errors have occurred. Likewise, for channel estimation, the receiver will be designed to do a sequence of measurements that would result in a *channel syndrome*. The resultant channel syndrome shall identify the positions (and values) of non-zero angular channel components.

Contributions of this work can be summarized as follows:

- We set an analogy between beam discovery and channel coding to utilize low-complexity decoding techniques for efficient beam discovery.
- We provide rigorous criteria for setting the number of channel measurements based on the size of the channel and its sparsity level.
- We show that the number of measurements required for beam discovery is linked to the rate of a used linear block code. Hence, maximizing the rate of the underlying code is equivalent to minimizing the number of measurements.
- We develop a simple receiver architecture that enables us to measure signals arriving from multiple directions.

Related Work: The main objective of mm-wave channel estimation is to find a mechanism that can reliably estimate the channel using as few measurements as possible. For instance, in [7], a compressed sensing based algorithm to estimate single-path channels is proposed and an upper bound on its estimation error is derived. Further, the authors propose a multipath channel estimation algorithm based on that of single-path channels. The proposed algorithms in [7] use an adaptive approach with a hierarchical codebook[2] of increasing resolution. Similarly, the work in [16] proposes an adaptive compressive sensing channel estimation algorithm that accounts for off-the-grid AoAs and AoDs by using continuous basis pursuit [20] dictionaries. Such adaptive algorithms divide the estimation process into stages and demand frequent feed back to the TX after each stage. Hence, while the number of required measurements are shown to decrease, these methods may add a considerable overhead.

Other works like [21], [22] and [23] have proposed channel estimation algorithms using overlapped beam patterns. For instance, the algorithm in [23] can estimate multipath channel components by sequentially estimating each path gain using an algorithm designed to estimate single-path channels followed by recursively removing the estimated paths' effect from subsequent measurements. Similar to [7], [16] adaptive beams with increasing resolution that require feedback to TX are used to refine the AoA/AoD estimates. On the other hand, the beam alignment algorithms proposed in [21] and [22] assume a multipath mm-wave channel. These algorithms, with a high probability, can find the best beam alignment in a logarithmic number of measurements (with respect to the total number of available AoA directions). Nonetheless, despite the possible existence of multiple paths, those algorithms are designed to find one path to TX.

Exploiting the results of previous beam alignment operations could be used to reduce the overhead of subsequent alignments. For instance, assuming that successive beam alignments are statistically correlated, Hashemi *et al.* [24] use this contextual information to improve beamforming delay via Multi-Armed Bandit based models. Similarly, changes in the channel can occur due to user's mobility. Zhou *et al.* [13] exploit the spatio-temporal channel correlations due to mobility to enhance beamforming performance. Similarly, [14] used a Markov process to model the evolution of the gains of channel paths. This enabled the reduction of the number of pilots necessary for channel estimation. These studies address the channel tracking problem and did not focus on the first phase of estimation, i.e., beam discovery. In the existing art, the beam discovery problem is typically addressed via exhaustive search/scanning.

Most research efforts in the field of mm-wave channel estimation use the magnitude and phase information of the acquired channel measurements. Nevertheless, if a carrier frequency offset (CFO) error occurs in the transceiver hardware, the phase information might be unreliable. Hence, the work in [21], [22], and [25] tackle this problem by ignoring the phase information. Similar to [21], [22], the solution in [25] can only obtain one (dominant) path between TX and RX using a compressed sensing based technique. The CFO problem is tackled in [26] by considering it as a variable to be estimated.

While the power consumption problem of mmwave systems is commonly alleviated using analog or hybrid beamforming transceivers, an alternative solution is to use low resolution ADCs in fully digital architectures. Owing to the fact that low resolution ADCs operate at much lower power than their high resolution counterparts, the work in [10]–[12] and [27] employ low resolution (single-bit) ADCs in digital transceivers. The work in [17] and [28] study the channel estimation problem using such architectures. Moreover, other solutions include integrated mm-wave and sub-6 GHz systems [29] to provide reliable and energy efficient communication systems.

Notations: A vector and a matrix are denoted by $\boldsymbol{x}$ and $\boldsymbol{X}$, respectively, while x denotes a scalar or a complex number depending on the context. The transpose, conjugate transpose and frobenius norm of $\boldsymbol{X}$ are given by $\boldsymbol{X}^T$, $\boldsymbol{X}^H$ and $\|\boldsymbol{X}\|_F$, respectively. The sets of real and complex numbers are $\mathbb{R}$ and $\mathbb{C}$. The $k\times k$ identity matrix is $\boldsymbol{I_k}$. A set is denoted by $\mathcal{X}$, while $|\mathcal{X}|$ is its cardinality. Finally, $\mathbb{1}()$ is the indicator function.

II. Motivating Example

To elaborate, we present the following example: consider a point to point communication link between a TX with single

[2]A codebook refers to the set of all possible beamforming vectors.

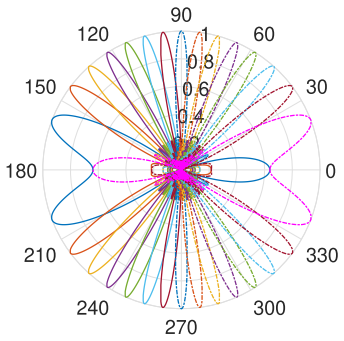

Fig. 1. Beam patterns of all possible angular directions.

antenna ($n_t = 1$) and RX with $n_r = 15$ antennas. Therefore, the vector of channel gains,[3] $\boldsymbol{q}$, is a 15×1 vector, and its corresponding angular (virtual) channel, $\boldsymbol{q}^a$, is a vector of the same size which can be derived using the DFT matrix $\boldsymbol{U_r}$ as $\boldsymbol{q}^a = \boldsymbol{U}_{\boldsymbol{r}}^H\, \boldsymbol{q}$ [30] (this is merely a linear transformation that maps the sequence of channel gains into a sequence of gains from different AoAs. This mapping will be presented in more detail in Section III). Assume a single-path channel, i.e., the channel has only one cluster with a single path in it. Let the path gain be denoted by α. For simplicity assume $\alpha = 1$. Further, let us assume perfect sparsity such that the AoA is along one of the directions defined in the DFT matrix $\boldsymbol{U_r}$, i.e., the channel path will only contribute to one angular bin. Finally, let us also neglect the channel noise.

Based on the channel description above, we get an angular channel vector of the form

$$\boldsymbol{q}^a = \left(q_0^a\ q_1^a\ \cdots\ q_{14}^a\right)^T, \tag{1}$$

such that $q_i^a \in \{0, 1\}$ and the number of non-zero components in $\boldsymbol{q}^a$ is 1. Any component of $\boldsymbol{q}^a$ can be measured using one of the beam patterns shown in Fig. 1.

Objective: Suppose the transmitter sends pilot symbols of the form x=1. Thus, the received vector $\boldsymbol{y}$ can be obtained as

$$\boldsymbol{y} = \boldsymbol{q}x = \boldsymbol{q} \Longleftrightarrow \boldsymbol{y}^a = \boldsymbol{q}^a \tag{2}$$

where $\boldsymbol{y}^a$ is the received vector in the angular domain. So, with change of basis, we can think of $\boldsymbol{q}^a$ as a received sequence with just one non-zero component. To identify the position of this non-zero component, the receiver performs a sequence of channel measurements. Let y_{s_i} denote the i^{th} measurement:

$$y_{s_i} = \boldsymbol{w}_i^H \boldsymbol{y} = \boldsymbol{w}_i^H \boldsymbol{q}, \tag{3}$$

where $\boldsymbol{w}_i$ denotes the i^{th} receive (rx-)combining vector. Our aim is to design channel measurements (i.e., $\boldsymbol{w_i}$'s) such that the correct AoA is identified using a small number of measurements (compared to the number of available AoAs).

Proposed Solution: We consider this non-zero component to be an anomaly to a normally all-zero 15-bin angular channel. Hence, the goal of identifying its position is analogous to finding the most likely 1-bit error pattern of a 15-bit codeword in a linear block code. Now, we need to identify an error correction code with codewords of length 15 and with 1-bit error correction capability [31]. Hence, we can use the binary $(15, 11, 3)$ Hamming code with parity check matrix $\boldsymbol{H}$ of size 4×15 and given by

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{4}$$

where $h_{i,j}$ represents the component at the intersection of row i and column j of $\boldsymbol{H}$. Using hard decision decoding of LBCs, error syndrome vectors of length 4 are obtained. Every possible syndrome vector maps to only one correctable error pattern.[4] Similarly, for channel estimation, several measurements should be performed at RX where each measurement mimics the behavior of a corresponding component in the error syndrome vector. Each measurement boils down to adding signals from a subset of the available 15 directions. Since each measurement can either include the direction of the incoming strong path of gain $\alpha = 1$ or no strong paths at all, then the components of the channel syndrome vector are in $\{0, 1\}$.

For every measurement y_{s_i}, we design $\boldsymbol{w}_i$ based on the entries of the i^{th} row of $\boldsymbol{H}$ such that: if $h_{i,j} = 1$, then we include the beam pattern that points to direction j in $\boldsymbol{w}_i$. For example, the 0^{th} row of $\boldsymbol{H}$ is given by $[100010011010111]$. Hence, $\boldsymbol{w}_0$ should include beam patterns pointing to the set of directions $\{0, 4, 7, 8, 10, 12, 13, 14\}$.

Fig. 2 illustrates this operation for $\boldsymbol{w}_0$. We can see that the resultant beam pattern of $\boldsymbol{w}_i$ combines signals coming from a set of selected directions dictated by the i^{th} row of $\boldsymbol{H}$. We call the obtained measurement vector, $\boldsymbol{y_s}$, the *channel syndrome* which is analogous to error syndromes in hard decision decoding of LBCs. Then, a table that maps every possible channel syndrome to a unique corresponding channel can be constructed. Table I shows this mapping.

In this example, we are able to estimate the channel based on only 4 measurements as opposed to 15, which is the number of measurements with exhaustive search. Important aspects of our proposed method include the choice of codes, the design of precoding and rx-combining measurement vectors, the effect of variable gains and phases of different paths and the occurrence of measurement errors.

Remark (Receiver Architecture): Note that, to achieve beam patterns similar to the one shown in Fig. 2, the receiver architecture needs controllable low-power amplifiers (variable gain amplifiers (VGA)) at each antenna element. The reason is that, unlike analog beamforming where the components of rx-combining vectors $\boldsymbol{w_i}$'s have equal magnitudes but different phases, the rx-combing vectors that produce beam patterns

TABLE I
MAPPING OF CHANNEL SYNDROMES TO ANGULAR CHANNELS

Channel Syndrome $\boldsymbol{y_s}^T$	Angular Channel $\boldsymbol{q}^{aT}$
[0 0 0 0]	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[1 0 0 0]	[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[0 1 0 0]	[0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
[0 0 1 0]	[0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 1]	[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
⋮	⋮
[1 0 0 1]	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]

[3]Let all the channels have one single significant tap.

[4]A correctable error pattern of a $(15, 11, 3)$ Hamming code is any 15×1 binary vector that contains only one '1' (at the error's position).

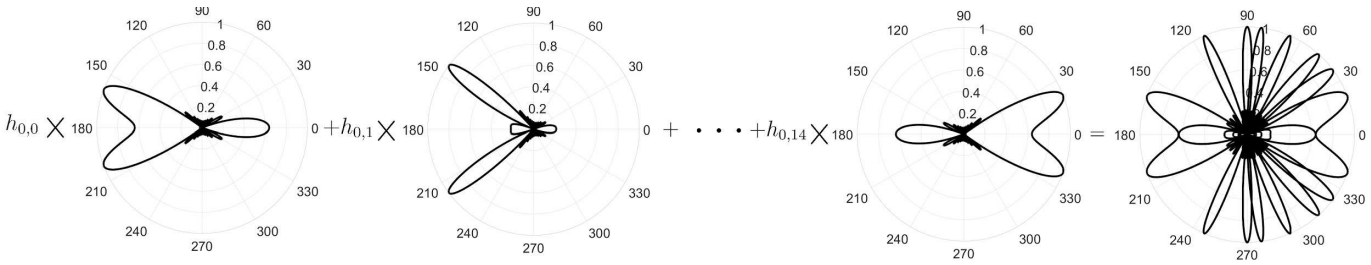

Fig. 2. Beam pattern of receive combining vector $\boldsymbol{w_0}$.

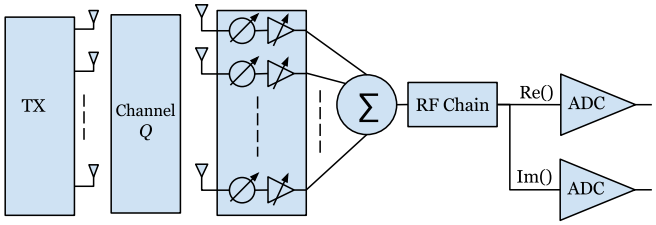

Fig. 3. Hardware Block Diagram: Every antenna is connected to a phase shifter and low-power variable gain amplifier. Then, all outputs are combined using an adder and passed to an RF chain with in-phase and quadrature channels.

as in Fig. 2 would have different magnitude values, as well. In fact, it is typical for existing implementations to include VGAs to correct for imperfections of the transceiver's electronic components [32]. Such amplitude control is also presumed possible in mm-Wave standards, e.g., IEEE 802.11ad, and is available in commercial off-the-shelf devices, e.g. Wilocity Wil6200 network cards [33].

Motivation for LBC-inspired approach: LBCs are designed to discover and correct a certain maximum number of errors in a codeword of a specified length. This objective is achieved by adding redundant *parity check* bits to the original information sequence. The number of parity bits is dependent on the length of the original information sequence and the required error correction capability. What makes our devised approach attractive is that the number of measurements needed for channel estimation can be shown to be equal to the number of parity bits of some corresponding code. Hence, we can control the estimation performance via appropriate code selection. In this work, we will propose a method to specify the number of necessary channel measurements as a function of the rate of the underlying code.

III. System Model

Consider a point-to-point millimeter-wave wireless communication system with a transmitter (TX) equipped with n_t antennas and a receiver (RX) with n_r antennas placed at fixed locations. Uniform Linear Arrays (ULA) are assumed at both TX and RX where each antenna element is connected to a phase shifter and a variable gain amplifier. A single RF chain at the receiver, with in-phase (I) and quadrature (Q) channels, is fed through a linear combiner (see Fig. 3). Only two mid-tread ADCs, with $2^b{+}1$ quantization levels, are utilized, where quantization levels take values from the set $\mathcal{Y} = \{-2^{b-1}, \ldots, -1, 0, 1, \ldots, 2^{b-1}\}$.

We adopt a single-tap channel model where $\boldsymbol{Q} \in \mathbb{C}^{n_r \times n_t}$ denotes the channel matrix between TX and RX. Assume that the channel has L clusters, where each cluster contains a single path with gain α_l, AoD θ_l, and AoA ϕ_l. The channel is assumed to be sparse such that $L \ll n_t, n_r$. Let $\alpha_l^b \in \mathbb{C}$ denote the baseband channel gain and is defined as

$$\alpha_l^b{=}\alpha_l \sqrt{n_t n_r}\, e^{-j\frac{2\pi\rho_l}{\lambda_c}}, \tag{5}$$

where ρ_l is the length of path l and λ_c is the carrier wavelength. The angular cosines of AoD and AoA associated with path l are denoted by Ω_{tl} and Ω_{rl}, respectively. The transmit and receive spatial signatures along the direction Ω are given by $\boldsymbol{\aleph_t}(\Omega)$ and $\boldsymbol{\aleph_r}(\Omega)$ such that

$$\boldsymbol{\aleph_t}(\Omega){=}\frac{1}{\sqrt{n_t}}\left(1{,}e^{-j2\pi\Delta_t\Omega}{,}e^{-j2\pi 2\Delta_t\Omega}{,}\ldots{,}e^{-j2\pi(n_t-1)\Delta_t\Omega}\right), \tag{6}$$

where $\boldsymbol{\aleph_r}(\Omega)$ has a similar definition to $\boldsymbol{\aleph_t}(\Omega)$, and Δ_t and Δ_r are the antenna separations at TX and RX normalized by the wavelength λ_c. Let the average path loss be denoted by μ. Thus, $\boldsymbol{Q}$ can be written as

$$\boldsymbol{Q} = \sum_{l=1}^{L} \frac{\alpha_l^b}{\mu}\, \boldsymbol{\aleph_r}(\Omega_{rl})\, \boldsymbol{\aleph_t}^H(\Omega_{tl}). \tag{7}$$

We define $\boldsymbol{U_t}$ and $\boldsymbol{U_r}$ as the unitary Discrete Fourier Transform (DFT) matrices whose columns constitute an orthonormal basis for the transmit and receive signal spaces $\mathbb{C}^{n_t}$ and $\mathbb{C}^{n_r}$, respectively. $\boldsymbol{U_t}$ (and similarly $\boldsymbol{U_r}$) is given by

$$\boldsymbol{U_t} = \left(\boldsymbol{\aleph_t}(0)\ \boldsymbol{\aleph_t}(\tfrac{1}{\varkappa_t})\ \ldots\ \boldsymbol{\aleph_t}(\tfrac{n_t-1}{\varkappa_t})\right), \tag{8}$$

where $\varkappa_t$ ($\varkappa_r$) is the normalized length of the transmit (receive) antenna array such that $\varkappa_t{=}n_t\Delta_t$ ($\varkappa_r{=}n_r\Delta_r$). Let $\boldsymbol{Q^a}$ be the channel matrix in the angular domain [30], where

$$\boldsymbol{Q^a} = \boldsymbol{U_r}^H \boldsymbol{Q} \boldsymbol{U_t}. \tag{9}$$

The rows and columns of $\boldsymbol{Q}^a$ divide the channel into resolvable RX and TX bins, respectively. Further, we assume a perfect sparsity model in which AoDs θ_l, and AoA ϕ_l, are along the directions defined in $\boldsymbol{U_t}$ and $\boldsymbol{U_r}$ [7], [17], [23]. Hence, each channel cluster will only contribute to a single pair of TX and RX bins. Therefore, $\boldsymbol{Q}^a$ has a maximum of L non-zero TX and RX bins.

The baseband channel model is given by

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{n}, \tag{10}$$

where $\boldsymbol{x} \in \mathbb{C}^{n_t}$ is the transmitted signal, $\boldsymbol{y} \in \mathbb{C}^{n_r}$ is the received signal and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, N_0 \boldsymbol{I}_{n_r})$ is an i.i.d. complex Gaussian noise vector.

Let $\boldsymbol{f} \in \mathbb{C}^{n_t}$ and $\boldsymbol{w} \in \mathbb{C}^{n_r}$ be the precoding and rx-combining vectors, respectively. The transmit signal $\boldsymbol{x}$ is

given by $\boldsymbol{x} = \boldsymbol{f}s$ where s is the transmitted symbol with average power $\mathbb{E}(ss^H)=P$. After the receiver applies the rx-combining vector $\boldsymbol{w}$, the resultant symbol u can be written as

$$u = \boldsymbol{w}^H \boldsymbol{Q} \boldsymbol{f} s + \boldsymbol{w}^H \boldsymbol{n}. \tag{11}$$

Afterwards, u is passed forward to the ADCs. There, a quantized version, u_s, of u is obtained such that

$$u_s = [\boldsymbol{w}^H \boldsymbol{Q} \boldsymbol{f} s + \boldsymbol{w}^H \boldsymbol{n}]_+ , \tag{12}$$

where $[\,\cdot\,]_+$ represents the quantization function. Now, u_s constitutes a single, quantized, unit measurement obtained using specific $\boldsymbol{f}$ and $\boldsymbol{w}$ vectors such that $u_s = u_s^{(r)} + i u_s^{(i)} \subseteq \mathbb{C}$ where $\left(u_s^{(r)}, u_s^{(i)}\right) \in \mathcal{Y}^2$ are the real and imaginary components of u_s, respectively.

We assume that $\boldsymbol{Q}$ remains fixed throughout the entire estimation process. The noise component $\boldsymbol{w}^H \boldsymbol{n}$ normalized by $\|\boldsymbol{w}\|$ is also a complex gaussian random variable such that $\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\boldsymbol{n} \sim \mathcal{CN}(0, N_0)$. We define the signal to noise ratio (*SNR*) on a per path basis such that *SNR* of path l is given by

$$SNR_l = \frac{P}{N_0} \left| \frac{\alpha_l^b}{\mu} \right|^2 . \tag{13}$$

Note that the actual received *SNR* depends on all path gains included in a measurement.

IV. Problem Statement

Suppose a maximum number of L clusters need to be discovered in the channel where $L \ll n_t, n_r$. Under the prefect sparsity assumption, $\boldsymbol{Q}^a$ has a maximum of L non-zero RX and TX angular bins. Our objective is to identify the angular positions at which channel clusters exist and identify their path gain values using the **least possible number of measurements**. Let the number of measurements be m such that each measurement, $u_{s_{i,j}}$, is obtained using the precoder $\boldsymbol{f_j}$ and rx-combiner $\boldsymbol{w_i}$. Let the number of rx-combiners and precoders be m_1 and m_2, respectively. Measurements take the form $u_{s_{i,j}} = [\boldsymbol{w_i}^H \boldsymbol{Q} \boldsymbol{f_j} s + \boldsymbol{w_i}^H \boldsymbol{n}]_+$. Let $\xi()$ be a mapping function that takes in the measurements $\{u_{s_{i,j}}\}_{\forall i,j}$ as inputs and returns the estimated channel $\widehat{\boldsymbol{Q}}^{\boldsymbol{a}}$. For each j, we stack the measurements $\{u_{s_{i,j}}\}_{\forall i}$ in a single (syndrome) vector such that $\boldsymbol{u_{s_j}} = [u_{s_{0,j}}\ u_{s_{1,j}} \dots u_{s_{m_1-1,j}}]^T$. Our design variables are the precoding vectors $\boldsymbol{f_j}$, rx-combining vectors $\boldsymbol{w_i}$, the number of measurements m, the mapping function $\xi()$, and the transmitted symbol power P.

In its essence, solving this problem boils down to finding the optimal set of measurements $\{u_{s_{i,j}}\}_{\forall i,j}$ and the mapping function $\xi()$ such that $\boldsymbol{Q}^a$ can be estimated using the minimum number of measurements. For ease of explanation, we first consider a channel with a single transmit antenna and n_r receive antennas. Therefore, no precoding is needed and the design of measurements is reduced to designing the rx-combining vectors $\boldsymbol{w_i}$. Recall that in the motivating example in Section II, we dealt with a special case of $n_r \times 1$ channels where we sought to find the direction of arrival of a channel with a single path of known gain, $\alpha = 1$. In the general case, we should consider arbitrary path gains $\alpha \in \mathbb{C}$ and channels with multiple paths.

V. Beam Discovery

In this section, we present our proposed solution. As an initial step, we solve a simplified version of the problem where communication channels have a single transmit antenna and multiple receive antennas. Afterwards, we will build on it to provide the solution for general channels with multiple transmit and receive antennas.

A. Beam Detection using LBC-inspired approach

To identify the exact number of measurements and their corresponding design, we follow a decoding-like approach of LBC.[5] First, we need to find an LBC, C, that has an error correction capability e_n such that i) the maximum number of clusters in the channel, L, is equal to e_n and ii) the length of its codewords n is equal to the number of antennas n_r (n_r is also the number of resolvable directions). The code C has a parity check matrix $\boldsymbol{H}$ which represents the link between channel decoding and beam detection problems. Binary codes deal with data and error sequences defined over the finite field $GF(2)$, i.e., addition and multiplication operations are defined over $GF(2)$ with binary inputs and outputs, i.e., 1's and 0's. However, mm-wave channel parameters are defined over the complex numbers field $\mathbb{C}$. Therefore, to account for arbitrary path gains, we should be able to extend this concept to $\mathbb{C}$.

Although $\boldsymbol{H}$ is defined over $GF(2)$, we interpret its $'1'$ and $'0'$ entries as real numbers. Then, similar to channel decoding, we seek to obtain a channel syndrome, $\boldsymbol{y_s}$, such that $(\boldsymbol{y}_s)^T \equiv (\boldsymbol{q}^a)^T \boldsymbol{H}^T \implies \boldsymbol{y}_s \equiv \boldsymbol{H}\boldsymbol{q}^a$. This matrix multiplication can be realized using channel measurements such that each measurement gives one component in $\boldsymbol{y_s}$. Measurements $\{y_{s_i}\}_{\forall i}$ make up the components of the channel syndrome vector $\boldsymbol{y_s}$. Then, we need to find a mapping function $\xi()$ that takes in the channel syndrome vector $\{\boldsymbol{y_s}\}$ as an input and returns the estimated channel $\widehat{\boldsymbol{q}}^{\boldsymbol{a}}$. The position of each non-zero component in $\widehat{\boldsymbol{q}}^{\boldsymbol{a}}$ identifies a path's AoA, and its value identifies its baseband path gain. Finally, for this to work, we need to show that such channel measurements provide one-to-one mapping to the channel. In other words, $\boldsymbol{y_s}$ must be a *sufficient statistic* for estimating the channel. In Section V-C, we will show that our design results in the sufficient statistic we seek to achieve.

Remark (Difference between $\boldsymbol{y_s}$ and $\boldsymbol{u_s}$): Both $\boldsymbol{y_s}$ and $\boldsymbol{u_s}$ refer to vectors of measurement symbols, however, $\boldsymbol{u_s}$ is considered to be the noise corrupted and quantized version of $\boldsymbol{y_s}$. Specifically, $\boldsymbol{u_s} = [\boldsymbol{y_s} + \boldsymbol{z}]_+$ such that $\boldsymbol{z}$ is the measurement noise vector. While $\boldsymbol{u_s}$ is what we expect to observe, our design of measurements focuses on finding $\boldsymbol{y_s}$; an error-free symbol. Of course, errors degrade beam discovery performance. Thus, in Section VI, we will deal with the effect of measurement errors separately and present a solution that

[5] In channel coding, the convention is to use row vectors. Thus, let $\boldsymbol{x}$ and $\boldsymbol{c}$ be $1\times k$ and $1\times n$ binary row vectors that represent an information sequence and its corresponding codeword of an LBC, respectively. Also let $\boldsymbol{r}=\boldsymbol{c}+\boldsymbol{e}$ be a received sequence corrupted by $1\times n$ error pattern $\boldsymbol{e}$. To decode $\boldsymbol{r}$, we calculate an error syndrome vector $\boldsymbol{s}$, of size $1\times n-k$, such that $\boldsymbol{s}=\boldsymbol{r}\boldsymbol{H}^T$, where $\boldsymbol{H}$ is the parity check matrix of the used LBC. Then, a most likely error pattern $\hat{\boldsymbol{e}}$ can be uniquely identified by $\boldsymbol{s}$ using a look-up table called the *standard array*. Finally, the decoded codeword is obtained using $\hat{\boldsymbol{c}}=\boldsymbol{r}-\hat{\boldsymbol{e}}$. A decoding error occurs if the number of errors, identified using 1's in $\boldsymbol{e}$, is beyond the error correction capability of the used code, denoted by e_n. Note that in this context, all vectors, matrices and math operations are over GF(2).

increases reliability of beam discovery. The separate treatment of measurement errors simplifies the design and provides a clear understanding of the nature of our solution.

Remark (Number of Measurements): The solution we obtain is dependent on channel parameters, namely, the number of antennas and the sparsity level of the channel. That is, at a fixed sparsity level, i.e., fixed number of clusters L, a larger number of antennas necessitates more channel measurements. In other words, the high resolution realized by large n_r comes at a price of an increased number of measurements. Similarly, at fixed n_r, more channel clusters involve more measurements for correct channel estimation.

B. Measurements Design

Recall that each component in $\boldsymbol{q}^a$ represents a resolvable angular direction at the receiver. Let each resolvable direction be given an identification number (*dir$_{rx}$#i*). Also let *beam$_{rx}$#i* denote the beam pattern pointing to *dir$_{rx}$#i*, i.e., a signal coming from *dir$_{rx}$#i* can be individually measured using *beam$_{rx}$#i* (similar to beam patterns in Fig. 1).

Now, using careful design of $\boldsymbol{w_i}$'s, we seek to obtain

$$\boldsymbol{y_s} = (y_{s_0}, y_{s_1}, \ldots, y_{s_{m-1}})^T \equiv \boldsymbol{H}\boldsymbol{q^a}, \text{ where} \tag{14}$$

$$y_{s_i} = \boldsymbol{w_i^H}\boldsymbol{q} \quad , \forall\, 0 \le i \le m-1 \tag{15}$$

To achieve this, each rx-combining vector $\boldsymbol{w_i}$ is designed as a multi-armed beam, i.e., composed of several sub-beams similar to the beam pattern in Fig. 2. The sub-beams included in each $\boldsymbol{w_i}$ are identified by the i^{th} row of the matrix $\boldsymbol{H}$. That is, only if $h_{i,j}$, the intersection of the i^{th} row and j^{th} column, is $= 1$, do we include *beam$_{rx}$#j* as a sub-beam in $\boldsymbol{w_i}$ (also refer to our discussion in Section II).

The design of rx-combining vectors is a crucial aspect of this work. As an initial step towards obtaining proper rx-combining vectors, we consider designing $\boldsymbol{w_i}$'s using linear summation of all analog beamformers that correspond to *beam$_{rx}$#j*'s $\forall j \;:\; h_{i,j} = 1$. Let $\Omega_j = \cos(\phi_j) = \frac{j}{L_r} \;\; \forall j \in \{0, \ldots, n_r - 1\}$, such that $\boldsymbol{\aleph_r}(\Omega_j)$ is the spatial signature of *beam$_{rx}$#j*. Then, $\boldsymbol{w_i}$ can be designed as

$$\boldsymbol{w_i} = \sum_{j=0}^{n_r-1} \mathbb{1}_{\{h_{i,j}=1\}} \boldsymbol{\aleph_r}(\Omega_j) \tag{16}$$

C. Sufficient Statistic

We will show in this section that each channel syndrome can only be mapped to a single *measurable channel*. A measurable channel in this context refers to $n_r \times 1$ channels with L non-zero components such that $L \le e_n$, where e_n is the error correction capability of the underlying code C and $n_r = n$ is its CWs length. Let $\mathcal{Q}^a$ be the set of all measurable channels:

$$\mathcal{Q}^a \triangleq \{\boldsymbol{q^a} \in \mathbb{C}^{n_r} : |q_i^a : q_i^a \neq 0| \le e_n\}. \tag{17}$$

Since each measurement combines signals coming from multiple directions, each element in the channel syndrome vector is a linear combination of a subset of the available paths. In other words, each measurement has the possibility that one or more paths are included in it. This setting is rather challenging. To understand why, consider a channel that has two paths with gains $\alpha_1, \alpha_2 \in \mathbb{C}$. Suppose that α_1 and α_2 are of equal magnitudes but are out-of-phase (i.e., phase shift $= 180^\circ$). Hence, if signals coming from both paths are combined in a single measurement, the resultant value is 0 which is similar to the result we get if no paths exist in the measured directions. Also each channel measurement can be a result of endless possibilities for the combined path gain values. So, a natural question to ask is: does this ambiguity cause measurement errors? The direct answer to this question is: **No**. In the sequel we will show that the resulting channel syndrome, i.e., the combination of all channel measurements, is sufficient to correctly estimate the channel.

First, recall our discussion in Footnote 5. Then, consider all *single-bit error patterns* $\boldsymbol{e^{(i)}}$ of some code C, with maximum number of correctable errors $= e_n$ such that

$$e_k^{(i)} = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases}, \tag{18}$$

where $e_k^{(i)}$ is the k^{th} component of $\boldsymbol{e^{(i)}}$. Also let $\boldsymbol{s^{(i)}}$ be the corresponding error syndrome of $\boldsymbol{e^{(i)}}$. Recall that $\boldsymbol{s^{(i)}} = \boldsymbol{e^{(i)}}\boldsymbol{H}^T$. Hence, we can see that $\boldsymbol{s^{(i)}}$ is exactly the i^{th} row of $\boldsymbol{H}^T$, i.e., i^{th} column of $\boldsymbol{H}$. Let $\mathcal{E}_C$ denote the set of correctable error patterns of the code C such that

$$\mathcal{E}_C \triangleq \left\{\boldsymbol{e} \in \{0,1\}^n : \boldsymbol{e} = \sum_{i=1}^{n} \varrho_i \boldsymbol{e}^{(i)}, \varrho_i \in \{0,1\} : |\varrho_i : \varrho_i = 1| \le e_n\right\}. \tag{19}$$

Now, we can write any correctable error pattern $\boldsymbol{e} \in \mathcal{E}_C$ as a linear combination of all single-bit error patterns over the finite field $GF(2)$ such that $\boldsymbol{e} = \sum_{i=1}^{n} \varrho_i \boldsymbol{e^{(i)}}$, and its corresponding error syndrome is $\boldsymbol{s} = \sum_{i=1}^{n} \varrho_i \boldsymbol{s^{(i)}}$.

Lemma 1: For an error pattern $\boldsymbol{e}_t$ with number of bit errors identical to e_n, its syndrome $\boldsymbol{s}_t$ is a linear combination of e_n linearly independent vectors $\boldsymbol{s}^{(i)}$.

Proof: We are going to prove this lemma by contradiction. First, assume that $\boldsymbol{s}_t$ is a linear combination of e_n linearly **dependent** vectors $\boldsymbol{s}^{(i)}$ over $GF(2)$. Therefore, there exists another error syndrome $\boldsymbol{s}_t^*$ composed of only linear combination of independent vectors $\boldsymbol{s}^{(i)}$ such that $\boldsymbol{s}_t = \boldsymbol{s}_t^*$. Therefore, there exists another error patter $\boldsymbol{e}_t^*$ with number of errors strictly less than e_n such that its syndrome $\boldsymbol{s}_t^* = \boldsymbol{s}_t$. Since $\boldsymbol{e}_t^*$ has a number of errors less than e_n, then it is a correctable error pattern, and since all error syndromes of correctable error patterns are different, then $\boldsymbol{s}_t^*$ should be $\neq \boldsymbol{s}_t$. Hence, we arrive at a contradiction. □

It is also easy to see that if $\boldsymbol{e}_{t_1}$ and $\boldsymbol{e}_{t_2}$ are two different correctable error patterns, then their error syndromes $\boldsymbol{s}_{t_1}$ and $\boldsymbol{s}_{t_2}$ are composed of a linear combination of different sets of single-bit error syndromes $\boldsymbol{s}^{(i)}$.

Lemma 2: Any $n-$dimensional linearly independent vectors over $GF(2)$, are also linearly independent over $\mathbb{C}^n$.

Proof: Let $\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}$ be a set of $n-$dimensional vectors defined over $GF(2)$. The vectors $\boldsymbol{v_i}$ can be made the columns of an $n \times m$ matrix $\boldsymbol{\Psi}$. Since all $\boldsymbol{v_i}$'s are linearly independent over $GF(2)$, then $\boldsymbol{\Psi}$ is a left-invertible matrix. Therefore, there exists a non-zero (modulo 2) $m \times m$ minor of $\boldsymbol{\Psi}$. Now, suppose the entries in $\boldsymbol{\Psi}$ are interpreted as real numbers. Therefore, $\boldsymbol{\Psi}$,

now taken over $\mathbb{R}$, has an $m\times m$ sub-matrix whose determinant is non-zero, which proves that it is invertible. Therefore, the vectors $\boldsymbol{v_i}$'s, i.e., columns of $\boldsymbol{\Psi}$, are linearly independent over $\mathbb{R}$ which, using the same argument, can also be shown to be linearly independent over $\mathbb{C}$. □

Suppose that entries of $\boldsymbol{H}$ and $\boldsymbol{e^{(i)}}$ are interpreted as real numbers, then we can write the channel $\boldsymbol{q^a}$ as

$$(\boldsymbol{q}^a)^T = \alpha_1 \boldsymbol{e}^{(1)} + \alpha_2 \boldsymbol{e}^{(2)} + \cdots + \alpha_n \boldsymbol{e}^{(n)} \tag{20}$$

where $\alpha_i \in \mathbb{C}$ and $\sum_{i=1}^{n} \mathbb{1}_{\{\alpha_i \neq 0\}} \leq e_n$. Therefore, each channel syndrome $(\boldsymbol{y_s})^T = (\boldsymbol{q}^a)^T \boldsymbol{H}^T \implies \boldsymbol{y_s} = \boldsymbol{H}\boldsymbol{q}^a$ is a linear combination of independent vectors in $\mathbb{C}^{n-k}$ (columns of $\boldsymbol{H}$). Therefore, all possible measurable channels yield unique channel syndromes which implies that they are sufficient for the channel estimation problem.

D. Mapping Function $\xi()$

After showing that each measurable channel can be mapped to a unique channel syndrome, we need to find this mapping function, i.e., $\xi : \boldsymbol{y_s} \rightarrow \widehat{\boldsymbol{q}}^{\boldsymbol{a}}$, where $\widehat{\boldsymbol{q}}^{\boldsymbol{a}}$ denotes the estimated channel. We propose two different approaches to find $\xi()$.

1) Look-up Table Method: Similar to hard decision decoding where a look-up table is constructed that maps error syndromes to their corresponding error patterns, we can construct a look-up table that indicates which channel corresponds to an obtained channel syndrome. The number of table entries depend on the employed ADC resolution and are based on error-free channel syndromes $\boldsymbol{y_s}$. The mapping between the actual error-corrupted syndrome $\boldsymbol{u_s}$ and $\boldsymbol{q}^a$ is obtained using l^2- norm minimization where $\xi()$ returns the channel whose corresponding $\boldsymbol{y_s}$ has the smallest distance to $\boldsymbol{u_s}$. The l^2 distance function $\delta(\cdot,\cdot)$ is given by

$$\delta(\boldsymbol{y_s}, \boldsymbol{u_s}) = \|\boldsymbol{y_s} - \boldsymbol{u_s}\|_2 = \sqrt{\sum_{i=0}^{m-1} |y_{s_i} - u_{s_i}|^2}. \tag{21}$$

Further details of the table construction can be found in [34].

2) Search Method: Recall that $\boldsymbol{y_s} = \boldsymbol{H}\boldsymbol{q^a}$ (Eq. (14)), and let the parity check matrix $\boldsymbol{H}$ be represented as:

$$\boldsymbol{H} = \left(\boldsymbol{h_1}\ \boldsymbol{h_2}\ \ldots\ \boldsymbol{h_{n_r}}\right), \tag{22}$$

where $\boldsymbol{h_i}$ is the i^{th} column of $\boldsymbol{H}$. Thus, we can write $\boldsymbol{y_s}$ as:

$$\boldsymbol{y_s} = q_1^a \boldsymbol{h_1} + q_2^a \boldsymbol{h_2} + \cdots + q_{n_r}^a \boldsymbol{h_{n_r}}, \tag{23}$$

where q_i^a is the i^{th} component of $\boldsymbol{q^a}$. Note that $\boldsymbol{q^a}$ is $L-$sparse, i.e., we have no more than L non-zero components q_i^a. Let the indices of the non-zero components be $x_1, x_2, \ldots, x_L$, hence, $\boldsymbol{y_s}$ can succinctly be written as:

$$\boldsymbol{y_s} = q_{x_1}^a \boldsymbol{h_{x_1}} + q_{x_2}^a \boldsymbol{h_{x_2}} + \cdots + q_{x_L}^a \boldsymbol{h_{x_L}}. \tag{24}$$

Then, we can write Eq. (24) in matrix form as:

$$\boldsymbol{y_s} = \boldsymbol{C}\boldsymbol{q_c^a}, \quad \text{where} \tag{25}$$

$$\boldsymbol{C} \triangleq \left(\boldsymbol{h_{x_1}} \quad \boldsymbol{h_{x_2}} \quad \ldots \quad \boldsymbol{h_{x_L}}\right) \tag{26}$$

and $\boldsymbol{q_c^a} = \left(q_{x_1}^a\ q_{x_2}^a\ \ldots q_{x_L}^a\right)^T$ is a shortened version of $\boldsymbol{q^a}$ that only has L dimensions. Also $\boldsymbol{C}$ is an $m\times L$ matrix of rank L, since $L<m$, and $\boldsymbol{h_{x_i}}$'s are linearly independent columns of $\boldsymbol{C}$ (recall our discussion in Section V-C). Therefore, $\boldsymbol{C}$ has a left Moore-Penrose inverse (pseudo inverse), $\boldsymbol{C}^+ = (\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T$ where $\boldsymbol{C}^+\boldsymbol{C} = \boldsymbol{I}$ of size $L\times L$. Thus, if we have knowledge of $\boldsymbol{C}$, we can then find $\boldsymbol{q_c^a}$ as:

$$\boldsymbol{q_c^a} = \boldsymbol{C}^+ \boldsymbol{y_s}. \tag{27}$$

The problem we need to solve is obtaining the matrix $\boldsymbol{C}$. We can solve this problem using an exhaustive search method which can be explained as follows:

(i) Candidate matrices $\boldsymbol{C_j}$ are generated by choosing different L combinations of columns of $\boldsymbol{H}$ where $1 \leq j \leq \binom{n_r}{L}$.

(ii) Find $\boldsymbol{q_{c_j}^a} = \boldsymbol{C_j^+}\boldsymbol{y_s} = \boldsymbol{C_j^+}\boldsymbol{C}\boldsymbol{q_c^a}$. Note that at this step, we obtain a vector $\boldsymbol{q_{c_j}^a}$ identical to $\boldsymbol{q_c^a}$ if and only if $\boldsymbol{C_j^+}\boldsymbol{C} = \boldsymbol{I} \Leftrightarrow \boldsymbol{C_j} = \boldsymbol{C}$.

(iii) Let $\boldsymbol{\beta}_j$ be such that

$$\boldsymbol{\beta}_j = \boldsymbol{C_j}\boldsymbol{q_{c_j}^a} = \boldsymbol{C_j}\boldsymbol{C_j^+}\boldsymbol{y_s}, \tag{28}$$

Hence, if the correct choice $\boldsymbol{C_j} = \boldsymbol{C}$ is made, then $\boldsymbol{\beta}_j = \boldsymbol{C_j}\boldsymbol{C_j^+}\boldsymbol{C}\boldsymbol{q_c^a} = \boldsymbol{y_s}$. Else, if $\boldsymbol{C_j} \neq \boldsymbol{C}$, then[6] $\boldsymbol{\beta}_j = \boldsymbol{C_j}\boldsymbol{C_j^+}\boldsymbol{C}\boldsymbol{q_c^a} \neq \boldsymbol{y_s}$

Hence, if $\boldsymbol{\beta}_{j^*} = \boldsymbol{y_s}$, we declare its corresponding matrix $\boldsymbol{C}_{j^*}$ the true matrix $\boldsymbol{C}$ defined in Eq. (26) which satisfies Eq. (25). Also, we have that $\boldsymbol{q_c^a} = \boldsymbol{q_{c_{j^*}}^a}$. Since, identifying $\boldsymbol{C}$ is equivalent to identifying the indexes $x_1, \ldots, x_L$. Thus, we found the angular channel $\boldsymbol{q^a}$ which is all zeros except - potentially[7] - for the components $q_{x_1}^a, \ldots, q_{x_L}^a$.

The previous discussion dealt with an idealized version of the measurements (i.e., $\boldsymbol{y_s}$), however, in practice, we observe $\boldsymbol{u_s}$ as an error-corrupted version of $\boldsymbol{y_s}$. Define $\boldsymbol{z_s}$ to be the error vector that captures the effect of both channel noise and quantization error which satisfies

$$\boldsymbol{u_s} = \boldsymbol{y_s} + \boldsymbol{z_s}. \tag{29}$$

Suppose that we know the matrix $\boldsymbol{C}$ for which we have

$$\boldsymbol{u_s} = \boldsymbol{C}\boldsymbol{q_c^a} + \boldsymbol{z_s}, \tag{30}$$

then, we can find $\boldsymbol{C}^+\boldsymbol{u_s}$ (compare to Eq. (27)) as follows

$$\boldsymbol{C}^+\boldsymbol{u_s} = \boldsymbol{C}^+\boldsymbol{C}\boldsymbol{q_c^a} + \boldsymbol{C}^+\boldsymbol{z_s} = \boldsymbol{q_c^a} + \boldsymbol{C}^+\boldsymbol{z_s},$$

to be a noise-corrupted version of $\boldsymbol{q_c^a}$.

Now, to find an estimate $\hat{\boldsymbol{q}}^{\boldsymbol{a}}$ of $\boldsymbol{q^a}$, we follow a very similar procedure to the one described before as follows:

(i) Matrices $\boldsymbol{C_j}$ are generated similar to the 1^{st} step before.

(ii) Define $\boldsymbol{E_j}$ to be the difference between $\boldsymbol{C}$ and $\boldsymbol{C_j}$ where

$$\boldsymbol{C} = \boldsymbol{C_j} + \boldsymbol{E_j}. \tag{31}$$

That is, $\boldsymbol{E_j} = \boldsymbol{0}$ (all zero matrix) $\Leftrightarrow \boldsymbol{C_j} = \boldsymbol{C}$.

(iii) Find $\boldsymbol{q_{c_j}^a}$ such that

$$\boldsymbol{q_{c_j}^a} = \boldsymbol{C_j^+}\boldsymbol{u_s} = \boldsymbol{q_c^a} + \boldsymbol{C_j^+}(\boldsymbol{E_j}\boldsymbol{q_c^a} + \boldsymbol{z_s}). \tag{32}$$

[6]Since $\boldsymbol{C_j^+}$ is the left pseudo-inverse of $\boldsymbol{C_j}$, and since $\boldsymbol{C_j}$ is not a square matrix, then $\boldsymbol{C_j}\boldsymbol{C_j^+} \neq \boldsymbol{I} = \boldsymbol{C_j^+}\boldsymbol{C_j} \implies \boldsymbol{C_j}\boldsymbol{C_j^+}\boldsymbol{C} \neq \boldsymbol{C}$

[7]This means that if the number of paths is less than L, then some $q_{x_i}^a$'s might have zero values as well.

Unlike the 2^{nd} step of the no-error case, $\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}_j}$ will not be identical to the true $\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}}$ with probability 1, since $\boldsymbol{z_s}$ is not identical to $\mathbf{0}$ with probability 1 ($\boldsymbol{z_s}$ is the difference between continuous and discrete quantities).

(iv) Let $\boldsymbol{\beta}_j$ be such that

$$\boldsymbol{\beta}_j = \boldsymbol{C_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}_j} = \boldsymbol{C_j}\left(\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} + \boldsymbol{C}^{+}_{\boldsymbol{j}}\left(\boldsymbol{E_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} + \boldsymbol{z_s}\right)\right) \quad (33)$$

$$= \begin{cases} \boldsymbol{y_s} + \boldsymbol{C_j}\boldsymbol{C}^{+}_{\boldsymbol{j}}\boldsymbol{z_s} & , \boldsymbol{C}=\boldsymbol{C_j} \\ \boldsymbol{C_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} + \boldsymbol{C_j}\boldsymbol{C}^{+}_{\boldsymbol{j}}\left(\boldsymbol{E_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} + \boldsymbol{z_s}\right) & , \boldsymbol{C}\neq\boldsymbol{C_j} \end{cases} \quad (34)$$

Then find j^* such that $j^* = \arg\min_j \left\|\boldsymbol{\beta}_j - \boldsymbol{u_s}\right\|$, where $\boldsymbol{\beta}_j - \boldsymbol{u_s}$ is given by

$$\boldsymbol{\beta}_j - \boldsymbol{u_s} = \boldsymbol{C_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} - \boldsymbol{y_s} + \boldsymbol{C_j}\boldsymbol{C}^{+}_{\boldsymbol{j}}\boldsymbol{E_j}\boldsymbol{q}^{\boldsymbol{a}}_{\boldsymbol{c}} + \left(\boldsymbol{C_j}\boldsymbol{C}^{+}_{\boldsymbol{j}} - \boldsymbol{I}\right)\boldsymbol{z_s} \quad (35)$$

which at $\boldsymbol{C} = \boldsymbol{C_{j^*}}$ is further reduced to

$$\boldsymbol{\beta}_{j^*} - \boldsymbol{u_s} = \left(\boldsymbol{C_j}\boldsymbol{C}^{+}_{\boldsymbol{j}} - \boldsymbol{I}\right)\boldsymbol{z_s} \quad (36)$$

E. Multiple Transmit and Receive Antennas

So far, we have considered channels with single transmit antennas and shown how to perform beam discovery at RX. To extend our approach to a general setting, we consider channels with n_t antennas at TX, and n_r antennas at RX. Thus, instead of the TX just sending signals omnidirectionally, now it can perform highly directional transmission. Recall that the RX is able to perform channel measurements using multi-armed beams. Similarly, the TX can send signals using multi-armed beams to simultaneously focus on multiple directions using precoding vectors $\boldsymbol{f_j}$.

The design of precoding vectors can also be obtained using an LBC approach. Similar to the method of designing rx-combining vectors $\boldsymbol{w_i}$, we look for an LBC, C_2, that has CWs of length $n_2{=}n_t$ and can correct for $e_n{=}L$ errors. Let the parity check matrix of C_2 be $\boldsymbol{H}_2$, using which, we will design the precoding vectors $\boldsymbol{f_j}$. Let *beam$_{tx}$#i* denote the i^{th} TX beam which points to TX direction *dir$_{tx}$ #i*. Then, just as before, we envisage $\boldsymbol{H}_2$ as an array whose columns are associated with resolvable TX directions such that: i) its j^{th} column corresponds to *dir$_{tx}$#j*, and ii) its i^{th} row corresponds to the i^{th} measurement. We note that no actual measurements are performed at TX; we use the word *measurement* to refer to precoding, consistent with the case of RX. That is, the i^{th} TX measurement is actually the i^{th} precoder $\boldsymbol{f}_i$. Thereby, we design the i^{th} precoder as a multi-armed TX beam such that, only if $h_{i,j}$, the intersection of the i^{th} row and j^{th} columns of $\boldsymbol{H}_2$, is $= 1$, do we include sub-beam *beam$_{tx}$#j* in $\boldsymbol{f_i}$. Each TX measurement provides a component in a TX channel syndrome vector $\boldsymbol{y}^{\boldsymbol{TX}}_{\boldsymbol{s}}$. The total number of TX measurements (i.e., precoding vectors), denoted by m_2, is equal to the number of parity check bits of the code C_2. That is, $m_2{=}n_2{-}k_2$, where k_2 is the length of C_2's information sequences. To obtain AoDs of strong paths at TX, we define the function $\xi_2()$ as the mapping function between all possible TX channel syndromes and their corresponding angular channels denoted by $\boldsymbol{q}^{\boldsymbol{aTX}}$. Note that, for every *dir$_{rx}$#i*, there exists a corresponding $\boldsymbol{q}^{\boldsymbol{aTX(i)}}$ which represents the i^{th} row of $\boldsymbol{Q^a}$. Also, since the maximum number of clusters is L, then, the number of non-zero vectors $\boldsymbol{q}^{\boldsymbol{aTX(i)}}$ is $\leq L$.

To see the whole picture, assume that a code C_1, with CWs of length $n_1{=}n_r$, is an LBC code associated with beam discovery at RX side. Let the number of RX measurements, i.e., the number of rx-combining vectors, be m_1 such that $m_1{=}n_1{-}k_1$, where k_1 is the length of information sequences of C_1. Also let $\xi_1()$ be the mapping function between RX channel syndromes and its corresponding angular channel. Under this setting, the beam discovery problem is performed as follows: i) The TX starts sending its training sequence using the precoder $\boldsymbol{f_j}, \forall j \in \{0, \ldots, m_2{-}1\}$. ii) The RX performs m_1 channel measurements while $\boldsymbol{f_j}$ is being used at TX and obtains a channel syndrome $\boldsymbol{y_{s_j}}$. iii) Based on $\boldsymbol{y_{s_j}}$, the RX obtains a corresponding channel, $\boldsymbol{q^{a(j)}}$ with path components $\{q_p^{a(j)}\}_{\forall p \in \{1,\ldots,n_r\}}$. Notice that $\boldsymbol{q^{a(j)}}$'s do not necessarily represent individual path gains, but rather, combinations of paths accumulating at a single *dir$_{rx}$#*. Therefore, there exists a resemblance to channel syndromes which we exploit. iv) We construct a set of n_r TX channel syndromes, $\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}}$ where their j^{th} component $y^{TX(p)}_{s_j}{=}q_p^{a(j)}$, i.e., $[\boldsymbol{y}^{\boldsymbol{TX(1)}}_{\boldsymbol{s}}, \boldsymbol{y}^{\boldsymbol{TX(2)}}_{\boldsymbol{s}}, \ldots, \boldsymbol{y}^{\boldsymbol{TX(n_r)}}_{\boldsymbol{s}}] = [\boldsymbol{q^{a(1)}}, \boldsymbol{q^{a(2)}}, \ldots, \boldsymbol{q^{a(m_2)}}]^T$. v) Finally, we find the angular TX channel for every *dir$_{rx}$ #p*, i.e., p^{th} row of $\boldsymbol{Q^a}$, using the mapping function $\boldsymbol{q^{aTX(p)}}{=}\xi_2(\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}})$. Notice that, since no more than $L \ll n_r$ clusters exist, and since $\mathbf{0}$ channels correspond to $\mathbf{0}$ channel syndromes, we only need to apply $\xi_2()$ a maximum of L times -unless measurement error occurs. This whole process is highlighted in Algorithm 1.

Remark (Sufficiency of Syndromes): Recall that finding $\boldsymbol{Q^a}$ from $\{\boldsymbol{y_{s_j}}\}_{\forall j}$ is done using two stages of unique mappings; First: For every $\boldsymbol{y_{s_j}}$ we find $\boldsymbol{q^{a(j)}}$ using $\xi_1()$. Second: After constructing the vectors $\{\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}}\}_{\forall 1 \leq p \leq n_r}$ from $\{\boldsymbol{q^{a(j)}}\}_{\forall 0 \leq j \leq m_1 - 1}$ (using a simple matrix transpose operation), we map $\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}}$ using $\xi_2()$ to $\boldsymbol{q^{aTX(p)}}$ where the latter constitute the columns of $\boldsymbol{Q^a}$. The sufficiency of the syndromes $\{\boldsymbol{y_{s_j}}\}_{\forall j}$ for estimating $\boldsymbol{Q^a}$ can be easily proven by showing that $\{\boldsymbol{y_{s_j}}\}_{\forall j}$ are sufficient to estimate $\{\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}}\}_{\forall p}$, and that $\{\boldsymbol{y}^{\boldsymbol{TX(p)}}_{\boldsymbol{s}}\}_{\forall p}$ are sufficient to estimate $\{\boldsymbol{q^{aTX(p)}}\}_{\forall p}$. Both of these steps are similar to the single-transmit multiple-receive antenna scenario discussed in Section V-C.

VI. Error Correction

So far, the main focus of this paper has been on beam discovery using a "small" number of measurements. We have not evaluated the detection error under noisy observations. In fact, with no measurement errors, our proposed solution can estimate the channel matrix perfectly. However, the presence of channel noise and quantization degrades the beam discovery performance. The key question we investigate here is whether increasing the number of channel measurements—by essentially adding redundancy—would improve the beam detection performance. The answer to this is: **Yes**. In the sequel we will present a method that allows for increasing the number

Algorithm 1: Beam Discovery of Multiple TX/RX Antennas

input : $\{\boldsymbol{w}_i\}_{\forall i\in\{1,\ldots,m_1\}}$, $\{\boldsymbol{f}_j\}_{,\forall j\in\{1,\ldots,m_2\}}$, $\xi_1() : \boldsymbol{y_s} \rightarrow \widehat{\boldsymbol{q}}^a$, $\xi_2() : \boldsymbol{y_s^{TX}} \rightarrow \hat{\boldsymbol{q}}^{a^{\boldsymbol{TX}}}$

output: $\{\boldsymbol{y_{s_i}}\}_{\forall i\in\{1,\ldots,m_2\}}$

begin
  $j = 0$;
  while $j < m_2$ **do**
    $i = 0$;
    while $i < m_1$ **do**
      // channel measurement
      $y_{s_{i,j}} = \boldsymbol{w}_i^H \boldsymbol{Q} \boldsymbol{f}_j s$;
      $i \leftarrow i + 1$;
    end
    // construct channel syndrome $\boldsymbol{y_{s_j}}$
    $\boldsymbol{y_{s_j}} \leftarrow \{y_{s_{i,j}}\}_{\forall i\in\{1,\ldots,m_1\}}$;
    /* find corresponding channel $\boldsymbol{q^{a(j)}} = [q_1^{a(j)}, q_2^{a(j)}, \ldots, q_{n_r}^{a\ (j)}]^T$ */
    $\boldsymbol{q^{a(j)}} \leftarrow \xi_1(\boldsymbol{y_{s_j}})$;
    for $p \leftarrow 1$ **to** n_r **do**
      /* construct TX channel syndromes $\boldsymbol{y_s^{TX(p)}}$, where $\boldsymbol{y_s^{TX(p)}} = [y_{s_1}^{TX(p)}, y_{s_2}^{TX(p)}, \ldots, y_{s_{m_2}}^{TX(p)}]^T$ */
      $y_{s_j}^{TX(p)} \leftarrow q_p^{a(j)}$
    end
    $j \leftarrow j + 1$;
  end
  for $p \leftarrow 1$ **to** n_r **do**
    $\boldsymbol{q^{aTX(p)}} \leftarrow \xi_2(\boldsymbol{y_s^{TX(p)}})$
  end
  $\widehat{\boldsymbol{Q}}^{\boldsymbol{a}} = \left(\boldsymbol{q^{aTX(1)}}\ \boldsymbol{q^{aTX(2)}}\ \ldots\ \boldsymbol{q^{aTX(n_r)}}\right)^T$
end

of measurements and trades it for higher reliability. The very concept of adding redundant information to combat noisy observations is the foundation of channel coding. Hence, it is appealing to use channel coding ideas to achieve more reliable beam discovery. For simplicity we again present our proposed solution for the simple setting of one transmit antenna and multiple receive antennas. The general multiple transmit and receive antennas setting can be dealt with in the same fashion described in Section V-E.

Recall that a received symbol u_s is given by Eq. (12) as $u_s{=}[y_s + \boldsymbol{w}^H\boldsymbol{n}]_+$ where $y_s{=}\boldsymbol{w}^H\boldsymbol{Q}\boldsymbol{f}s$ is the error-free measurement symbol. We write $u_s{=}y_s + z_s$ where z_s is the measurement error (Eq. (29)). Also recall that, for $\boldsymbol{f}{=}1$ (one transmit antenna), and $\boldsymbol{w_i}$, we form the channel syndrome vector $\boldsymbol{u_s}{=}[u_{s_0}\ u_{s_1}\ \ldots\ u_{s_{m-1}}]^T$ such that $u_{s_i}{=}\boldsymbol{w_i}^H\boldsymbol{q}s{+}z_{s_i}$ where $\boldsymbol{q}$ is the $n_r{\times}1$ channel vector. Equivalently, we have that $\boldsymbol{u_s}{=}\boldsymbol{H}\boldsymbol{q^a}{+}\boldsymbol{z_s}$ where $\boldsymbol{z_s}$ is formed by stacking $\{z_{s_i}\}_{\forall i=0,\ldots,m-1}$. Recall that this is exactly Eq. (14) but with the noise terms added.

In fact, we can perceive the channel syndrome $\boldsymbol{y_s}$ as raw information sequence that need to be transmitted over a noisy channel, and $\boldsymbol{u_s}$ is the noise-corrupted received sequence. The syndrome, $\boldsymbol{y_s}$, is a vector that lies in an $m-$dimensional vector space. By exploiting channel codes, we can map $\boldsymbol{y_s}$ to longer sequences $\boldsymbol{y_s^\nu}$ (encoded channel syndrome) that lie in an $m-$dimensional subspace of an m_c-dimensional vector space. The longer sequences $\boldsymbol{y_s^\nu}$ should have increased distance which allows for higher resilience against measurement errors. Hence, $\boldsymbol{u_s}$ can now be written as $\boldsymbol{u_s} = \boldsymbol{y_s^\nu} + \boldsymbol{z_s}$. Our goal is to design $\boldsymbol{y_s^\nu}$. Once we achieve that, the rest of the problem can be tackled as discussed in section V.

Towards that end, let us use an error correction code C_c, with generator matrix $\boldsymbol{G_c}$ and error correction capability e_c. Note that we use the subscript c to refer to *correction*. The size of $\boldsymbol{G_c}$ is $m{\times}m_c$. Thus, the encoded channel syndromes can be represented as $\boldsymbol{y_s^\nu} = \boldsymbol{G_c}^T\boldsymbol{y_s}$, where $\boldsymbol{y_s}$ and $\boldsymbol{y_s^\nu}$ are of sizes $m{\times}1$ and $m_c{\times}1$, respectively. Thus, $\boldsymbol{y_s^\nu}$ can be written as $\boldsymbol{y_s^\nu} = \boldsymbol{G_c}^T\boldsymbol{H}\boldsymbol{q^a}$. Then, similar to Eqn. (14) we want to use the matrix $\boldsymbol{G_c}^T\boldsymbol{H}$ to design $\boldsymbol{y_s^\nu}$. However, the problem here is that this matrix in not necessarily a binary matrix (i.e., with elements of '1's and '0's). Hence, let us denote by $\boldsymbol{H^\nu}$, the matrix $\boldsymbol{G_c}^T\boldsymbol{H} \pmod 2$ and use it to design $\boldsymbol{y_s^\nu}$ such that

$$\boldsymbol{y_s^\nu} = \boldsymbol{H^\nu}\boldsymbol{q^a}. \tag{37}$$

In other words, $\boldsymbol{H^\nu}{=}\boldsymbol{G_c}^T\boldsymbol{H}$ is the matrix product over $GF(2)$. Therefore, instead of designing the channel measurements based on $\boldsymbol{H}$, we propose to design them based on $\boldsymbol{H^\nu}$ with that being the only difference to the design proposed earlier.

At this point, it remains to show that the new measurements design still provides a one-to-one mapping to every angular channel (i.e., if $\boldsymbol{q_1^a} \neq \boldsymbol{q_2^a}$, then their corresponding channel syndromes $\boldsymbol{y_{s1}^\nu} \neq \boldsymbol{y_{s2}^\nu}$). Furthermore, we will show that the new design provides a better resilience to measurement errors. That is, we will show that if $\boldsymbol{q_1^a} \neq \boldsymbol{q_2^a}$, then $\delta(\boldsymbol{y_{s1}}, \boldsymbol{y_{s2}}) \leq \delta(\boldsymbol{y_{s1}^\nu}, \boldsymbol{y_{s2}^\nu})$, where $\delta(\cdot,\cdot)$ is defined as in Eq. (21).

A. Sufficient Statistic

We start off by showing that the new measurements provide a sufficient statistic for beam discovery. We will follow a similar approach to that of Section V-C. Specifically, we will first consider error patterns, matrices, and operators over the finite field $GF(2)$. Afterwards, we will extend those concepts to the complex field where all channel matrices and measurements lie.

Let us consider a code C, with codewords of length n and error correction capability e_n. The parity check and generator matrices of C are given by $\boldsymbol{H}$ and $\boldsymbol{G}$, respectively. The error syndromes of C are given by $\boldsymbol{s} = \boldsymbol{r}\boldsymbol{H}^T = \boldsymbol{e}\boldsymbol{H}^T$, where $\boldsymbol{r}$ is the received sequence and e is the error pattern corrupting the transmitted codeword $\boldsymbol{c}$ (recall Footnote 5). Now suppose we encode $\boldsymbol{s}$ using another error correction code C_c. The parity check and generator matrices of C_c are given by $\boldsymbol{H_c}$ and $\boldsymbol{G_c}$, respectively. The encoded syndromes $\boldsymbol{s^\nu}$ are given as

$$\boldsymbol{s^\nu} \overset{(a)}{=} \boldsymbol{s}\boldsymbol{G_c} \overset{(b)}{=} \boldsymbol{e}\boldsymbol{H}^T\boldsymbol{G_c} \overset{(c)}{=} \boldsymbol{e}\boldsymbol{H^\nu}^T, \tag{38}$$

Consider all single bit error patterns $\boldsymbol{e}^{(i)}$ and let the encoded syndrome that corresponds to $\boldsymbol{e}^{(i)}$ be $\boldsymbol{s^{\nu(i)}} = \boldsymbol{e}^{(i)}\boldsymbol{H^\nu}^T$. Thus, $\boldsymbol{s^{\nu(i)}}$ is exactly the i^{th} row of $\boldsymbol{H^\nu}^T$, i.e., i^{th} column of $\boldsymbol{H^\nu}$.

Lemma 3: For any error sequence $\boldsymbol{e_t}$ with number of bit errors identical to e_n, its encoded syndrome $\boldsymbol{s_t}^{\boldsymbol{\nu}}$ is a linear combination of e_n linearly independent vectors $\boldsymbol{s}^{\boldsymbol{\nu}(i)}$.

The proof of this lemma is similar to that of Lemma 1, and hence, is omitted here (see [34] for proof).

Lemma 3 allows us to use the result of Lemma 2 which states that if we have a collection, $\left\{\boldsymbol{s}^{\boldsymbol{\nu}(i)}\right\}_{i=x_1}^{x_{e_n}}$, of linearly independent vectors over $GF(2)$. Then, if their $'1'$ and $'0'$ entries are interpreted as real numbers, then they are also linearly independent over $\mathbb{C}$.

Let us interpret the elements of $\boldsymbol{H^{\nu}}$ and $\boldsymbol{e^{(i)}}$ as real numbers. Then, we can write the channel $\boldsymbol{q^a}$ as

$$(\boldsymbol{q}^a)^T = \alpha_1 \boldsymbol{e}^{(1)} + \alpha_2 \boldsymbol{e}^{(2)} + \cdots + \alpha_n \boldsymbol{e}^{(n)} \tag{39}$$

where $\alpha_i \in \mathbb{C}$ and $\sum_{i=1}^{n} \mathbb{1}_{\{\alpha_i \neq 0\}} \leq e_n$. Therefore, each encoded channel syndrome $(\boldsymbol{y_s^{\nu}})^T = (\boldsymbol{q}^a)^T \boldsymbol{H^{\nu}}^T \implies \boldsymbol{y_s^{\nu}} = \boldsymbol{H^{\nu}} \boldsymbol{q}^a$ is a linear combination of independent vectors in $\mathbb{C}^{m_c}$ (columns of $\boldsymbol{H^{\nu}}$). Therefore, for all measurable channels $\boldsymbol{q_1^a} \neq \boldsymbol{q_2^a}$, we have that $\boldsymbol{y_{s1}^{\nu}} \neq \boldsymbol{y_{s2}^{\nu}}$. Therefore, measurements designed based on $\boldsymbol{H^{\nu}}$ are sufficient for beam discovery.

B. Resilience to Errors

We are going to show that the encoded syndromes $\boldsymbol{y_s^{\nu}}$ are more tolerant to the occurrence of measurement errors. Since the mapping functions $\xi()$ finds the correct $\boldsymbol{q^a}$ using l^2-norm minimization methods (this is true for both look-up table and search methods), then it is intuitively beneficial to separate the channel syndrome vectors, in the l^2-norm sense, as much as possible. Thus, we want to show that if two channel syndromes $\boldsymbol{y_{s_1}}$ and $\boldsymbol{y_{s_2}}$ (corresponding to channel vectors $\boldsymbol{q_1^a}$ and $\boldsymbol{q_1^a}$) have distance $\delta(\boldsymbol{y_{s_1}}, \boldsymbol{y_{s_2}})$, then their corresponding encoded syndromes are such that

$$\delta(\boldsymbol{y_{s_1}^{\nu}}, \boldsymbol{y_{s_2}^{\nu}}) \geq \delta(\boldsymbol{y_{s_1}}, \boldsymbol{y_{s_2}}) \tag{40}$$

$$\iff \quad \left\|\boldsymbol{y_{s_1}^{\nu}} - \boldsymbol{y_{s_2}^{\nu}}\right\| \geq \left\|\boldsymbol{y_{s_1}} - \boldsymbol{y_{s_2}}\right\| \tag{41}$$

Proposition 4: Let $\boldsymbol{G_c}$ be the generator matrix of some LBC code C. Then, if $\boldsymbol{G_c}$ is represented in the standard form and $\boldsymbol{H^{\nu}}$ is generated using $\boldsymbol{G_c}$ (as $\boldsymbol{G_c^T H} \pmod 2$), then $\left\|\boldsymbol{y_{s_1}^{\nu}} - \boldsymbol{y_{s_2}^{\nu}}\right\| \geq \left\|\boldsymbol{y_{s_1}} - \boldsymbol{y_{s_2}}\right\|$.

Proposition 4 shows that by using any appropriate systematic code, we obtain encoded channel syndromes, $\boldsymbol{y_s^{\nu}}$, that have greater l^2-distance than the original syndromes $\boldsymbol{y_s}$. Hence, we increase the space allowed for the noise-corrupted measurement vector $\boldsymbol{u_s}$ to lie in, while still being able to identify its true corresponding, error-free, channel syndrome. The proof of Proposition 4 can be found in [34] and is omitted here due to space limitation.

VII. Performance Evaluation

A. Simulation Setup and Parameters

We consider an $n_r \times n_t$ mm-wave channel with carrier frequency $f_c = 73$GHz and average path loss $\mu = 136$dB. A channel path is assumed to exist if its path attenuation is not higher than 14dB above the average path loss, i.e. total path loss is at most 150dB. These values are based on the measurement results carried out in New York City and published in [2, Table I]. Let L be the **maximum** number of (strong) paths that exist between TX and RX.[8]

Let τ be the time duration of a pilot sequence of one measurement. For simplicity, let $\tau=1$. Also, recall that *SNR* is defined for a single path (see equation (13)), and that P is the corresponding transmitted power (i.e., per path). Let the total transmitted power be P_t, where P_t is an integer multiple of P that depends on the number of combined transmit and receive directions (recall Fig. 2). Then, the total energy required for beam discovery is $E = mP_t\tau = mP_t$, where m is the total number of measurements.[9] Let the normalized energy be $E_t \triangleq \frac{E}{N_0}|\alpha_{\min}/\mu|^2$.

To map the channel measurements to their corresponding angular channels, we use the search method presented in Section V-D.2. Finally, for every simulation scenario, we obtain the average performance across 10^5 runs.

B. Performance metrics

To asses the performance of the proposed beam discovery method, we mainly focus on three basic criteria, namely, ***accuracy of beam discovery***, ***number of measurements***, and ***accuracy of path gain value estimates***. To that end, we use the following performance metrics:

i) ***Number of measurements:*** This represents the number of pilots sent from TX to discover the paths to RX.

ii) ***Probability of strongest k beams discovery:*** This denotes the probability of correctly identifying the directions of k strong reflectors among L. There are two cases we consider pertaining to the possibility of the algorithm identifying exactly k directions or more than k directions:
1) perfect k beam discovery: where exactly k true paths are discovered with no incorrect paths among them.
2) all k beam discovery: where k true paths are discovered with potentially more incorrectly identified paths.

iii) ***Number of incorrect beams:*** Due to the possibility of obtaining a combination of correct and incorrect paths, it is important that we have as few incorrect beams as possible since further refinement would be made easier.

iv) ***Normalized mean squared error (MSE):*** $\frac{\left\|\boldsymbol{Q^a} - \widehat{\boldsymbol{Q}}^{\boldsymbol{a}}\right\|_F^2}{\|\boldsymbol{Q^a}\|_F^2}$. Measurement errors occur in the form of 1) imperfect estimates of path gains and phases, and 2) incorrect beam discovery. Hence, MSE provides an inclusive metric for how close the estimated channel matrix is to the true one.

[8]Different channel environments are expected to have different values of L. For instance, in an urban environment (New York City), [2] shows that a maximum number of 4 paths exist while for indoor environments like offices/corridors/conference rooms, [33] shows that ≤ 5 paths exist. Moreover, if we are only interested in LoS communication (e.g., IEEE 802.11ad), then we can reasonably assume that $L=1$ since the path loss of non-LoS (NLoS) paths is significantly higher than that of LoS paths (≈ 30 dB higher [2]). If more than L paths exist, our approach will only strive to find the L strongest paths, with weaker paths having an effect similar to measurement errors.

[9]This formula for total energy assumes equal P_t for all measurements. Depending on the employed LBC, this might not always be the case. More generally, we can find the total energy to be: $E = \sum_i m_i P_{t_i}$, where m_i is the number of measurements with total transmit power P_{t_i}.

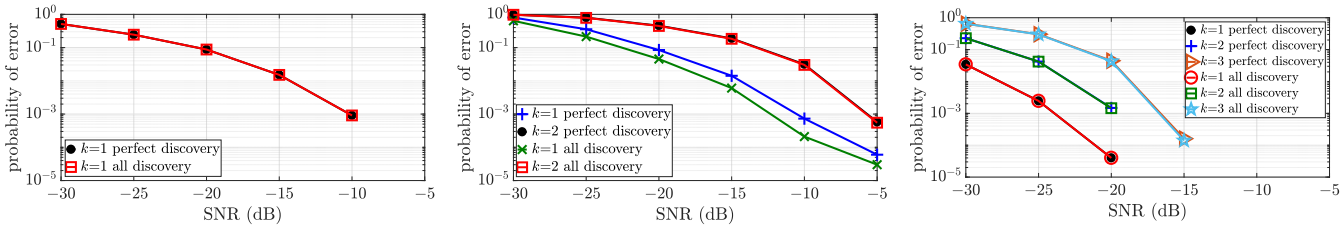

Fig. 4. Beam detection probability. (a) 15×15 channel with $L=1$. (b) 8×8 channel with $L=2$. (c) 32×32 channel with $L=3$.

Measurement errors mainly occur due to two contributing factors. The first is ***measurement noise***, and the second is ***quantization*** (recall that we assume the measurements to be quantized using mid-tread ADC quantizers with 2^b+1 levels). In Sections VII-C and VII-D, we assess the performance of Beam Discovery approach against ***only*** the effect of measurements noise. We do so by assuming a perfect, infinite resolution ADC. Then, in Section VII-E, we investigate the system performance at different ADC resolution levels. This separate investigation of sources of errors allows us to understand how each source affects the performance. Thus, enabling full realization of potential gains of Beam Discovery approach.

C. Single-path channels

Consider a 15×15 mm-wave channel with $L=1$ path between TX and RX. Hence, the parity check matrix of $(15, 11, 3)$ Hamming code can be used to design both the precoders, $\boldsymbol{f_j}$, and rx-combiners, $\boldsymbol{w_i}$, i.e., $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$, are identical. Hence, we need a number of TX measurements m_1, which is identical to the number of RX measurements $m_2=15-11=4$. Hence, the total number of measurements is $m=16$. On the other hand, the exhaustive scanning method requires 225 measurements to inspect every possible TX-RX beam combination. Thus, our approach results in $\approx92.8\%$ reduction in the required number of measurements.

We plot the ***probability of error*** vs. *SNR* for: i) *Perfect* beam discovery where only the single strongest path ($k=1$) is correctly identified, and ii) *all* beam discovery where the strongest path is correctly identified among potentially other misidentified paths. Fig. 4 shows those curves. We observe that both curves are on top of each other which indicates that the strongest path is either correctly detected or is completely missed. Moreover, at all *SNR* values ≥ -5dB, the probability of error is lower than 10^{-5}, and hence, is not shown here since the shown figures are the averages of 10^5 simulation runs.

In Fig. 5, we plot the normalized mean squared error of the channel estimate $\widehat{\boldsymbol{Q}}^{\boldsymbol{a}}$. The very high values at low signal to noise ratios indicate that $\widehat{\boldsymbol{Q}}^{\boldsymbol{a}}$ has large components at truly zero components in $\boldsymbol{Q}^{\boldsymbol{a}}$ and/or large components in $\boldsymbol{Q}^{\boldsymbol{a}}$ are not represented in $\widehat{\boldsymbol{Q}}^{\boldsymbol{a}}$. Nevertheless, MSE drops steadily fast as *SNR* increases; indicating improved channel estimation.

When we talk about the possibility of misidentified beams for the all beam discovery metric, it is crucial to have a small number of incorrect beams which would facilitate further refinement. Interestingly, for this scenario, since the error performance of perfect and all beam discovery are the same, we do not have any misidentified paths besides the correct one. Nevertheless, this is not always the case as we will see in further investigated scenarios.

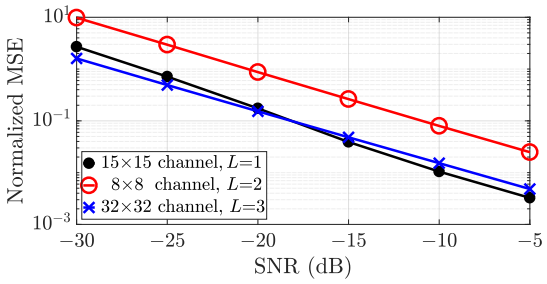

Fig. 5. Normalized mean squared error (MSE).

D. Multi-path Channels

First, consider an 8×8 channel with $L=2$ paths. For this scenario, we use an $(8, 2, 5)$ code for both $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$. With this code, a total number, 36, of measurements is needed for beam discovery. Compared with the 64 measurements needed for exhaustive scanning, we achieve $\approx43.7\%$ reduction in the number of measurements under this scenario.

Since we investigate a channel that potentially has two strong paths, we evaluate the probability of error of picking one correct strong path ($k=1$) as well as picking two strong paths ($k=2$). Fig. 4b depicts the corresponding probability of error of the perfect and all k beam discovery metrics. Unlike single-path channels, there exists a wider gap between perfect and all beam discovery curves for the $k=1$ metric; which indicates higher vulnerability to picking incorrect paths. On the other hand, for $k=2$, the error performance of the perfect and all beam discovery metrics are almost on top of each other. In Fig. 5, similar trend for normalized MSE is obtained where MSE steadily drops as *SNR* increases.

Recall that in the 15×15 single-path channel investigation, no incorrect paths were obtained alongside correctly identified strong paths. This behavior is not replicated for the 8×8 channel under investigation. For instance, at -10dB we obtain a maximum of 2 misidentified paths. Further, the probability of obtaining incorrect paths at -10dB is ≈ 0.04637.

We further consider a larger array with dimensions 32×32 and $L=3$ paths. We use a $(32, 16, 8)$ Reed-Muller code to design both $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$. This corresponds to $m_1=m_2=16$, i.e., total number of measurements $m=265$. This is 75% fewer measurements needed compared to exhaustive scanning which requires 1024 measurements for beam discovery.

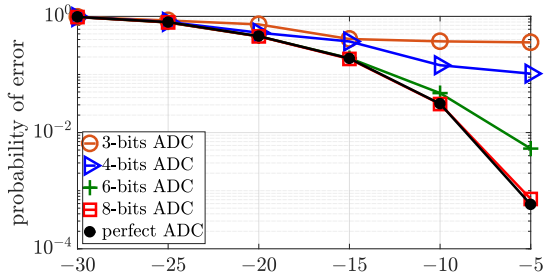

Fig. 6. Perfect k=2 beam discovery at different quantization resolution (8×8 channel with L=2).

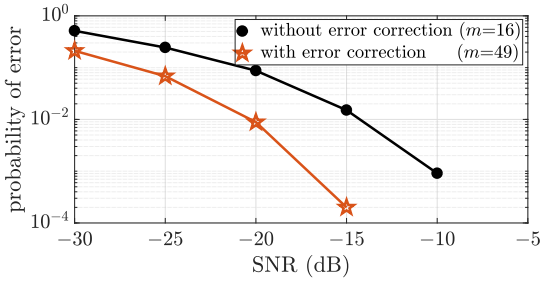

Fig. 7. Beam detection probability (15×15 channel with L=1).

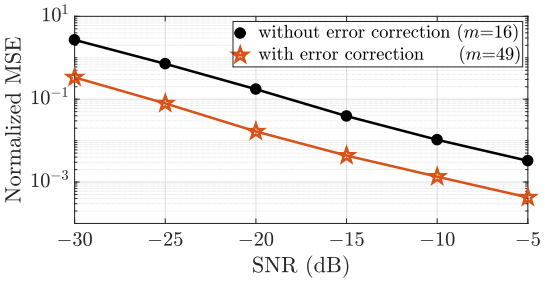

Fig. 8. MSE (15×15 channel with L=1)

The probability of error for perfect and all k=1, 2, 3 beam discovery are shown in Fig. 4c. We notice a faster decay rate for the probability of error. This behavior is due to the higher gain of the TX and RX antenna arrays; which increases the receive signal to noise ratios compared to small arrays. The normalized MSE is shown in 5 have similar trend to the previous investigated scenarios.

E. Effect of Quantization

In this section, both sources of errors are incorporated. Specifically, we analyze the system performance at different ADC resolution levels. We will show that very low resolution ADCs can have detrimental effect on performance. Thus, a natural question that we try to answer in this study is: ***How far should we increase the resolution of quantizers in order to unlock the full potential of the Beam Discovery approach?***

Recall that we use mid-tread ADCs with 2^b+1 quantization levels (b stands for the number of bits required to represent the ADC output (approximately)). We limit our discussion to the case of 8×8 channels with L=2 paths since its results are representative of the other previously investigated scenarios. For clarity and legibility of figures, we only plot the perfect k=2 beam discovery for $b = 3, 4, 6, 8$ bits i.e., the corresponding number of quanization levels is $9, 17, 65, 257$, respectively. We also plot the corresponding probability of error using a perfect ADC (i.e., $b \to \infty$). These curves are shown in Fig. 6.

We find that, at b=3, the probability of error is very high and does not improve with increasing *SNR*. Hence, quantization is the dominant source of errors. Then, as the resolution of ADCs increase, significant performance improvement can be achieved. For instance, while b=4 still do not produce very good probability of error (with increasing *SNR*), a huge leap in performance can be obtained using ADCs with only b=6 bits. Moreover, at b=8, we approach the performance of perfect ADCs. Recall that only 2 ADCs are required (see Fig. 3).

F. Error Correction

In this section, we investigate the performance of Beam Discovery with the ***error correction*** technique proposed in SectionVI. Recall that error correction is a *channel-coding-like* technique that allows for improving the error performance on the expense of increased number of measurements.

1) Single-Path Channel: Consider the 15×15 single-path channel we studied in Section VII-C. Recall that we used the parity check matrix of $(15, 11, 3)$ Hamming code for both $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$ which resulted in syndromes $\boldsymbol{y_s}$ of length m_1=m_2=4. Now, we need to encode sequences of length 4 into longer sequences $\boldsymbol{y_s^\nu}$ using a systematic code. Conveniently, we can use the $(7, 4, 3)$ Hamming code which maps sequences of length 4 into sequences of length 7. The corresponding $\boldsymbol{H_1^\nu}$ and $\boldsymbol{H_2^\nu}$ matrices are of size 7×15 and we have that m_{c_1}=m_{c_2}=7. Hence we have a total number of measurements for Beam Discovery with error correction m_c=49. This is ≈78.2% fewer measurements compared to exhaustive scanning. Recall that the number of measurements without error correction is 16.

The probability of error for perfect k=1 beam discovery is depicted in Fig. 7. A notable performance improvement over the m=16 case is obtained. That is, at the same *SNR*, significantly lower probability of error is achieved. This performance improvement is also reflected in the MSE curves in Fig. 8.

2) Multi-Path Channel: We study 32×32 channels with L= 3 paths. Recall that, in SectionVII-D, we use a $(32, 16, 8)$ Reed-Muller code for which the parity check matrices $\boldsymbol{H_1}$=$\boldsymbol{H_2}$ are of size 16×32. Under this setting we obtain 75% reduction in the number of channel measurements compared to exhaustive scanning (256 instead of 1024 measurements). To add the error correction capability, we encode the channel syndromes using a $(21, 16, 3)$ code (a subcode of the $(31, 26, 3)$ Hamming code). We obtain $\boldsymbol{H_1^\nu}$=$\boldsymbol{H_2^\nu}$ of size 21×32. Thus, m_{c_1}=m_{c_2}=21 (m_c=441), which gives a reduction of ≈ 57% in number of measurements compared to exhaustive scanning.

For clarity, we only plot the probability of error for perfect k=1, 2, 3 beam discovery shown in Fig. 9. Note that at m_c=441, the k=1 perfect beam discovery achieves error probability below 10^{-5}, hence, it is not shown in Fig. 9. We notice a huge performance improvement over the m=265 case, that is, at fixed *SNR* we obtain at least an order of magnitude improvement in the probability of error. We also obtain a corresponding improvement in MSE (figure omitted, see [34]).

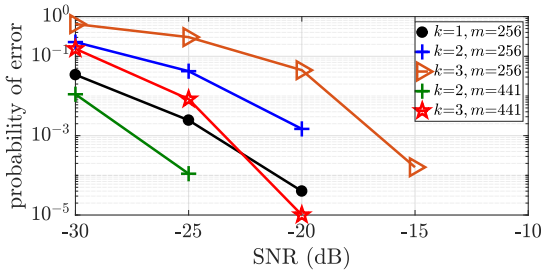

Fig. 9. Beam detection probability (32×32 channel with L=3).

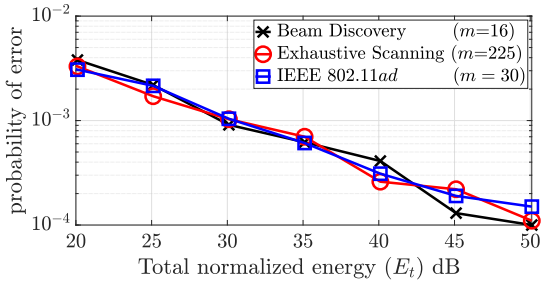

Fig. 10. Perfect k=1 beam discovery. Comparison with Scanning / 802.11ad (15×15 channel with L=1).

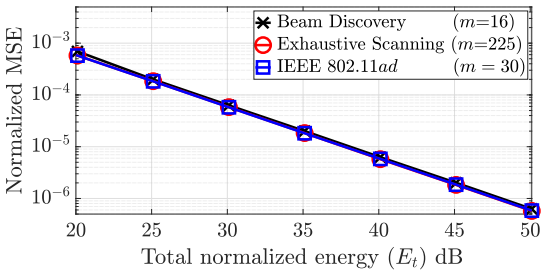

Fig. 11. Normalized MSE. Comparison with Scanning / 802.11ad (15×15 channel with L=1).

G. Comparison to Exhaustive Scanning and IEEE 802.11ad

We compare the performance of our proposed beam discovery approach against i) the Exhaustive Scanning method, and ii) the channel estimation of IEEE 802.11ad. ***In order to have a fair comparison, we adjust the transmission power of every scheme such that the total consumed energy for the measurement phase is identical for all schemes***. Note that at equal transmission power, each one of those three schemes would operate at a different SNR value. Moreover, due to the different number of measurements required by each scheme, the total transmit energy would also be different. To understand why this happens, recall that each one of these approaches operate with distinct transmit/receive antenna beam patterns, with different beam-widths. As the beam-width becomes wider, the received SNR decreases.[10] In other words, variations in beam-width causes the SNR to change. Since our proposed approach requires fewer measurements than Exhaustive scanning, we can increase the transmission power while still keeping the total transmit energy unchanged. The same is true for the IEEE 802.11ad channel estimation. Finally, we assume that perfect ADCs are used for all schemes.

Due to space limitation, we limit our discussion to only 15×15 single-path channels. In Fig. 10, we plot the probability of error for perfect k=1 beam discovery for all three schemes. The range used for the normalized energy, E_t, corresponds to SNR of ≈ -5 to 25dB for the Exhaustive Scanning method. We find that our approach achieves almost the same error performance as both Scanning and 802.11ad, yet, with 92.8% and 46.7% fewer measurements, respectively. This is further illustrated by the MSE curves shown in Fig. 11. In terms of computational complexity, the Exhaustive Scanning and 802.11ad can both *independently* process their measurement results in order to discover the channel's strongest paths. Our proposed approach, on the other hand, requires joint processing of all measurements, which, in turn, requires heavier computations. Specifically, the number of required computation steps is $m_2 \times \binom{n_r}{L} + L \times \binom{n_t}{L}$ (see Algorithm 1).

VIII. Conclusion

This work provides a solution for the mm-wave channel estimation problem by exploiting its sparse nature in the angular domain. The proposed solution is a beam discovery technique that is similar to error discovery in channel coding. We show that our proposed technique can significantly reduce the number of measurements required for reliable channel estimation. Our solution takes into account the size of TX/RX arrays and the sparsity level of the channel. We determine the number of measurements and the design of each measurement in a deterministic way; based on parity check matrices of appropriately selected LBCs. Under no measurement errors, our solution is guaranteed to find all available beams (paths) between TX and RX. However, due to the presence of channel noise and quantization (ADCs), measurement errors occur, which might cause incorrect beam discovery. Hence, we assess the performance of the proposed scheme under different levels of *SNR* and ADC resolutions. We further provide a technique for error correction that is also inspired by channel coding.
A special case of uncoded discovery within our general coded discovery framework is Exhaustive Scanning. We compare our solution against Scanning and find that we approach its error performance under the same total energy expenditure.

References

[1] Y. Shabara, C. E. Koksal, and E. Ekici, "Linear block coding for efficient beam discovery in millimeter wave communication networks," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 2285–2293.
[2] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
[4] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 922–928, May 2004.
[5] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
[6] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[10] In fact, performing mm-wave channel measurements at low SNR is common for non-adaptive channel measurement schemes due to the persistent use of wide antenna beams. This phenomenon is argued in [35] for compressed-sensing-based methods.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[8] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 90–97.

[9] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[10] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2015, pp. 191–198.

[11] J. Mo and R. W. Heath, Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2014, pp. 1–5.

[12] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Adaptive one-bit compressive sensing with application to low-precision receivers at mmWave," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[13] A. Zhou, X. Zhang, and H. Ma, "Beam-forecast: Facilitating mobile 60 GHz networks via model-driven beam steering," in *Proc. IEEE Conf. Comput. Commun.*, May 2017, pp. 1–9.

[14] H. S. Ghadikolaei, H. Ghauch, and C. Fischione, "Learning-based tracking of AoAs and AoDs in mmWave networks," in *Proc. 2nd ACM Workshop Millim. Wave Netw. Sens. Syst.*, Oct. 2018, pp. 45–50.

[15] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[16] S. Sun and T. S. Rappaport, "Millimeter wave MIMO channel estimation based on adaptive compressed sensing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 47–53.

[17] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 957–961.

[18] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 3rd Quart., 2017.

[19] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[20] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[21] H. Hassanieh, O. Abari, M. Rodreguez, M. Abdelghany, D. Katabi, and P. Indyk, "Agile millimeter wave networks with provable guarantees," Jun. 2017, *arXiv:1706.06935*. [Online]. Available: https://arxiv.org/abs/1706.06935

[22] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proc. 15th ACM Workshop Hot Topics Netw.*, Nov. 2016, pp. 169–175.

[23] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, "Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 601–616, Feb. 2017.

[24] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 2393–2401.

[25] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. 18th Int. Workshop Mobile Comput. Syst. Appl.* , Feb. 2017, pp. 13–18.

[26] N. J. Myers and R. W. Heath, Jr., "A compressive channel estimation technique robust to synchronization impairments," Jul. 017, *arXiv:1707.09441*. [Online]. Available: https://arxiv.org/abs/1707.09441

[27] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[28] C. N. Barati *et al.*, "Directional initial access for millimeter wave cellular systems," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2015, pp. 307–311.

[29] M. Hashemi, C. E. Koksal, and N. B. Shroff, "Rate-optimal power and bandwidth allocation in an integrated sub-6 GHz—Millimeter wave architecture," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Oct. 2017, pp. 1207–1211. doi: 10.1109/ACSSC.2017.8335543.

[30] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[31] J. H. Van Lint, *Introduction to Coding Theory* (Graduate Texts in Mathematics), vol. 86, 3rd ed. Berlin, Germany: Springer, 1999.

[32] W.-T. Li, Y.-C. Chiang, J.-H. Tsai, H.-Y. Yang, J.-H. Cheng, and T.-W. Huang, "60-GHz 5-bit phase shifter with integrated VGA phase-error compensation," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 3, pp. 1224–1235, Mar. 2013.

[33] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "BeamSpy: Enabling robust 60 GHz links under blockage," in *Proc. 13th Usenix Conf. Netw. Syst. Design Implement.*, Mar. 2016, pp. 193–206.

[34] Y. Shabara, C. E. Koksal, and E. Ekici, "Beam discovery using linear block codes for millimeter wave communication networks," May 2018, *arXiv:1805.12009*. [Online]. Available: https://arxiv.org/abs/1805.12009

[35] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 860–864.

Yahia Shabara received the B.Sc. degree in electrical engineering from Alexandria University, Alexandria, Egypt, in 2012, and the M.Sc. degree in wireless communications from Nile University, Giza, Egypt, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA. His research interests include wireless communications, computer networks, and machine learning.

C. Emre Koksal (S'96–M'03–SM'13) received the S.M. and Ph.D. degrees in electrical engineering and computer science from MIT in 1998 and 2002, respectively. He was a Post-Doctoral Fellow at MIT until 2004, and a Senior Researcher at EPFL until 2006. Since 2006, he has been with the Electrical and Computer Engineering Department, The Ohio State University, where he is currently a Professor. He is also the Founder of DAtAnchor, a data security start-up, and a number of his other technologies are also being commercialized by various companies. His general areas of interest are wireless communication, cybersecurity, communication networks, information theory, stochastic processes, and financial economics. He was a recipient of the National Science Foundation CAREER Award in 2011, OSU CoE Innovator Award in 2016, OSU CoE Lumley Research Award in 2011 and 2017, a finalist of the Bell Labs Prize in 2015, and the co-recipient of an HP Labs—Innovation Research Award in 2011. His coauthored papers received the Best Paper Award in IEEE WiOpt 2018 and the Best Student Paper Candidate in ACM MOBICOM 2005. He served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and *Computer Networks*.

Eylem Ekici (S'99–M'02–SM'11–F'17) received the B.S. and M.S. degrees in computer engineering from Boğaziçi University, Turkey, in 1997 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology in 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, The Ohio State University. His current research interests include dynamic spectrum access, vehicular communication systems, and next-generation wireless systems, with a focus on algorithm design and resource management. He was the General Co-Chair of ACM MobiCom 2012. He was also the TPC Co-Chair of the IEEE INFOCOM 2017. He is an Associate Editor-in-Chief of the IEEE TRANSACTIONS ON MOBILE COMPUTING, and a former Associate Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and *Computer Networks*.